DISSERTATION

in partial fulfillment of the requirements for the degree of
Doktor der Naturwissenschaften

Statistical Inference for Intensity-Based Load Sharing Models With Damage Accumulation

by

M.Sc. Mirko Alexander Jakubzik

Referees:	Prof. Dr. Christine Müller
	Prof. Dr. Markus Pauly
Commission chairperson:	JProf. Dr. Kirsten Schorning
Assessor:	Prof. Dr. Carsten Jentsch
Submitted:	December 4, 2023

Department of Statistics
Statistics with Applications in Engineering Sciences
TU Dortmund University

Acknowledgments

First of all, I would like to thank my supervisor Prof. Dr. Christine Müller for her continuous support and guidance in writing this thesis. When the initial research question seemed stuck, she gave me the freedom and encouraged me to start over and pursue a different direction. It is thanks to her numerous suggestions and our mathematical discussions that today both parts find their way into this finished work. I am grateful for her inspiring tenacity and tireless curiosity that made this possible.

Moreover, my gratitude belongs to the members and former members of the StABI chair at TU Dortmund University. A special thanks goes to Dr. Kevin Leckey, Dr. Dennis Malcherczyk and Dr. Melanie Horn for their helpful advice and for becoming more than just colleagues to me, and to Heike Große-Oetringhaus, who assisted me in many organizational matters and always brings a cheerful attitude to the office.

I am also thankful for my family and friends, they mean the world to me. I owe the greatest thanks to Dr. Nicole Hufnagel for her help, love and patience. She not only gave me the strength to keep going, but also contributed to the completion of this thesis with her proofreading and mathematical advice. A special thank you to Katharina Bata and Dr. Sascha Knüttel for not only accompanying me on my journey through the study of mathematics, but also for being my closest friends. Their support and encouragement is invaluable to me. I would like to write my final thanks to my family, especially my parents Martina and Wilfried, in German:

Mama, Papa, ohne eure Liebe und euren Glauben an mich wäre ich niemals so weit gekommen. Danke für eure unermüdliche Unterstützung auf all meinen Wegen. Dodo, ich weiß, dass du diesen Moment gerne erlebt hättest. Ich danke dir für all die Liebe, die du mir auf diese Reise mitgegeben hast.

Contents

1. Introduction

An important task in reliability studies is the lifetime testing of systems composed of dependent or interacting components (cf. Li and Lynch 2011, p. 2811). In such a system, the failure of a component affects the performance of the surviving components (Yaonan and Zhisheng 2015, p. 1). For example, Figure 1 shows a civil engineering experiment in which a cyclic load is applied to a system consisting of 35 tension wires. Whenever one of these wires breaks, the total exerted load is redistributed across the remaining tension wires. This in turn increases the individual load applied to each of the surviving wires and therefore their risk of failure. We call any system with this kind of behavior a *load sharing system*. Load sharing systems are by no means limited to the engineering

Figure 1: Tension wires broken due to cyclic loadings; Müller, Szugat and Maurer 2016.

sciences: we encounter them in a variety of fields such as software development (Kim and Kvam 2004, p. 84), organ subsystems (Yaonan and Zhisheng 2015, p. 1) or the strength testing of composite materials. Early considerations of load sharing systems date back to Daniels 1945, who developed a statistical framework to relate the strength of a textile to its constituent threads, and Rosen 1964, who employed Daniel's model in his experimental treatment of fibrous composites. These pioneering works assume an *equal load sharing rule* under which the load is equally redistributed among the surviving components whenever one component fails. Subsequent research has led to more general, but also increasingly complex classes of load sharing rules, see Lee, Durham and Lynch 1995, with Harlow and Phoenix 1982 and the preceding works of Phoenix being important references on *local load sharing rules*. Within this thesis, we always operate under an equal load sharing rule.

In an equal load sharing system, the failure risk of a surviving component increases with the *number* of failed components. However, the failure risk is likely to also depend on *how long* the surviving components were exposed to the redistributed load. Even between two consecutive component failures, the *accumulation of damage* within the system can therefore cause a continuous increase in the risk of component failure. The primary objective of this thesis is to provide statistical inference in the context of load sharing systems that are subject to damage accumulation. The key to this is to recognize the component failure rate as a stochastic process, for which we can establish a parametric model. In a system with I components, let $0 < T_1 < T_2 < \ldots < T_I$ denote the random failure times of the components. Then,

$$N_t := \sum_{i=1}^{I} \mathbb{1}_{(0,t]} (T_i)$$

counts the number of failures up to time t and we can regard $N = (N_t)_{t\geq 0}$ as a counting process. The Doob-Meyer decomposition allows us to decompose this process into a predictable process $\Lambda = (\Lambda_t)_{t\geq 0}$, known as the compensator of N, and a martingale $M = (M_t)_{t\geq 0}$ so that

$$N_t = \Lambda_t + M_t \,.$$

The predictable process Λ contributes the qualitative behavior of N and compensates for its monotonicity, while the trend-free martingale M provides the "unpredictable" randomness (cf. Kopperschmidt and Stute 2013, p. 1273). In particular,

$$\mathbb{E}\left(N_t\right) = \mathbb{E}\left(\Lambda_t\right) \,,$$

so that Λ serves as a basic predictor for the counting process N. If Λ admits a Lebesgue density $\Lambda(\mathrm{d}t) = \lambda_t \,\mathrm{d}t$, then $\lambda = (\lambda_t)_{t\geq 0}$ is called the stochastic intensity of N and satisfies

$$\lambda_t \,\mathrm{d}t = \mathbb{E}\left[N(\mathrm{d}t) \,|\, \sigma(\{N_s : s < t\})\right] = \mathbb{P}\left[N(\mathrm{d}t) = 1 \,|\, \sigma(\{N_s : s < t\})\right] \,,$$

where $\sigma(N_s : s < t)$ denotes the history of N up to but not including time t. Since $N(\mathrm{d}t) = 1$ means that a component failure occurs in $\mathrm{d}t$, λ_t can be interpreted as the instantaneous failure rate of the load sharing system represented by N at time t. Moreover, the stochastic intensity λ determines the probability structure of N uniquely. This means that a load sharing system is fully characterized by a stochastic intensity, or in other words, its component failure rate. Consequently, a model for load sharing systems with damage accumulation can be described by a parametric family of stochastic intensities, and we refer to such models as *intensity-based load sharing models with damage accumulation.*

While intensity-based models for load sharing systems have been studied by Kvam and Peña 2005, Spizzichino 2019 and Zhang, Zhao and Ma 2020, among others, the incorporation of damage accumulation into these models is relatively unexplored. Wang, Jiang and Park 2019, for instance, investigated load sharing systems with memory, that "consider how long a surviving component has worked for prior to the redistribution [of workload]" (Wang, Jiang and Park 2019, p. 341). Further examples were recently proposed by Müller and Meyer 2022. We adopt one of them - the Basquin load sharing model with multiplicative damage accumulation - as the core model of this dissertation. The model owes its name to the Basquin link function, which is derived from Basquin's exponential law of endurance (Basquin 1910) and relates the expected lifetime of a component to the current stress exerted on the system. This specific choice of a link function allows the model to be formulated as a relative risk regression model, a generalization of the well-known Cox proportional hazards model, and to provide statistical inference via the (partial) likelihood. The frequentist approach in Müller and Meyer 2022, who also consider Bayesian inference, relies on these likelihood-based methods.

The dissertation ties in with this point. We consider two further approaches to statistical inference for intensity-based models. The first is the minimum distance estimator introduced by Kopperschmidt and Stute 2013, the second is based on the K-sign depth of Leckey et al. 2023.

(i) *Minimum Distance Estimator.* This estimator presented by Kopperschmidt and Stute 2013 builds on the idea of minimizing the distance between a model compensator Λ_θ and the counting process N with respect to a parameter θ. Kopperschmidt and Stute 2013 claim the strong consistency and asymptotic normality of their estimator, but we demonstrate that their proof of the asymptotic distribution is flawed. We present a corrected proof under slightly adjusted requirements. Furthermore, we show that these requirements are met by the Basquin load sharing model with multiplicative damage accumulation.

(ii) *K-Sign Depth.* The K-sign depth of Leckey et al. 2023 emerged from a combination of the regression depth of Rousseeuw and Hubert 1999 and the simplicial depth of Liu 1990. The corresponding K-sign depth test is a more powerful but similarly robust generalization of the classical sign test. Leckey et al. 2023 applies the K-sign depth test to the residuals of a linear model. We explain how "residuals" can be obtained in an intensity-based point process model via the hazard transformation. We then derive conditions on the model under which the 3-sign depth test is consistent and prove that these conditions are satisfied by the Basquin load sharing model with multiplicative damage accumulation.

As our final contribution, we compare these methods in a simulation study with the likelihood approach previously studied by Müller and Meyer 2022. We place particular emphasis on a robustness study that evaluates the performance of the methods when applied to contaminated data. The study confirms that, in contrast to the competing methods, the 3-sign depth test offers both a powerful and robust tool for statistical inference in intensity-based point process models.

The dissertation is structured as follows: In Chapter 2, we introduce the framework for intensity-based models. We familiarize ourselves with the important notations and learn about relative risk regression models. We then motivate load sharing models both with and without damage accumulation and highlight the Basquin load sharing model with multiplicative damage accumulation. We briefly touch upon related models with damage accumulation, before we move on to develop uniform bounds for the intensity and its partial derivatives that are required for the asymptotic normality of the minimum distance estimator. Finally, we address the hazard transformation for intensity-based point process models and both discuss the distributional properties and give explicit formulae for the hazard transforms in the Basquin load sharing model with multiplicative damage accumulation. Chapter 3 deals with the minimum distance estimator of Kopperschmidt and Stute 2013. We give its definition and restate their result on strong consistency. The rest of the chapter is mainly devoted to the proof of the asymptotic normality and concludes with the application to our specific model. Chapter 4 revolves around the K-sign depth. After a short treatise on the origin of the K-sign depth, we get to know its definition and the asymptotic distribution, from which the K-sign depth test can be derived. We present the general consistency conditions for the 3-sign depth test given by Leckey, Jakubzik and Müller 2023, and identify criteria under which these are fulfilled by an intensity-based point process model. This involves the aforementioned hazard transforms, and necessitates an ordering of the transforms that is specific to load sharing models with damage accumulation. We also verify that the 3-sign depth test for the significance of damage accumulation is consistent in the Basquin load sharing model with multiplicative damage accumulation. In Chapter 5, we recapitulate the likelihood-based approach followed by Müller and Meyer 2022. We provide further insights on partial likelihoods, but beyond that only summarize and expand on their results. The comparison of the methods from Chapters 3, 4 and 5 via a simulation study is carried out in Chapter 6. We first describe how point process realizations from a given parametric intensity-based model can be simulated. Throughout this study, this will always be the Basquin load sharing model with multiplicative damage accumulation. From these point process realizations, for each of our methods we compute confidence regions for the true parameter of the parametric model and compare them in terms of size and coverage rate. We then evaluate the power of the respective tests for the significance of damage accumulation. We conclude the chapter with a robustness study in which we contaminate part of the data and study the effects on the competing methods. In the final Chapter 7, we close this thesis with a brief outlook for future research.

2. Intensity-Based Modelling

Chapter 2 is dedicated to the foundation of this thesis: the intensity-based models. We first introduce a framework for them in Section 2.1. Next, we learn about multiplicative intensity models and relative risk regression models in Section 2.2, before introducing specific models known as load sharing models in Section 2.3. In the process, we also become acquainted with the core model of our consideration, namely the Basquin load sharing model with damage accumulation. Some properties of this particular model are highlighted in the remaining two sections of the chapter. In Section 2.4 we provide uniform bounds for the intensity and its partial derivatives. They later become useful in the context of minimum distance estimation in Chapter 3. The final Section 2.5 deals with the hazard transformation of an intensity-based load sharing model, which forms the basis of our depth-related approaches in Chapter 4.

Throughout this thesis we will presume knowledge of common terminology such as simple point processes, compensators, or stochastic intensities. For the reader not familiar with these terms, we have compiled a “comprehensive introduction” in Appendix A, which motivates, defines and explains the concepts underlying this dissertation. Whilst we refer to this overview whenever we introduce a new object, we would like to encourage the experienced reader not to let it interfere with the flow of reading and ignore these references. For convenience, we also maintain a list of recurring symbols in the order of their first appearance, which can be viewed in Table 7 in Appendix C.

2.1. Framework for Intensity-Based Models

An intensity-based model aims to capture the qualitative behavior of counting processes. For our basic framework, let $N^{(1)}, \ldots, N^{(J)}$ with $J \in \mathbb{N}$ be stochastically independent copies of a counting process $N = (N_t)_{t\in\mathcal{I}}$ over some interval $\mathcal{I} \subset \mathbb{R}$ (see Definition A.5). In most cases, we consider either the compact interval $\mathcal{I} = [0, \tau]$, where $\tau \in (0, \infty)$ marks - for instance - the end of an experiment, or the positive real line $\mathcal{I} = [0, \infty) = \mathbb{R}_+$.
We assume that the processes are defined on a common filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in\mathcal{I}}, \mathbb{P})$ satisfying the *usual conditions* (see Definition A.14 for details). We always demand that N is adapted with respect to the filtration $\{\mathcal{F}_t\}_{t\in\mathcal{I}}$, and often opt for the natural or an intrinsic filtration of the counting process, that is, $\mathcal{F}_t = \mathcal{F}_t^N \vee \mathcal{G}_0$ (cf. Definitions A.10, A.11 and A.13). The counting processes are allowed to depend on random external covariates, which are then taken to be $\mathcal{G}_0$-measurable and i.i.d. for all repetitions of N. In the general case, we denote such a vector of real-valued random variables of arbitrary dimension by X or X_j. When we encounter random covariates in the specific models, a separate notation will be introduced.
The simple point processes associated with $N, N^{(1)}, \ldots, N^{(J)}$ (see Definition A.3) are denoted by $T, T^{(1)}, \ldots, T^{(J)}$, so that $T_i^{(j)}$ is the time of the ith event in the jth iteration of N, where $i \in \mathbb{N}$ and $j \in \{1, \ldots, J\}$. We indicate a realization with a lowercase letter, for example $T_i^{(j)}(\omega) = t_i^{(j)}$. Furthermore, we use $\Lambda, \Lambda^{(1)}, \ldots, \Lambda^{(J)}$ for the compensators given by the Doob-Meyer decomposition (see Theorem A.23), which means that

$$M^{(j)} := N^{(j)} - \Lambda^{(j)}, \quad j \in \{1, \ldots, J\},$$

and $M := N - \Lambda$ are $(\mathbb{P}, \{\mathcal{F}_t\}_{t\in\mathcal{I}})$-martingales. We usually assume that the counting processes admit $(\mathbb{P}, \{\mathcal{F}_t\}_{t\in\mathcal{I}})$-intensities $\lambda, \lambda^{(1)}, \ldots, \lambda^{(J)}$, so that their compensators are absolutely continuous with respect to the Lebesgue measure and satisfy (cf. Appendix A.2.4)

$$\Lambda^{(j)}(t) = \int_0^t \lambda^{(j)}(u)\,\mathrm{d}u\,, \quad j \in \{1,\ldots,J\}\,.$$

We thus refer to the compensators in the following as "cumulative intensity processes" or, for brevity, as "cumulative intensities". We call any model for these cumulative intensities an "intensity-based model".

The intensity-based models in this thesis are parametric. Let $\theta \in \Theta \subset \mathbb{R}^d$ be the parameter of interest, where $d \in \mathbb{N}$ and the parameter space Θ is assumed to be compact or at least bounded throughout most applications. An intensity-based model $\mathcal{M}$ is a parametric class of cumulative intensities,

$$\mathcal{M} := \{\Lambda_\theta \,:\, \theta \in \Theta\}\,, \tag{2.1}$$

such that the true cumulative intensity process Λ of N satisfies

$$\Lambda = \Lambda_{\theta^*}\,, \quad \text{for some parameter } \theta^* \in \Theta\,. \tag{2.2}$$

Accordingly, we refer to θ^* as the "true parameter". By the fundamental theorem of calculus, the model in Equation (2.1) can equivalently be stated in terms of stochastic intensities λ_θ, as then it holds:

$$\Lambda_\theta(t) = \int_0^t \lambda_\theta(u)\,\mathrm{d}u\,. \tag{2.3}$$

In fact, we construct all models within this thesis by virtue of the easier-to-interpret intensity process. The actual model in the sense of Equation (2.1) can then be obtained by applying Equation (2.3).

Because the intensities are subject to randomness, this also applies to the model $\mathcal{M}$. In order to specify a model completely, we have to be able to state it for each individual j,

$$\mathcal{M}_j := \{\Lambda_\theta^{(j)} \,:\, \theta \in \Theta\}\,, \tag{2.1$*$}$$

and demand that Equation (2.2) is fulfilled for every j, that is,

$$\Lambda^{(j)} = \Lambda_{\theta^*}^{(j)}\,, \quad \text{for some parameter } \theta^* \in \Theta \text{ and all } j \in \{1,\ldots,J\}. \tag{2.2$*$}$$

Hereafter, we will usually not state models as parametric classes of cumulative intensities, but rather specify the cumulative intensities $\Lambda_\theta^{(j)}$ or the intensity processes $\lambda_\theta^{(j)}$ directly. The most accessible route to an intensity-based model is to conceive the intensity process as a piecewise amalgamation of conditional hazard functions. We will demonstrate this approach below. Since we frequently work with conditional probabilities, we introduce an abbreviated notation first.

Remark 2.1 (Abbreviated Notation for Conditional Probabilities)**.**
To shorten the formulas involving conditional probabilities, we use the notation

$$T_{1:(i-1)} := (T_1,\ldots,T_{i-1})^\top\,,$$

that can be similarly defined for $(t_1, \ldots, t_{i-1})^\top$ and other vectors. We then write

$$\mathbb{P}(\,\cdot\,|\,T_{1:(i-1)} = t_{1:(i-1)}) \quad \text{instead of} \quad \mathbb{P}\left(\cdot\,|\,T_1 = t_1, \ldots, T_{i-1} = t_{i-1}\right) \,,$$

and analogously for conditional densities, cumulative distribution or hazard functions.

According to Jacod's formula, the conditional distribution of the ith point T_i given the past of the corresponding counting process N follows a continuous distribution (see Corollary A.34). This past comprises knowledge of $T_1, \ldots, T_{i-1}$ as well as the random covariate X. Let therefore $f_i(t\,|\,t_{1:(i-1)}, x)$ be the conditional density function of T_i after the observation of $T_{1:(i-1)} = t_{1:(i-1)}$ and $X = x$. Let $S_i(t\,|\,t_{1:(i-1)}, x)$ be the associated survival function (details are available in Summary 1 of Appendix A.2.4).
The conditional hazard function is the continuous function given by

$$h_i(t\,|\,t_{1:(i-1)}, x) := \frac{f_i(t\,|\,t_{1:(i-1)}, x)}{S_i(t\,|\,t_{1:(i-1)}, x)} \,, \quad t \geq t_{i-1} \,. \tag{2.4}$$

The conditional intensity function λ^* is defined piecewise as follows:

$$\lambda^*(t) := \begin{cases} h_1(t\,|\,x) \,, & 0 \leq t < t_1 \,, \\ h_i(t\,|t_{1:(i-1)}\,, x) \,, & t_{i-1} \leq t < t_i \,, i \geq 2 \,. \end{cases} \tag{2.5}$$

The function λ^* is right-continuous and fulfills the identity

$$\Lambda(t) = \int_0^t \lambda^*(u)\,\mathrm{d}u \,, \quad \text{for all } t \in \mathcal{I}. \tag{2.6}$$

Recall that the intensity process λ emerges as the Radon-Nikodym derivative of the compensator Λ with respect to the Lebesgue measure, so that

$$\Lambda(t) = \int_0^t \lambda(u)\,\mathrm{d}u \,, \quad \text{for all } t \in \mathcal{I}.$$

The conditional intensity function λ^* therefore coincides with the intensity process λ almost everywhere. Since both Equations (2.5) and (2.6) are specific to the particular realization, λ^* is itself a stochastic process. By taking a left-continuous (and thus predictable) modification of λ^*, it can hence be identified with λ.
As a result, Equation (2.5) enables us to construct models for stochastic intensities that incorporate both external influences in the form of X and internal changes following an event T_i. These internal changes can also access the further past $T_1, \ldots, T_{i-1}$.
Note, however, that we may not allow for extrinsic shocks over the course of time as long as an intrinsic filtration is considered. Thus, all information that is not generated by the counting process itself must already be available at the beginning. In other words, any random variable representing external information must be $\mathcal{G}_0$-measurable.

2.2. Multiplicative Intensity and Relative Risk Regression Models

In Section 2.1, we laid the foundation for specific intensity-based models. The idea to model the dynamics of a counting process by virtue of its stochastic intensity dates back to the 1970s, when Aalen introduced the multiplicative intensity model in his article Aalen 1978, p. 707 and the less accessible dissertation Aalen 1975, respectively. In the multiplicative intensity model, $\lambda^{(j)}$ is assumed to take the form

$$\lambda^{(j)}(t) := \alpha(t) Y_j(t)\,, \quad j = 1, \ldots, J\,, \tag{2.7}$$

where α is an unknown non-negative deterministic function, the *baseline intensity*, while Y_j is a predictable stochastic process for each $j \in \{1, \ldots, J\}$ (see Definition A.20). Aalen studied the non-parametric setting "that arises by letting α vary freely" (Aalen 1978, p. 707) in a set $\mathcal{A}$ that is subject to further constraints. We obtain a first parametric model if we instead take $\mathcal{A}$ to be a parametric family of suitable functions, that is,

$$\mathcal{A} = \{\alpha(\cdot, \theta) \,:\, \theta \in \Theta\}\,.$$

Thus, the intensity process also depends on the parameter θ,

$$\lambda_\theta^{(j)}(t) = \alpha(t, \theta) Y_j(t)\,, \quad j = 1, \ldots, J\,, \tag{2.8}$$

and we obtain an intensity-based model $\mathcal{M}_j$ by integrating these intensities.
The multiplicative intensity model is known to be the "broadest setting in which there exists a good asymptotic theory given i.i.d. copies of an underlying process" (Karr 1991, p. 172). However, here the parameter θ only affects the deterministic baseline intensity and not the random component Y_j specific for the jth "individual" (i.e., the jth iteration of the underlying counting process N).
A popular model that complements the approach of parametric multiplicative intensities is the basic regression model of Cox 1972. This model incorporates covariates influencing the intensity "in such a way that hazard functions for different individuals are mutually proportional" (Karr 1991, p. 200). Because of this characteristic, it is also known as the proportional hazards model. Cox proposed the following intensity for the jth individual (see Cox 1972, p. 189):

$$\lambda_\theta^{(j)}(t) := \alpha(t) \exp\left(\theta^\top z_j\right), \quad j = 1, \ldots, J\,, \tag{2.9}$$

where α is again the baseline intensity and $z_j \in \mathbb{R}^d$ is a vector of deterministic covariates. Cox studied the asymptotic properties of a maximum partial likelihood estimator for θ (Cox 1972 and Cox 1975). Andersen and Gill later studied an extension of Cox's model with time-dependent covariate processes and proved that the consistency and asymptotic normality of Cox's estimator carry over to their model (Andersen and Gill 1982). They assume that the intensity process $\lambda_\theta^{(j)}$ has the shape (cf. Andersen and Gill 1982, p. 1102)

$$\lambda_\theta^{(j)}(t) := \alpha(t) \exp\left(\theta^\top Z_j(t)\right) Y_j(t)\,, \quad j = 1, \ldots, J\,. \tag{2.10}$$

Here, Y_j is again a predictable stochastic process, but is assumed to be taking values in $\{0, 1\}$. $Y_j(t)$ usually indicates whether the jth individual is under observation at time t and thus implements a censoring scheme. The covariate process Z_j is supposed to be predictable and locally bounded, which is always the case if it is adapted and left-continuous with right limits (cf. Andersen et al. 1993, pp. 62–63). If $Z_j \equiv 0$ with probability one, the model reduces to the multiplicative intensity model (2.7). If on the other hand $Z_j \equiv z_j$ and $Y_j \equiv 1$ with probability one, the Cox regression model (2.9) is obtained. The extension of Andersen and Gill thus emerges as the canonical confluence of these two approaches.

Two final modifications to the model (2.10) can still be made: First, in analogy to the parametric multiplicative intensity model (2.8), the baseline intensity may also depend on a parameter, and second, an arbitrary non-negative twice differentiable function can be used in place of the exponential function (the latter is due to Prentice and Self 1983). The resulting class of models is known as relative risk regression models, whose intensity process is given by (compare Andersen et al. 1993, pp. 477–478)

$$\lambda_\theta^{(j)}(t) := \alpha(t,\gamma)\, r(\beta^\top Z_j(t))Y_j(t)\,, \quad j = 1,\dots,J\,, \tag{2.11}$$

where $\theta = \left(\gamma^\top, \beta^\top\right)^\top$ and r is a non-negative twice differentiable function. The model's name originates from the observation that for any two individuals $j_1, j_2 \in \{1,\dots,J\}$, the ratio of their intensities on $\{Y_{j_1}(t) = 1\} \cap \{Y_{j_2}(t) = 1\}$, that is,

$$\frac{\lambda_\theta^{(j_1)}(t)}{\lambda_\theta^{(j_2)}(t)} = \frac{r(\beta^\top Z_{j_1}(t))}{r(\beta^\top Z_{j_2}(t))}\,,$$

is determined only by the regression part of the model (i.e., the covariate processes Z_{j_1}, Z_{j_2} and the regression coefficients β) and specified via the *relative risk function* r (see Andersen et al. 1993, p. 477).
All models studied in this dissertation can be assigned to one of the classes presented in this section. In particular, this implies that we will be able to compare our methods of statistical inference with the established maximum likelihood estimation.

2.3. Load Sharing Models

This section is divided into two parts: In the first part, we learn about some models for implementing equal load sharing rules. A very basic model is introduced to motivate the concept of load sharing systems. It provides the foundation for the model of Leckey et al. 2020, in which lifetimes of components are related to the current stress of a system through a parametric link function. We motivative the choice of the Basquin link, which is derived from S-N curve models of fatigue testing. We also address the flexible semiparametric model of Kvam and Peña 2005, although it is of less importance to us because of the larger sample sizes it typically requires. However, it essentially encompasses all equal load sharing models within the scope of this thesis as special cases. The formulation of the Basquin load sharing system without damage accumulation concludes the first part.
In the second part, we illustrate the need for a concept of accumulating stress. Based on the article of Müller and Meyer 2022, we introduce a damage accumulation term that extends the Basquin load sharing system to a model with damage accumulation. Finally, we address related models with damage accumulation and hint at a class of generalized models as a starting point of future research.

2.3.1. Load Sharing Models Without Damage Accumulation

We consider J independent systems, each consisting of $I_j \in \mathbb{N}$, $j = 1,\dots,J$ parallel components to which a load is applied (e.g., J concrete beams with I tension wires each; see Maurer, Heeke and Marzahn 2012 and Szugat et al. 2016 for reference on this example). In these system, components successively fail due to the resulting stress. Under an *equal load sharing rule*, the exerted load is equally redistributed among the surviving

components whenever one component fails. In contrast, under a *local load sharing rule*, the load is instead transferred to the nearest surviving neighbours in order to account for local stress concentrations (cf. Lee, Durham and Lynch 1995 on general load sharing rules). While in some fields of application local load sharing rules may be advantageous, we focus solely on equal load sharing rules in this thesis.
We suppose that the failure times of the jth system form a simple point process $T^{(j)}$. Consequently, we assume that never two or more components fail at the same time. The associated counting process $N^{(j)}$ then counts the component failures, so that $N_t^{(j)}$ indicates the number of failed components until time $t \in \mathcal{I}$.
The central idea for modelling a load sharing system is that any change in the current stress level translates directly into a change in the underlying intensity process. A very basic equal load sharing system can then be realized as a multiplicative intensity model, where

$$\lambda_\theta^{(j)}(t) = \alpha(t, \theta) \underbrace{\left(s_j \frac{I_j}{I_j - N_{t^-}^{(j)}} \right) \mathbb{1}_{\left\{ N_{t^-}^{(j)} < I_j \right\}}}_{=Y_j(t)} . \tag{2.12}$$

Here, $s_j > 0$ is the individual initial load or stress in the jth system and

$$N_{t^-} := \lim_{s \uparrow t} N_s$$

is the left-hand limit of N at t, which ensures the left-continuity of $\lambda_\theta^{(j)}$. At each point of time t, the initial stress s_j is thus distributed equally over the remaining $I_j - N_{t^-}^{(j)}$ components.
To comply with the i.i.d. assumptions, we take the s_j to be random and i.i.d. as well, while usually assuming $I_j \equiv I$ with probability one (i.e., the total number of components remains constant over all repetitions). Either way, the (s_j, I_j) take the role of the external covariates X from Section 2.1. According to Equation (2.5), the intensity function can be specified in terms of the conditional hazard functions. Here we obtain:

$$h_i^\theta\Big(t \,\big|\, t_{1:(i-1)}^{(j)}, s_j, I_j\Big) = \begin{cases} \alpha(t, \theta) \left(s_j \frac{I_j}{I_j - (i-1)} \right) , & \text{if } i \le I_j, \\ 0 \, , & \text{otherwise,} \end{cases} \tag{2.13}$$

because $N_{t^-}^{(j)} = i - 1$ for $t \in \big(T_{i-1}^{(j)}(\omega), T_i^{(j)}(\omega)\big]$. We observe that these conditional hazard functions do not depend on j in the i.i.d. situation, but their arguments obviously do.
Equation (2.13) shows us that the risk of a component failure in the jth system is given by a baseline hazard $s_j \alpha(t, \theta)$ times a factor $\frac{I_j}{I_j - (i-1)}$ that increases in inverse proportion to the fraction of components remaining. For example, when half of the components have failed, the conditional hazard rate has increased to $2 s_j \alpha(t, \theta)$. This is consistent with the intuition that the total load of the system is then distributed among half of its components, so that the individual stress of each component has doubled.
A major drawback of this simple model is that the parameter only affects the baseline hazard. As a result, the magnitude of change that follows each component failure is known. More sophisticated approaches, on the other hand, allow for unknown load sharing rules (cf. Kim and Kvam 2004 and Kvam and Peña 2005 for introductory motivation). Kvam and Peña 2005, p. 264 propose the following model, which we adjusted only with respect

to the notation:

$$h_i^\theta\Big(t\,|\,t_{1:(i-1)}^{(j)},\tau_j\Big) = \begin{cases}\alpha(t)\theta_{i-1}\left(I-(i-1)\right)\cdot \mathbb{1}_{\{t\leq\tau_j\}}\,, & \text{if } i\leq I,\\ 0\,, & \text{otherwise.}\end{cases} \tag{2.14}$$

Here, $\alpha(t)$ is again a baseline hazard and $I_j \equiv I$. The random covariate τ_j marks the end of the jth experiment, so that the predictable process $\mathbb{1}_{\{t\leq\tau_j\}}$ implements a type I censoring scheme, as was already foreshadowed in the previous section (see explanations following Equation (2.10)). The unknown load sharing rule is represented by the parameter vector $\theta = (\theta_0,\theta_1,\ldots,\theta_{I-1})^\top$ that governs how the failure rate changes after each component failure. The conditional hazard function of the load sharing system is then composed of the individual risk $\alpha(t)\theta_{i-1}$ of a single component after $i-1$ failures multiplied by the number $I-(i-1)$ of remaining components.
Model (2.14) provides a load sharing rule with utmost flexibility, but in return relies on a usually unknown baseline hazard as well as the I unknown parameters $\theta_0,\theta_1,\ldots,\theta_{I-1}$. Statistical inference in this model therefore mostly requires large data samples. However, in many fields such as civil engineering, experiments are costly and only few observations are available. In such situations, the number of parameters can be reduced through link functions (cf. Balakrishnan, Beutner and Kamps 2011, p. 605).
In Leckey et al. 2020, p. 1897 a model is considered that extends the basic load sharing model (2.12) by a parametric link function, that is,

$$\lambda_\theta^{(j)}(t) = g_\theta\left(s_j\frac{I}{I-N_{t-}^{(j)}}\right)\mathbb{1}_{\left\{N_{t-}^{(j)}<I\right\}}\,, \tag{2.15}$$

where g_θ is a deterministic non-decreasing function for each $\theta\in\Theta$. Note that once again $I_j\equiv I$. Furthermore, a homogeneous (i.e., time-independent) baseline hazard is used, so that $\alpha(t,\theta)\equiv\alpha(\theta)$ can be incorporated into the link function g_θ. All models considered here have in common that their intensities depend on the past of the counting process only by the number of failed components. This means that the associated counting process is a *pure birth process* (see for example Snyder and Miller 1991, pp. 95–97). In particular, model (2.15) leads to a homogeneous birth process, where the conditional hazard functions are constant,

$$h_i^\theta\Big(t\,|\,t_{1:(i-1)}^{(j)},s_j\Big) = \begin{cases} g_\theta\left(s_j\frac{I}{I-(i-1)}\right)\,, & \text{if } i\leq I,\\ 0\,, & \text{otherwise.}\end{cases}$$

In such a case, the intensity function is piecewise constant, and the interarrival times (i.e., the waiting times between two component failures),

$$W_i^{(j)} := T_i^{(j)} - T_{i-1}^{(j)}\,,\quad i=1,\ldots,I\,,\ j=1,\ldots,J\,,$$

follow an exponential distribution according to Corollary A.47,

$$W_i^{(j)}\sim\mathcal{E}\left(g_\theta\left(s_j\frac{I}{I-(i-1)}\right)\right)\,.$$

The logarithmic expected waiting times can then be calculated to be

$$\ln\left(\mathbb{E}\left(W_i^{(j)}\right)\right) = -\ln\left(g_\theta\left(s_j\frac{I}{I-(i-1)}\right)\right)\,, \tag{2.16}$$

so that the link function g_θ relates expected lifetimes to the current stress exerted on the system. In the search for an appropriate link function, we are faced with the task of "establishing an equation which represents the relation between applied stress and some average value of the fatigue life" (Weibull 1961, p. 174). In fatigue testing, such an equation is called an S-N curve model, where S and N in the acronym stand for the stress and the number of loading cycles at the time of failure (i.e., the "lifetime"), respectively (cf. Burhan and Kim 2018, p. 1). The name is derived from the S-N diagrams, dating back to August Wöhler, in which the results of fatigue tests are plotted in lin-log or log-log scale (cf. Weibull 1961, p. 147). Starting from Basquin's "exponential law of endurance", see Basquin 1910, numerous S-N equations have been proposed. For example, a selection can be found in Weibull 1961, pp. 175–178 or Kohout and Věchet 2001, p. 176, while an evaluation is available in Burhan and Kim 2018. In this thesis, we will largely focus on the aforementioned Basquin link.

Definition 2.2 (Basquin Link Function; cf. Basquin 1910)**.**
The Basquin link is defined as the parametric function $g_\theta : [0,\infty) \to [0,\infty)$ *with*

$$g_\theta(x) := \theta_1 x^{\theta_2}\,, \quad \theta = (\theta_1, \theta_2)^\top \in [0,\infty)^2\,. \tag{2.17}$$

Plugging Equation (2.17) into Equation (2.16) yields

$$\ln\left(\mathbb{E}\left(W_i^{(j)}\right)\right) = -\ln(\theta_1) - \theta_2 \ln(x)\,.$$

If we identify $\mathbb{E}(W_i^{(j)})$ with the "lifetime" N (usually the <u>n</u>umber of load cycles until failure in engineering sciences; not to be confused with the counting process!) and x with the "stress" S, this can be equivalently stated as

$$\ln \mathrm{N} = -\ln(\theta_1) - \theta_2 \ln \mathrm{S} \quad\Longleftrightarrow\quad \mathrm{S} = \theta_1^{-\frac{1}{\theta_2}} \mathrm{N}^{-\frac{1}{\theta_2}}\,,$$

which closely resembles the original power law of Basquin (cf. Weibull 1961, p. 174).
An alternative parametrization for the Basquin link is given by

$$g_{\tilde\theta}(x) = \exp\left(-\tilde\theta_1 + \tilde\theta_2 \ln(x)\right), \quad \tilde\theta = (\tilde\theta_1, \tilde\theta_2)^\top \subset \mathbb{R} \times [0,\infty)\,. \tag{2.18}$$

For $\theta_1 > 0$, it coincides on $(0,\infty)$ with Equation (2.17) by setting $\tilde\theta = (-\ln(\theta_1), \theta_2)$. The usage of the exponential function has the advantage that the positivity of $g_{\tilde\theta}(x)$ is always guaranteed. Equation (2.18) is also the preferred representation of Leckey et al. 2020, who consider the Basquin link too. For this particular choice of g_θ, model (2.15) can be written as

$$\lambda_{\tilde\theta}^{(j)}(t) = \exp\left(\tilde\theta^\top Z_j(t)\right) Y_j(t)\,,$$

where

$$Z_j(t) = \left(-1,\, \ln\left(s_j \frac{I}{I - N_{t^-}^{(j)}}\right)\right)^\top \qquad \text{and} \qquad Y_j(t) = \mathbb{1}_{\left\{N_{t^-}^{(j)} < I\right\}}\,.$$

Conveniently, the intensity process $\lambda_{\tilde\theta}^{(j)}$ therefore has the form of a relative risk regression model, compare Equation (2.10). In our effort to conceptualize load sharing systems as specific relative risk regression models, we will adopt the model of Leckey et al. 2020 as our central load sharing model without damage accumulation.

The Basquin Load Sharing Model Without Damage Accumulation

Before we can specify the model itself, we introduce some model-specific random covariates. We assume throughout this paragraph that each system consists of I components (i.e., $I_j \equiv I$). Every such system corresponds to an experimental run in which component failures are monitored up to a random time τ_j. Thus, τ_j marks the end of the jth experiment, although not all components must have failed at this point: the observable data is therefore subject to random type I censoring, as we have previously seen with model (2.14) of Kvam and Peña 2005 (cf. Klein and Moeschberger 2003, pp. 64–70 for more details on right censoring). In order to comply with our assumptions, we require that

$$\tau_1, \ldots, \tau_J \overset{\text{i.i.d.}}{\sim} \mathbb{P}^{\tau_0}$$

for some probability measure $\mathbb{P}^{\tau_0}$ on $\mathcal{B}(\mathcal{I})$, the Borel σ-algebra over $\mathcal{I}$. In the case $\mathcal{I} = [0, \tau]$ it is ensured that $\tau_j \leq \tau$ holds for $j = 1, \ldots, J$. Moreover, by choosing, for example, $\mathbb{P}^{\tau_0} = \delta_\tau$, the non-random case can be covered as well, where δ_τ denotes the Dirac measure centred on τ.

Likewise, we allow for a second layer of censoring where the number $C_j \leq I$ of observable component failures for the jth system is randomly chosen in advance. We assume that

$$C_1, \ldots, C_J \overset{\text{i.i.d.}}{\sim} \mathbb{P}^{C_0},$$

where $\mathbb{P}^{C_0}$ is a probability measure on $2^{\{1,\ldots,I\}}$. In the absence of censoring, we can choose $\mathbb{P}^{C_0} = \delta_I$. Therefore, it holds $C_j = I$ almost surely so that the failure of all components is observable for each system. Note that this is not necessarily given in practice, where often the experiment has to be stopped once a critical amount of components has failed. For this reason, one often opts for $\mathbb{P}^{C_0} = \delta_{I_c}$ instead, where $I_c \in \{1, \ldots, I\}$ denotes the critical number of component failures. Similarly, other probability measures with support $\{1, \ldots, I\}$ can be considered if the critical number of component failures is itself random. While the first instance of censoring implemented random type I censoring at the system level, this second instance resembles random type II censoring at the component level within individual systems. In our setting, however, these components are usually not independent, so we will refrain from using this terminology and simply refer to random (right-)censoring instead. Moreover, we could weaken our assumptions to a non-informative independent censoring scheme, see Kalbfleisch and Prentice 2002, pp. 195–196, but for our purposes this distinction is inessential.

Finally, we suppose that the systems are exposed to different initial stress levels $s_1, \ldots, s_J$, on which we impose similar restrictions as before, that is,

$$s_1, \ldots, s_J \overset{\text{i.i.d.}}{\sim} \mathbb{P}^{s_0}$$

for some probability measure $\mathbb{P}^{s_0}$ on $\mathcal{B}([0, \infty))$. We already encountered these covariates in the basic model (2.12) as well as the model (2.15) of Leckey et al. 2020. In most applications, we can consider $\mathbb{P}^{s_0}$ to be a discrete distribution whose support consists of a preset assortment of positive stress levels.

Hereafter, we will mostly assume that the τ_j, C_j and s_j are stochastically independent. This may not always be consistent with practice, where higher stress levels favour shorter experiment durations, indicating a negative correlation between s_j and τ_j. Nevertheless, the independence assumption primarily serves to verify certain preconditions and simplifies

several model-specific proofs later on. It may in general be dropped as long as the applicability of the discussed methods can still be ensured.
We now summarize the essential assumptions in condensed form for future reference.

Assumption 2.3 (Random Covariates in the Basquin Load Sharing Model)**.**
We consider an intrinsic filtration $\{\mathcal{F}_t\}_{t\in\mathcal{I}}$, so that $\mathcal{F}_t = \mathcal{F}_t^N \vee \mathcal{G}_0$ for some σ-algebra $\mathcal{G}_0 \subset \mathcal{F}$. Let

$$\begin{aligned}
\tau_0 &: (\Omega, \mathcal{F}) \to (\mathcal{I}, \mathcal{B}(\mathcal{I}))\ , && \text{(end of the experiment)}\\
C_0 &: (\Omega, \mathcal{F}) \to \left(\{1,\dots,I\}, 2^{\{1,\dots,I\}}\right), && \text{(number of observable failures)}\\
s_0 &: (\Omega, \mathcal{F}) \to (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))\ , && \text{(initial stress level)}
\end{aligned}$$

be stochastically independent and $\mathcal{G}_0$-measurable random variables. Set $X = (\tau_0, C_0, s_0)$ and let $X_1, \dots, X_J$ denote i.i.d. copies of X, such that $(N, X), (N^{(1)}, X_1), \dots, (N^{(J)}, X_J)$ are stochastically independent. The random covariate associated with the jth experiment is then given by X_j, $j = 1, \dots, J$.

With these final assumptions in mind, we can now state the Basquin load sharing model without damage accumulation. As pointed out earlier, its formulation without random covariates is due to Leckey et al. 2020.

Definition 2.4 (Basquin Load Sharing Model Without Damage Accumulation)**.**
In the framework of Section 2.1 and under Assumptions 2.3, the Basquin load sharing model without damage accumulation is given via the intensity process

$$^{\mathrm{B}}\lambda_\theta^{(j)}(t) := \theta_1 \left(s_j \frac{I}{I - N_{t^-}^{(j)}} \right)^{\theta_2} \cdot \mathbb{1}_{\left\{N_{t^-}^{(j)} < C_j\right\} \cap \left\{t \le \tau_j\right\}}\ , \quad \theta = (\theta_1, \theta_2)^\top \in \mathbb{R}_+^2\ , \tag{2.19}$$

where the superscript B *(for* B*asquin) serves as a model indicator to help distinguish the intensities of multiple models later on.*

We have previously seen that the notion of load redistribution is implemented through the fraction $s_j \frac{I}{I - N_{t^-}^{(j)}}$. We hereafter refer to it as the “load sharing term”.

Definition 2.5 (Load Sharing Term)**.**
In the setting of the Basquin load sharing model without damage accumulation, the load sharing term of the jth system is defined as

$$B_j(t) := s_j \frac{I}{I - N_{t^-}^{(j)}}\ , \quad j = 1, \dots, J\ , \quad t \in \mathcal{I}\ . \tag{2.20}$$

Under Assumptions 2.3, $\{\mathcal{F}_t\}_{t\in\mathcal{I}}$ is an intrinsic filtration of N, so that

(a) the counting process $N^{(j)}$ is $\{\mathcal{F}_t\}_{t\in\mathcal{I}}$-adapted as an i.i.d. copy of N, and

(b) the random variable s_j is $\mathcal{F}_t$-measurable for each $t \in \mathcal{I}$.

Then, $B_j(t)$ is $\mathcal{F}_t$-measurable as a (left-hand) limit of $\mathcal{F}_t$-measurable mappings. Moreover, $B_j(t)$ is left-continuous by construction. As a stochastic process, the load sharing term $(B_j(t))_{t\in\mathcal{I}}$ is therefore $\{\mathcal{F}_t\}_{t\in\mathcal{I}}$-predictable.

2.3.2. Load Sharing Models With Damage Accumulation

In the recent article Müller and Meyer 2022, it is suggested that the risk of a component failure should not be determined only by the *number* of remaining components. Instead, the risk is likely to depend on *how long* the surviving components were exposed to the redistributed load. They assume that degradation processes lead to an accumulation of damage (cf. Müller and Meyer 2022, p. 2). The idea behind this is visualized in Figure 2: Let us consider two independent experiments conducted under identical conditions. In

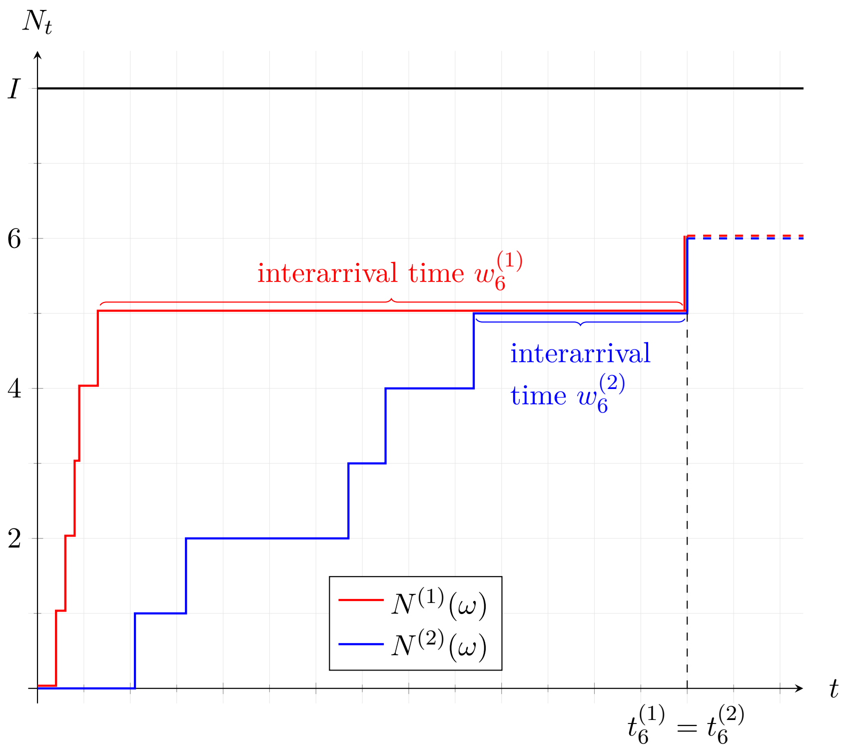


Figure 2: Schematic plot of two paths $t \mapsto N_t^{(1)}(\omega)$ (red) and $t \mapsto N_t^{(2)}(\omega)$ (blue) to demonstrate the idea of damage accumulation. We added vertical lines at the discontinuities (i.e., the "jumps" of the paths). The path of $N^{(1)}$ is shifted slightly in vertical direction to improve readability.

these experiments, a load is applied to a system consisting of $I = 8$ components. The number of component failures until time t in the jth experiment is denoted by $N_t^{(j)}$. In our example, we assume that the sixth component failed at the same time $t_6^{(1)} = t_6^{(2)}$ for both experiments. The paths of the counting processes $N^{(1)}$ and $N^{(2)}$ up to time $t_6^{(1)} = t_6^{(2)}$ are plotted in Figure 2. For any model from Subsection 2.3.1, $\lambda_\theta^{(1)}(t) = \lambda_\theta^{(2)}(t)$ then holds for all $\theta \in \Theta$ as long as $t_6^{(1)} < t \leq \min\{T_7^{(1)}, T_7^{(2)}\}$. This means that the risk of the seventh component failure is the same for both experiments. Nevertheless, the plot shows that a large number of components failed early on in the first experiment. The surviving components were thus exposed to the entire load of the system for almost the entire duration of the experiment, as can be seen from the interarrival time $w_6^{(1)}$. In contrast, in the second experiment, the load was distributed over a larger number of components most of the time (in particular, $w_6^{(2)} \ll w_6^{(1)}$). We would therefore expect that $\lambda_\theta^{(1)}(t) > \lambda_\theta^{(2)}(t)$ should apply. However, this behavior cannot be achieved by the previous models, and motivates the introduction of a "damage accumulation term".

Definition 2.6 (Damage Accumulation Term)**.**
In the setting of the Basquin load sharing model without damage accumulation of Definition 2.4, the damage accumulation term of the j*th system is defined as*

$$A_j(t) := \frac{1}{\tau}\int_0^t s_j \frac{I}{I - N_{u^-}^{(j)}}\,\mathrm{d}u\,, \quad j = 1, \dots, J\,. \tag{2.21}$$

As before, τ *marks the deterministic termination time of the experiment if* $\mathcal{I} = [0, \tau]$*. In the case* $\mathcal{I} = [0, \infty)$*,* τ *can be chosen arbitrarily.*

As a deterministic integral of a predictable process, the damage accumulation term is also predictable as a stochastic process $(A_j(t))_{t\in\mathcal{I}}$ (see remarks following Definition 2.5). The damage accumulation term $A_j(t)$ accumulates the stress until time t in the sense of load sharing. Thereby, $A_j(t)$ takes into account how long the stress was distributed over the remaining components. In the example from Figure 2, we obtain $A_1(t) > A_2(t)$ for all $t \leq t_6^{(1)}$, since the paths do not intersect.
We note that, at $t = \tau$, the damage accumulation term corresponds to the mean value of the load sharing term B_j on the interval $[0, \tau]$. We can therefore think of τ as a scaling factor to make the load sharing term B_j and the damage accumulation term A_j comparable.
The damage accumulation term can be conceived as a weighted sum of interarrival times. Before demonstrating this, we introduce two abbreviations for a more concise notation.

Remark 2.7 (Abbreviated Notation for Load Sharing and Damage Accumulation Terms)**.**
For $j = 1, \dots, J$ and $i = 1, \dots, I$, let

$$B_{j,i} := s_j \frac{I}{I - (i-1)} \tag{2.22}$$

be the current stress that is due to equal load sharing before the ith component failure in the jth system. We define:

$$A_{j,i} := \sum_{k=1}^{i-1} B_{j,k} W_k^{(j)} = \sum_{k=1}^{i-1} B_{j,k}\left(T_k^{(j)} - T_{k-1}^{(j)}\right)\,, \quad \text{where } T_0^{(j)} := 0. \tag{2.23}$$

We note that $A_{j,i}$ is a random variable by definition, while $B_{j,i}$ is not. Occasionally, we will implicitly use the same notation for a realization of this random variable. For any $t \in \mathcal{I}$, we let $i = N_t^{(j)} + 1$ and observe:

$$\begin{aligned}
\tau A_j(t) &= \int_0^t s_j \frac{I}{I - N_{u^-}^{(j)}}\,\mathrm{d}u \\
&= \int_{T_{i-1}^{(j)}}^t \underbrace{s_j \frac{I}{I - N_{u^-}^{(j)}}}_{=B_j(t)}\,\mathrm{d}u + \sum_{k=1}^{i-1} \int_{T_{k-1}^{(j)}}^{T_k^{(j)}} \underbrace{s_j \frac{I}{I - N_{u^-}^{(j)}}}_{=B_{j,k}}\,\mathrm{d}u \\
&= B_j(t)\left(t - T_{i-1}^{(j)}\right) + \sum_{k=1}^{i-1} B_{j,k}\left(T_k^{(j)} - T_{k-1}^{(j)}\right)\,.
\end{aligned} \tag{2.24}$$

In particular, Equation (2.24) is also valid for $t = T^{(j)}_{i-1}$ on $\{i-1 \le C_j\} \cap \{T^{(j)}_{i-1} \le \tau_j\}$ (any $\omega \in \{i-1 > C_j\} \cup \{T^{(j)}_{i-1} > \tau_j\}$ yields $T^{(j)}_{i-1}(\omega) = \infty$). Hence,

$$\tau A_j(T^{(j)}_{i-1}) = \sum_{k=1}^{i-1} B_{j,k}\left(T^{(j)}_k - T^{(j)}_{k-1}\right) = A_{j,i}\,, \tag{2.25}$$

so that $A_{j,i}$ is the rescaled accumulated stress until $i-1$ components have failed. This justifies setting $A_{j,1} = 0$, because no stress was exerted before conducting the experiment. Much like we obtained the Basquin load sharing model without damage accumulation by applying the Basquin link g_θ to the load sharing term $B_j(t)$, plugging the damage accumulation term $A_j(t)$ into g_θ yields a simple model with damage accumulation. The resulting model, however, would not be an extension of the basic Basquin load sharing model from Definition 2.4. If we want to keep ${}^{\mathrm{B}}\lambda^{(j)}_\theta(t)$ as a special case of the augmented model, we can instead append the damage accumulation term multiplicatively. This provides the Basquin load sharing model with multiplicative damage accumulation.

Definition 2.8 (Basquin Load Sharing Model With Multiplicative Damage Accumulation; cf. Müller and Meyer 2022, p. 3)**.**
In the framework of Section 2.1 and under Assumptions 2.3, the Basquin load sharing model with multiplicative *damage accumulation is given via the intensity process*

$$\begin{aligned}{}^{\times\mathrm{D}}\lambda^{(j)}_\theta(t) &:= \theta_1\left(s_j\frac{I}{I-N^{(j)}_{t^-}}\right)^{\theta_2}\left(\frac{1}{\tau}\int_0^t s_j\frac{I}{I-N^{(j)}_{u^-}}\,\mathrm{d}u\right)^{\theta_3}\cdot \mathbb{1}_{\left\{N^{(j)}_{t^-}<C_j\right\}\cap\left\{t\le\tau_j\right\}}\\ &= \theta_1 B_j(t)^{\theta_2}A_j(t)^{\theta_3}\cdot\mathbb{1}_{\left\{N^{(j)}_{t^-}<C_j\right\}\cap\left\{t\le\tau_j\right\}}\,, \qquad \theta = (\theta_1,\theta_2,\theta_3)^\top \in \mathbb{R}^3_+\,.\end{aligned} \tag{2.26}$$

For any $\varepsilon > 0$, the parameter space Θ can be extended[1] *to subsets of* $\mathbb{R}^2_+ \times [-1+\varepsilon, \infty)$.

The schematic in Figure 3 explains the individual components of this model and serves as a diagrammatic summary of this subsection. From Equation (2.26), one obtains the intensity of the Basquin load sharing model *without* damage accumulation by setting $\theta_3 = 0$. Specifically,

$${}^{\mathrm{B}}\lambda^{(j)}_{(\theta_1,\theta_2)^\top}(t) = {}^{\times\mathrm{D}}\lambda^{(j)}_{(\theta_1,\theta_2,0)^\top}(t)\,, \quad \text{for all } \theta_1,\theta_2 \in \mathbb{R}_+. \tag{2.27}$$

Accordingly, the model with damage accumulation emerges as an actual extension of the model without damage accumulation. This proves to be useful especially when testing whether the damage accumulation effect is significant (i.e., whether $\theta_3 \neq 0$), an essential section of the upcoming statistical inference. Note that the respective model for the cumulative intensity process ${}^{\times\mathrm{D}}\Lambda_\theta$ can easily be derived in the fashion of Equation (2.3). If we consider the alternative parametrization of the Basquin link from Equation (2.18) (i.e., we use the parameter $\tilde\theta = (-\ln(\theta_1), \theta_2, \theta_3)$), this model can also be regarded as a relative risk regression model:

$${}^{\times\mathrm{D}}\lambda^{(j)}_\theta(t) = \theta_1 B_j(t)^{\theta_2}A_j(t)^{\theta_3}\cdot\mathbb{1}_{\left\{N^{(j)}_{t^-}<C_j\right\}\cap\left\{t\le\tau_j\right\}}$$

[1] A justification is provided retrospectively in Remark 2.18 in Section 2.4.

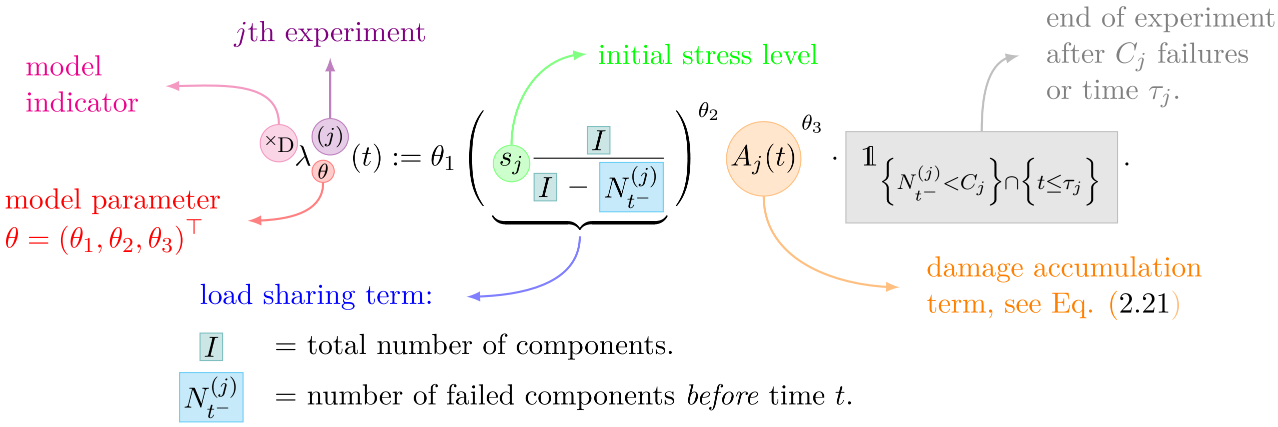


Figure 3: Schematic of the Basquin load sharing model with damage accumulation.

$$= \exp\left(-\tilde{\theta}_1 + \tilde{\theta}_2 \ln B_j(t) + \tilde{\theta}_3 \ln A_j(t)\right) \cdot \mathbb{1}_{\left\{N_{t-}^{(j)} < C_j\right\} \cap \left\{t \leq \tau_j\right\}} \tag{2.28}$$

$$= \exp\left(\tilde{\theta}^\top Z_j(t)\right) Y_j(t),$$

where

$$Z_j(t) = (-1, \ln B_j(t), \ln A_j(t))^\top \quad \text{and} \quad Y_j(t) = \mathbb{1}_{\left\{N_{t-}^{(j)} < C_j\right\} \cap \left\{t \leq \tau_j\right\}}.$$

Of course, other models incorporating damage accumulation are also conceivable. We conclude this subsection with a paragraph on further approaches pursued by Müller and Meyer 2022 and the author of this thesis.

Related Models With Damage Accumulation

The objective of extending the ordinary Basquin load sharing model with a damage accumulation term is to capture the effects of accumulating stress not only in the cumulative intensity, but the intensity itself: Where the intensity ${}^{\mathrm{B}}\lambda_{\theta}^{(j)}(t)$ of the Basquin load sharing model without damage accumulation at a given time t depends only on the total number of component failures prior to t and remains the same regardless of when these failures occured, the multiplicative damage accumulation term $A_j(t)$ accounts for earlier failures leading to a higher intensity over the course of the experiment. As a result, for $\theta_3 > 1$ we would generally expect an even greater impact of the damage accumulation on the instantaneous failure rate embodied in the intensity. But since $A_j(0) = 0$, we have ${}^{\times\mathrm{D}}\lambda_{\theta}^{(j)}(0) = 0$, whereas ${}^{\mathrm{B}}\lambda_{\theta}^{(j)}(t) = \theta_1 s_j^{\theta_2}$, so that the intensity drops relative to the base model before catching up due to the accumulation of damage. In order to circumvent the issue of $A_j(t)$ vanishing at $t = 0$ and ensure $A_j(t) \geq 1$ for all $t \in \mathcal{I}$, modifications of the damage accumulation term need to be considered. A selection of these is discussed here. If we want to force $A_j(t) \geq 1$ without fundamentally changing the inherent dynamics of the model, we can accomplish this by simply adding 1 to the non-negative damage accumulation term. This leads to a model with shifted damage accumulation.

Definition 2.9 (Basquin Load Sharing Model With Shifted Damage Accumulation)**.**
In the framework of Section 2.1 and under Assumptions 2.3, the Basquin load sharing model with shifted *damage accumulation is given via the intensity process*

$$^{\times\mathrm{S}}\lambda_\theta^{(j)}(t) := \theta_1 B_j(t)^{\theta_2}\left(1 + A_j(t)\right)^{\theta_3} \cdot \mathbb{1}_{\left\{N_{t^-}^{(j)} < C_j\right\} \cap \left\{t \le \tau_j\right\}}, \qquad \theta = (\theta_1, \theta_2, \theta_3)^\top \in \mathbb{R}_+^3 .$$

In this definition, the modified damage accumulation term is $\tilde{A}_j(t) := 1 + A_j(t)$. Because of $A_j(t) \geq 0$, other isotone transformations $\phi : [0, \infty) \to [1, \infty)$ serve a similar purpose. This leads to an entire class of intensity processes with $\tilde{A}_j(t) := \phi\left(A_j(t)\right)$, that is,

$$^{\times\phi}\lambda_\theta^{(j)}(t) := \theta_1 B_j(t)^{\theta_2}\phi\left(A_j(t)\right)^{\theta_3} \cdot \mathbb{1}_{\left\{N_{t^-}^{(j)} < C_j\right\} \cap \left\{t \le \tau_j\right\}}, \quad \text{for } \phi : [0, \infty) \to [1, \infty) \text{ isotone.}$$

However, a study of this class in its generality is beyond the scope of this thesis. Instead, we investigate the special case $\phi = \exp$, which leads to a model with exponential damage accumulation.

Definition 2.10 (Basquin Load Sharing Model With Exponential Damage Accumulation)**.**
In the framework of Section 2.1 and under Assumptions 2.3, the Basquin load sharing model with exponential *damage accumulation is given via the intensity process*

$$^{\times\mathrm{E}}\lambda_\theta^{(j)}(t) := \theta_1 B_j(t)^{\theta_2}\exp\left(\theta_3 A_j(t)\right) \cdot \mathbb{1}_{\left\{N_{t^-}^{(j)} < C_j\right\} \cap \left\{t \le \tau_j\right\}}, \qquad \theta = (\theta_1, \theta_2, \theta_3)^\top \in \mathbb{R}_+^3 .$$

All these models $^{\times\phi}\lambda_\theta^{(j)}(t)$ - which include the shifted and exponential damage accumulation - can again be represented as relative risk regression models. However, this does not apply to the subsequent model with additive damage accumulation studied in Müller and Meyer 2022. Therefore, it does not belong to the model class central to this thesis, which is why we introduce it here only for the sake of comparison.

Definition 2.11 (Basquin Load Sharing Model With Additive Damage Accumulation; cf. Müller and Meyer 2022, p. 3)**.**
In the framework of Section 2.1 and under Assumptions 2.3, the Basquin load sharing model with additive *damage accumulation is given via the intensity process*

$$^{+\mathrm{D}}\lambda_\theta^{(j)}(t) := \theta_1\left(B_j(t) + \theta_3 A_j(t)\right)^{\theta_2} \cdot \mathbb{1}_{\left\{N_{t^-}^{(j)} < C_j\right\} \cap \left\{t \le \tau_j\right\}}, \qquad \theta = (\theta_1, \theta_2, \theta_3)^\top \in \mathbb{R}_+^3 .$$

All models under consideration can be distinguished by their model indicator. These indicators are listed below:

B: Basquin load sharing model without damage accumulation,

$^\times$D: Model with basic multiplicative ($\times$) damage accumulation,

$^+$D: Model with additive ($+$) damage accumulation,

$^\times$S: Model with shifted multiplicative ($\times$) damage accumulation,

$^\times$E: Model with exponential multiplicative ($\times$) damage accumulation,

${}^{\times}\phi$: Model with multiplicative ($\times$) damage accumulation transformed by ϕ.

The main computations in this dissertation are performed on the models ${}^{\times}$D and B. Selected calculations are recapitulated in Appendix B.5 for the models ${}^{\times}$S and ${}^{\times}$E. The models ${}^{+}$D and ${}^{\times}\phi$ may be the subject of future research, with the model ${}^{\times}\phi$ in particular lending itself to further generalizations of the results presented in this thesis.

2.4. Uniform Bounds for the Intensity and its Partial Derivatives

One of the methods of statistical inference for intensity-based point process models considered in this thesis is the minimum distance estimator of Kopperschmidt and Stute. This estimator is based on minimizing a Cramér-von Mises distance between aggregate counting processes and their associated cumulative intensities. A major contribution of this work is the proof of asymptotic normality of the minimum distance estimator in Chapter 3, which corrects the original defective proof first given by Kopperschmidt 2005 and later published in Kopperschmidt and Stute 2013. In a nutshell, our proof relies on Taylor approximations of the Cramér-von Mises distance with respect to the parameter θ and therefore involves derivatives of cumulative intensities. The main preliminary work is thus to develop uniform bounds for the intensity function and its partial derivatives, especially in case of the Basquin load sharing model with damage accumulation.
We start with a basic proposition captured in the following lemma. It states that, for any fixed θ, the conditional intensity function ${}^{\times\mathrm{D}}\lambda_\theta^{(j)}(t)$ can be uniformly bounded by constraining both the load sharing term and the damage accumulation term.

Lemma 2.12 (Uniform Bounds for the Intensity in the Basquin Load Sharing Model with Damage Accumulation)**.**
For each fixed $\theta \in \Theta \cap \mathbb{R}_+^d$, the conditional intensity function ${}^{\times D}\lambda_\theta^{(j)}(t)$ of the Basquin load sharing model with multiplicative damage accumulation given in Definition 2.8 satisfies for all $t \in \mathcal{I}$:

$$\left| {}^{\times D}\lambda_\theta^{(j)}(t) \right| \le \theta_1 \left(s_j I\right)^{\theta_2+\theta_3} , \quad 1 \le j \le J . \tag{2.29}$$

Proof. Note that ${}^{\times\mathrm{D}}\lambda_\theta^{(j)}(t) = 0$ for $N_{t^-}^{(j)} \ge C_j$. By construction $C_j \le I$ holds, so only the case $N_{t^-}^{(j)} < C_j \le I$ needs to be considered. Since $N_{t^-}^{(j)} \in \mathbb{N}$, we have $N_{t^-}^{(j)} \le I-1$. Hence,

$$\frac{I}{I - N_{t^-}^{(j)}} \le \frac{I}{I-(I-1)} = I , \tag{2.30}$$

and substituting Equation (2.30) into Equation (2.26), we obtain for $t \in \mathcal{I} = [0,\tau]$:

$$\begin{aligned}
{}^{\times\mathrm{D}}\lambda_\theta^{(j)}(t) &= \theta_1 \left(s_j \underbrace{\frac{I}{I - N_{t^-}^{(j)}}}_{\le I} \right)^{\theta_2} \left(\frac{1}{\tau} \int_0^t s_j \underbrace{\frac{I}{I - N_{u^-}^{(j)}}}_{\le I} \,\mathrm{d}u \right)^{\theta_3} \cdot \underbrace{\mathbb{1}_{\left\{N_{t^-}^{(j)} < C_j\right\} \cap \left\{t \le \tau_j\right\}}}_{\le 1} \\
&\le \theta_1 \left(s_j I\right)^{\theta_2} \left(\frac{1}{\tau} \int_0^t s_j I \,\mathrm{d}u \right)^{\theta_3} \\
&= \theta_1 \left(s_j I\right)^{\theta_2} \left(\frac{t}{\tau} s_j I \right)^{\theta_3} \\
&\le \theta_1 \left(s_j I\right)^{\theta_2+\theta_3} .
\end{aligned}$$

Since ${}^{\times\mathrm{D}}\lambda_{\theta}^{(j)}(t) \geq 0$, Equation (2.29) follows. $\square$

We can then formulate a corollary of Theorem A.23 (the Doob-Meyer decomposition given in Appendix A) & Lemma 2.12 that introduces simple martingale bounds for the future minimum distance estimator. While parts (i) and (ii) contain more general statements, part (iii) is again tailored to the Basquin load sharing model with damage accumulation.

Corollary 2.13 (Simple Martingale Bounds in the Doob-Meyer Decomposition)**.** *Let $N = (N_t)_{t\in\mathcal{I}}$ be a counting process defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in\mathcal{I}}, \mathbb{P})$ and let $\Lambda = (\Lambda_t)_{t\in\mathcal{I}}$ denote its compensator according to the Doob-Meyer decomposition of Theorem A.23, where $\mathcal{I} = [0, \tau]$. Let $M = (M_t)_{t\in\mathcal{I}}$ given by $M_t = N_t - \Lambda_t$ be the associated $\{\mathcal{F}_t\}_{t\in\mathcal{I}}$-martingale. Then, the following statements hold:*

(i) If $\mathbb{E}(N_\tau) < \infty$ (i.e., $N_\tau < \infty$ holds $\mathbb{P}$-almost surely), then $\mathbb{E}(\Lambda_\tau) < \infty$. Furthermore, $\mathbb{E}|M_t| < \infty$ for all $t \in \mathcal{I}$.

(ii) If there exists a constant $C > 0$ such that $N_\tau \leq \frac{C}{2}$ and $\Lambda_\tau \leq \frac{C}{2}$ hold $\mathbb{P}$-almost surely, then, for all $t \in \mathcal{I}$, $|M_t| \leq C$ is satisfied $\mathbb{P}$-almost surely.

(iii) In the situation of Lemma 2.12, if $\Lambda = {}^{\times\mathrm{D}}\Lambda_{\theta}^{(j)}$ is the cumulative intensity of the Basquin load sharing model with damage accumulation and Θ is bounded, the constant

$$C = 2\max\left\{C_j,\ \tau \sup_{\theta\in\Theta} \theta_1 \left(s_j I\right)^{\theta_2+\theta_3}\right\} \tag{2.31}$$

can be chosen in (ii) and this constant does not depend on the value of θ.

Proof. The proof is rather basic and is presented here only for the sake of completeness.

(i) Using the martingale property, we have $\mathbb{E}(N_t) = \mathbb{E}(\Lambda_t)$ for all $t \in \mathcal{I}$, see Equation (A.15). Moreover, by applying the triangle inequality and exploiting the monotonicity of both N and Λ, we get for all $t \in \mathcal{I}$:

$$|M_t| = |N_t - \Lambda_t| \leq N_t + \Lambda_t \leq N_\tau + \Lambda_\tau\,, \tag{2.32}$$

so that $\mathbb{E}|M_t| \leq \mathbb{E}(N_\tau) + \mathbb{E}(\Lambda_\tau) < \infty$.

(ii) This statement follows immediately from Equation (2.32).

(iii) Utilizing the bounds for ${}^{\times\mathrm{D}}\lambda_{\theta}^{(j)}(t)$ provided in Lemma 2.12, we obtain

$${}^{\times\mathrm{D}}\Lambda_{\theta}^{(j)}(\tau) = \int_0^\tau {}^{\times\mathrm{D}}\lambda_{\theta}^{(j)}(u)\,\mathrm{d}u \leq \int_0^\tau \theta_1 \left(s_m I\right)^{\theta_2+\theta_3}\,\mathrm{d}u = \tau\theta_1 \left(s_j I\right)^{\theta_2+\theta_3}\,,$$

so that by applying the supremum[2] we get rid of the dependence on θ. By definition of the conditional intensity function ${}^{\times\mathrm{D}}\lambda_{\theta}^{(j)}(t)$, any jump beyond the C_jth jump of $N^{(j)}$ is prohibited and hence $N_\tau^{(j)} \leq C_j$ holds $\mathbb{P}$-almost surely. Accordingly, C given by Equation (2.31) satisfies the conditions of statement (ii).

[2]Note that while it is not necessary to take the supremum in order to achieve a constant satisfying (ii), doing so is convenient in that it allows us to derive uniform bounds with respect to both t and the generally unknown parameter θ.

□

From now on we will deal with the partial derivatives of the cumulative intensity function ${}^{\times\mathrm{D}}\Lambda_\theta^{(j)}$. To obtain uniform bounds for this function, we will instead compute bounds for the corresponding conditional intensity function ${}^{\times\mathrm{D}}\lambda_\theta^{(j)}(t)$ and integrate them with respect to t (cf. Equation (2.3)). In doing so, we will repeatedly encounter integrals of the natural powers of the logarithm ln. These integrals can be calculated by a general formula, which is the subject of the following lemma.

Lemma 2.14 (Integrals of the Natural Powers of ln)**.**
Let $t > 0$, $q > -1$ and $p \in \mathbb{N}_0$. Then,

$$\int_0^t x^q \cdot (\ln x)^p \,\mathrm{d}x = \frac{t^{1+q}}{1+q} \sum_{k=0}^{p} \left(\frac{-1}{1+q}\right)^{p-k} \frac{p!}{k!} (\ln t)^k \,. \tag{2.33}$$

Proof. The simple proof by induction can be found in Appendix B.1. □

After these preliminary considerations, we can proceed to the central lemma of this section. Unlike before, the following proofs tend to be technical and lengthy. We thus relocate them to Appendix B.1, along with more advanced findings that lack significant value for the main body of this dissertation, but may be useful for research building on this work.

Lemma 2.15 (Integrable Bounds for the Intensity Partial Derivatives in the Basquin Load Sharing Model with Multiplicative Damage Accumulation)**.**
Let ${}^{\times D}\lambda_\theta^{(j)}(t)$ be the conditional intensity function of the Basquin load sharing model with multiplicative damage accumulation given in Definition 2.8. Suppose that the sequence $(s_j)_{j\in\mathbb{N}}$ is bounded both downward by some $0 < s_{\text{low}} \le 1$ and upward by an arbitrary constant s_{upp} (e.g., if a preset assortment of initial stress levels $s_1, \ldots, s_L \ge 1$ is consecutively repeated). If we assume that $\Theta \subset \mathbb{R}_+^3$, then the following holds for all $t \in \mathcal{I}$, $\theta \in \Theta$ and $\omega \in \Omega$:

$$\left| \frac{\partial^{p+q+r}}{\partial\theta_1^r \partial\theta_2^p \partial\theta_3^q} {}^{\times D}\lambda_\theta^{(j)}(t) \right| \le \max\{1, \theta_1\} \cdot \left[\ln\left(\frac{s_j I}{s_{\text{low}}^2} \cdot \frac{\tau}{t} \right) \right]^{p+q} \cdot (s_j I)^{\theta_2+\theta_3} \,, \qquad p, q, r \in \mathbb{N}_0 \,, j \in \mathbb{N} \,. \tag{2.34}$$

If furthermore the parameter space Θ is bounded, there exists a constant C independent of $\theta \in \Theta$ and $j \in \mathbb{N}$ such that

$$\left| \frac{\partial^{p+q+r}}{\partial\theta_1^r \partial\theta_2^p \partial\theta_3^q} {}^{\times D}\lambda_\theta^{(j)}(t) \right| \le C \cdot \left[\ln\left(\frac{s_j I}{s_{\text{low}}^2} \cdot \frac{\tau}{t} \right) \right]^{p+q} , \quad p, q, r \in \mathbb{N}_0 \,, j \in \mathbb{N} \,. \tag{2.35}$$

Under these assumptions, differentiation of arbitrary order with respect to $\theta \in \Theta$ and integration with respect to $t \in \mathcal{I}$ are interchangeable, that is,

$$\frac{\mathrm{d}^p}{\mathrm{d}\theta^p} {}^{\times D}\Lambda_\theta^{(j)}(t) = \int_0^t \frac{\mathrm{d}^p}{\mathrm{d}\theta^p} {}^{\times D}\lambda_\theta^{(j)}(u)\, du \,, \quad p \in \mathbb{N} \,,$$

and we have

$$\left| \frac{\partial^{p+q+r}}{\partial\theta_1^r \partial\theta_2^p \partial\theta_3^q} {}^{\times D}\Lambda_\theta^{(j)}(t) \right| \le C\tau \sum_{k=0}^{p+q} \frac{(p+q)!}{k!} \left(\ln\left(\frac{s_j I}{s_{\text{low}}^2} \right) \right)^k , \tag{2.36}$$

where s_j can be replaced by s_{upp} whenever a uniform bound is desired.

Proof. The proof is purely technical and can be found in Appendix B.1. It makes use of the auxiliary Corollary B.1.1, an application of Lemma 2.14 also given in Appendix B.1, which is not needed in the main part of this thesis. □

Before turning to another Corollary of Lemma 2.15, we subject its premises to critical scrutiny and enlighten the particular choice of bounds for the sequence $(s_j)_{j\in\mathbb{N}}$.

Remark 2.16 (On the Prerequisites of Lemma 2.15)**.**
Let us first note that the requirement of $s_{\text{low}} \le 1$ is an artificial one: If any lower bound can be found for $(s_j)_{j\in\mathbb{N}}$, then obviously any smaller value will also provide a lower bound for the very same sequence. Without loss of generality, we could even assume that $(s_j)_{j\in\mathbb{N}}$ is bounded below by exactly 1 whenever a lower bound s_{low} exists. In order to see this, consider the following identity:

$$\begin{aligned}
&\theta_1 \left(s_j \frac{I}{I - N_{t^-}^{(j)}} \right)^{\theta_2} \left(\frac{1}{\tau} \int_0^t s_j \frac{I}{I - N_{u^-}^{(j)}} \, du \right)^{\theta_3} \\
&= \theta_1 \left(s_{\text{low}} \underbrace{\frac{s_j}{s_{\text{low}}}}_{=:\tilde{s}_j} \frac{I}{I - N_{t^-}^{(j)}} \right)^{\theta_2} \left(\frac{1}{\tau} \int_0^t s_{\text{low}} \frac{s_j}{s_{\text{low}}} \frac{I}{I - N_{u^-}^{(j)}} \, du \right)^{\theta_3} \\
&= \underbrace{\theta_1 s_{\text{low}}^{\theta_2+\theta_3}}_{=:\tilde{\theta}_1} \left(\tilde{s}_j \frac{I}{I - N_{t^-}^{(j)}} \right)^{\theta_2} \left(\frac{1}{\tau} \int_0^t \tilde{s}_j \frac{I}{I - N_{u^-}^{(j)}} \, du \right)^{\theta_3} \qquad (2.37) \\
&= \tilde{\theta}_1 \left(\tilde{s}_j \frac{I}{I - N_{t^-}^{(j)}} \right)^{\theta_2} \left(\frac{1}{\tau} \int_0^t \tilde{s}_j \frac{I}{I - N_{u^-}^{(j)}} \, du \right)^{\theta_3} ,
\end{aligned}$$

where $\tilde{s}_j$ is bounded below by 1. Accordingly, upon rescaling the initial loads s_j, only the first parameter of θ needs to be adjusted to obtain the exact same conditional intensity function. For this purpose it might be necessary to modify the parameter space Θ. However, if we suppose Θ to be bounded, so is the pth coordinate projection $\pi_p(\Theta)$ of Θ for $p = 1, 2, 3$. The same then holds true for the set

$$\left\{ \tilde{\theta}_1 = \theta_1 s_{\text{low}}^{\theta_2+\theta_3} \,:\, \theta = (\theta_1, \theta_2, \theta_3)^\top \in \Theta \right\} ,$$

since the image of a bounded set under a continuous mapping is itself bounded. Therefore, the rescaled parameter space

$$\tilde{\Theta} := \left\{ \left(\tilde{\theta}_1, \theta_2, \theta_3 \right)^\top \in \mathbb{R}^3 \,:\, \tilde{\theta}_1 = \theta_1 s_{\text{low}}^{\theta_2+\theta_3}, \theta = (\theta_1, \theta_2, \theta_3)^\top \in \Theta \right\}$$

remains bounded and hence the assumptions of Lemma 2.15 stay intact. Nonetheless, rescaling the initial stress levels is mostly not required, as the above identity can be exploited in the proof of Lemma 2.15. For details we recommend the reader to consult this exact proof in Appendix B.1.

For some applications, the differentiability (with respect to the parameter θ) of the intensity function is needed on the intersection of the plane

$$\pi_3^{-1}(\{0\}) = \{x = (x_1, x_2, x_3)^\top \in \mathbb{R}^3 \,:\, x_3 = 0\}$$

with the parameter space Θ. In particular, this requires that

$$\pi_3^{-1}(\{0\}) \cap \Theta = \{\theta = (\theta_1, \theta_2, \theta_3)^\top \in \Theta \,:\, \theta_3 = 0\}$$

lies in the interior of Θ. In this situation, $\Theta \subset \mathbb{R}_+^3$ is no longer satisfied, so that the integral bounds from Lemma 2.15 lose their validity. Nevertheless, we can easily formulate a corollary that accounts for parameter spaces extended beyond the plane $\pi_3^{-1}(\{0\})$.

Corollary 2.17 (Extension of Lemma 2.15)**.**
Let again ${}^{\times D}\lambda_\theta^{(j)}(t)$ *be the conditional intensity function of the Basquin load sharing model with multiplicative damage accumulation given in Definition 2.8 and suppose that the sequence* $(s_j)_{j\in\mathbb{N}}$ *is bounded as in Lemma 2.15. Moreover, we assume that* $\Theta \subset \mathbb{R}_+^2 \times (-1, \infty)$*. Then, for each* $\theta \in \Theta$ *with* $-1 < \theta_3 < 0$*, the following holds for all* $t \in \mathcal{I}$ *and* $\omega \in \Omega$*:*

$$\left| \frac{\partial^{p+q+r}}{\partial\theta_1^r \partial\theta_2^p \partial\theta_3^q} {}^{\times D}\lambda_\theta^{(j)}(t) \right| \leq \max\{1, \theta_1\} \cdot \left[\ln\left(\frac{s_j I}{s_{\text{low}}^2} \cdot \frac{\tau}{t} \right) \right]^{p+q} \cdot (s_j I)^{\theta_2} \left(s_{\text{low}} \frac{t}{\tau} \right)^{\theta_3} , \quad p, q, r \in \mathbb{N}_0 \,, j \in \mathbb{N} \,. \tag{2.38}$$

If further $0 < \varepsilon \leq 1$ *exists so that* $\Theta \subset \pi_3^{-1}\left([-1+\varepsilon, \infty)\right)$ *(i.e., the third parameter* θ_3 *is bounded away from* -1*) and* Θ *is bounded, then a constant* C *indepedent of* $\theta \in \Theta$ *and* $j \in \mathbb{N}$ *can be found such that, for all* $\theta \in \Theta$*,*

$$\left| \frac{\partial^{p+q+r}}{\partial\theta_1^r \partial\theta_2^p \partial\theta_3^q} {}^{\times D}\lambda_\theta^{(j)}(t) \right| \leq C \cdot \left[\ln\left(\frac{s_j I}{s_{\text{low}}^2} \cdot \frac{\tau}{t} \right) \right]^{p+q} \cdot \left(s_{\text{low}} \frac{t}{\tau} \right)^{\varepsilon - 1} , \quad p, q, r \in \mathbb{N}_0 \,, j \in \mathbb{N} \,. \tag{2.39}$$

Under these assumptions, differentiation of arbitrary order with respect to $\theta \in \Theta$ *and integration with respect to* $t \in \mathcal{I}$ *are also interchangeable. In addition, the following bound applies:*

$$\left| \frac{\partial^{p+q+r}}{\partial\theta_1^r \partial\theta_2^p \partial\theta_3^q} {}^{\times D}\Lambda_\theta^{(j)}(t) \right| \leq \frac{C \tau s_{\text{low}}^{\varepsilon-1}}{\varepsilon^{1+p+q}} \sum_{k=0}^{p+q} \frac{(p+q)!}{k!} \left(\varepsilon \ln\left(\frac{s_j I}{s_{\text{low}}^2} \right) \right)^k , \tag{2.40}$$

where replacing s_j *by* s_{upp} *yields a bound that is uniform with respect to* $j \in \mathbb{N}$*.*

Proof. The proof is reasonably concise, but relies heavily on the proof of the earlier Lemma 2.15. We therefore defer the proof again to Appendix B.1. ☐

Corollary 2.17 constitutes a starting point for further studies on the cumulative intensity of the Basquin load sharing model with multiplicative damage accumulation. Of particular emphasis are the Glivenko-Cantelli type convergence theorems that round out Appendix B.1. While Glivenko-Cantelli arguments play an important role in Kopperschmidt and Stute 2013 to proof the asymptotic normality of their minimum distance estimator, our

adapted proof no longer relies on them. As a consequence, we refrain here from discussing the further implications of Corollary 2.17 and instead point the interested reader to Appendix B.1.
Nonetheless, Corollary 2.17 still has an important application in that it allows us to extend the parameter space for the Basquin load sharing model with multiplicative damage accumulation to $\mathbb{R}_+^2 \times [-1+\varepsilon, \infty)$. We summarize this in a final remark that concludes this section.

Remark 2.18 (On Extensions of the Parameter Space for the Basquin Load Sharing Model With Multiplicative Damage Accumulation)**.**
We already noted in Definition 2.8 that for any $\varepsilon > 0$, the parameter space Θ of the Basquin load sharing model with multiplicative damage accumulation can be extended to arbitrary subsets of $\mathbb{R}_+^2 \times [-1+\varepsilon, \infty)$. In order for the cumulative intensity ${}^{\times D}\Lambda_\theta^{(j)}$ to be well-defined, the intensity process $\big({}^{\times D}\lambda_\theta^{(j)}(t)\big)_{t\in\mathcal{I}}$ must be integrable with respect to t on finite intervals, so that

$$ {}^{\times D}\Lambda_\theta^{(j)}(t) := \int_0^t {}^{\times D}\lambda_\theta^{(j)}(u)\,\mathrm{d}u < \infty\,, \quad t \in \mathcal{I}\,. $$

By setting $p = q = r = 0$, Equation (2.40) provides an upper bound for this integral,

$$ \int_0^t {}^{\times D}\lambda_\theta^{(j)}(u)\,\mathrm{d}u \leq \frac{C\tau s_{\mathrm{low}}^{\varepsilon-1}}{\varepsilon} < \infty. \tag{2.41} $$

The constant C can be determined explicitly if necessary, as shown in the proofs of Lemma 2.15 and Corollary 2.17. Note that Equation (2.41) requires $\varepsilon > 0$, so the parameter space Θ is delimited by the hyperplane $\pi_3^{-1}(\{-1\})$ (i.e., $\theta_3 \leq -1$ cannot be considered).
We also remark that for the well-definedness of the cumulative intensities, $\theta_2 < 0$ can be admitted. This would lead to the questionable effect that the individual component load *decreases* with each component failure, which directly contradicts our load sharing assumptions. Nevertheless, this extension still has its merits, because then the parameter space $\mathbb{R}^2 \times [-1+\varepsilon, \infty)$ becomes permissible in the alternative parametrization of the Basquin link, see Equation (2.18).

2.5. Hazard Transformation for Intensity-Based Load Sharing Models

Much like Section 2.4 lays the foundation for the distance-based parameter estimation in the Basquin load sharing model with multiplicative damage accumulation (see Chapter 3), the current Section 2.5 can be viewed as the basis for the depth-based hypotheses tests of Chapter 4. These robust tests generalize the classical sign test by taking into account only the signs of the observed data. Therefore, to perform depth-based statistical inference, we need a transformation of the point process data for which the probabilities of a positive sign can be quantified. The hazard transformation, which is the point process equivalent of the probability integral transformation, turns out to be a suitable choice. We obtain the required cumulative conditional hazard function by integrating the conditional hazard function of Equation (2.4),

$$ H_i(t \,|\, t_{1:(i-1)}, x) := \int_{t_{i-1}}^t h_i(u \,|\, t_{1:(i-1)}, x)\,\mathrm{d}u\,, \quad t \geq t_{i-1}\,, \tag{2.42} $$

quite analogous to how the cumulative intensities arise from the conditional intensity function by integration. Since the conditional intensity function is obtained by "concatenating" conditional hazard functions, compare Equation (2.5), a similar relationship translates to their cumulative counterparts. On the set $\{T_{i-1} \leq t \leq T_i\}$, it holds (details are provided in Summary 1 from Appendix A.2, see in particular Equation (A.34)):

$$H_i(t \mid T_{1:(i-1)}, X) = \Lambda(t) - \Lambda(T_{i-1}) \,. \tag{2.43}$$

This equation is crucial for our intensity-based modelling approach. While in most cases the true cumulative intensity process can not be determined, in the framework of an intensity-based model the general form of the cumulative intensity process is already specified by the model $\mathcal{M}$, see Equation (2.1). The implication of Equation (2.43) is that the shape of the cumulative conditional hazard function is inherent in the model specification, too. This allows us to calculate the hazard transform of a point process from only the cumulative intensities given by the model $\mathcal{M}$.

Definition 2.19 (Hazard Transform of a Point Process)**.**
In the framework of Section 2.1, the hazard transform $R^\theta = (R_i^\theta)_{i\in\mathbb{N}}$ at $\theta \in \Theta$ of the point process $T = (T_i)_{i\in\mathbb{N}}$ under the model $\mathcal{M} = \{\Lambda_\theta \,:\, \theta \in \Theta\}$ is defined as

$$R_i^\theta := \Lambda_\theta(T_i) - \Lambda_\theta(T_{i-1}) \,, \quad i \in \mathbb{N} \,,$$

where $T_0 := 0$. If $T^{(1)}, \ldots, T^{(J)}$ denote i.i.d. copies of T, then the hazard transform of $T^{(j)}$ at θ is analogously given by the process $R_j^\theta = (R_{j,i}^\theta)_{i\in\mathbb{N}}$ obtained via

$$R_{j,i}^\theta := \Lambda_\theta^{(j)}(T_i^{(j)}) - \Lambda_\theta^{(j)}(T_{i-1}^{(j)}) \,, \quad i \in \mathbb{N} \,, \quad j \in \{1, \ldots, J\} \,.$$

In Section 2.3, we occasionally stated the load sharing models in terms of parametric conditional hazard functions by virtue of Equation (2.5). The cumulative conditional hazard functions obtained by integrating these functions then satisfy

$$R_{j,i}^\theta = H_i^\theta(T_i^{(j)} \mid T_{1:(i-1)}^{(j)}, X_j) \,, \quad i \in \mathbb{N} \,, \quad j \in \{1, \ldots, J\} \,. \tag{2.44}$$

Equation (2.44) establishes the term "hazard transform". It also provides a simpler representation of the transformed process R_j^θ, which we prefer for many upcoming calculations. At the true parameter θ^*, $H_i^{\theta^*}(\,\cdot \mid t_{1:(i-1)}, x)$ is the true cumulative conditional hazard function of the ith point T_i after the observation of $T_{1:(i-1)} = t_{1:(i-1)}$ and $X = x$. This function is the same for the i.i.d. copies $T_i^{(1)}, \ldots, T_i^{(J)}$ of T_i. If the true cumulative conditional hazard functions are strictly increasing for all $i \in \mathbb{N}$, the distribution of the hazard transforms at θ^* can be determined. In the formulation of the following theorem, this requirement is expressed in terms of the intensity process λ_{θ^*}.

Theorem 2.20 (Exponential Distribution of the Hazard Transform at θ^*)**.**
In the framework of Section 2.1, let $R_{j,i}^\theta$ for $\theta \in \Theta$, $i \in \mathbb{N}$ and $j \in \{1, \ldots, J\}$ denote the hazard transforms from Definition 2.19. Suppose that the following conditions are satisfied:

(i) $\lambda_{\theta^*} > 0$ *almost everywhere $\mathbb{P}_{\theta^*}$-almost surely,*

(ii) $\int_0^\infty \lambda_{\theta^*}(u) \,\mathrm{d}u = \infty$ *$\mathbb{P}_{\theta^*}$-almost surely,*

where $\mathbb{P}_{\theta^*}$ indicates the probability measure under the true parameter θ^*. Then,

$$R_{j,i}^{\theta^*} \overset{i.i.d.}{\sim} \mathcal{E}(1)\,,$$

that is, the hazard transforms are independent and exponentially distributed with rate 1.

Proof. We give an elementary proof under slightly adjusted conditions in Theorem A.46 from Appendix A.3, where we highlight the parallels to the probability integral transform. However, the hazard transformation is usually treated in the literature as a random time change of a counting process N. Let N be the counting process associated with the point process T and let Λ denote the compensator of N (in the framework of Section 2.1, $\Lambda = \Lambda_{\theta^*}$). We define the generalized inverse of Λ by

$$\Lambda^{-1}(t) := \inf\{x \,:\, \Lambda(x) \geq t\}\,.$$

As shown in Brémaud 1981, pp. 41–42 and Daley and Vere-Jones 2008, pp. 420–421, condition (ii) then suffices to ensure that the transformed process $\tilde{N}$ defined by

$$\tilde{N}_t = N_{\Lambda^{-1}(t)} \tag{2.45}$$

is a homogeneous Poisson process with intensity 1, compare Theorem A.44 in Appendix A.3. Where the point T_i marks the time of the ith jump of N, Equation (2.45) implies that the ith jump of the transformed process $\tilde{N}$ is given at

$$\tilde{T}_i = \inf\{x \,:\, \Lambda^{-1}(x) \geq T_i\}\,.$$

Since the interarrival times of a homogeneous Poisson process with intensity λ are independent and follow an exponential distribution with parameter λ (cf. Equation (A.6)),

$$\tilde{T}_i - \tilde{T}_{i-1} \overset{\text{i.i.d.}}{\sim} \mathcal{E}(1)\,.$$

To complete the proof, we thus only need to show that $\tilde{T}_i = \Lambda(T_i)$. This is where condition (i) comes into effect: If the intensity process λ is strictly positive almost everywhere, then the cumulative intensity Λ is strictly increasing and therefore invertible. The inverse Λ^{-1} is also strictly increasing. From $\Lambda^{-1}(\Lambda(T_i)) = T_i$ it follows that

$$\{x \,:\, \Lambda^{-1}(x) \geq T_i\} = [\Lambda(T_i), \infty)\;,$$

which implies $\tilde{T}_i = \Lambda(T_i)$ and completes the proof. □

Remark 2.21 (On the Conditions of Theorem 2.20)**.**
It is shown in Lemma L17 of Brémaud 1981, p. 41 (the proof is found on p. 54 there) that condition (ii) from Theorem 2.20 is equivalent to the condition that

$$\lim_{t\to\infty} N_t = \infty \quad \mathbb{P}_{\theta^*}\text{-almost surely.}$$

If we aim to compute the hazard transform only for finitely many $i \in \{1, \dots, I\}$ with $I \in \mathbb{N}$, we can instead require that

$$\lim_{t\to\infty} N_t \geq I \quad \mathbb{P}_{\theta^*}\text{-almost surely.} \tag{2.46}$$

If equality holds in Equation (2.46), then $T_i = \infty$ for all $i > I$ and the hazard transforms for those i cannot be calculated. However,

$$R_i^{\theta^*} \overset{\text{i.i.d.}}{\sim} \mathcal{E}(1)\,, \quad i \in \{1, \dots, I\}\,,$$

still applies, as Daley and Vere-Jones 2003, p. 260 point out. This finding also has implications for condition (i) of Theorem 2.20: For (i) to hold, there has to exist a set $\Omega_0 \subset \Omega$ with $\mathbb{P}_{\theta^*}(\Omega_0) = 1$ such that for all $\omega \in \Omega_0$ the intensity function

$$t \mapsto \lambda(t, \omega)$$

attains the value 0 only on a Lebesgue null set. In other words, it has to be strictly positive almost everywhere on the sets

$$\underbrace{\{T_0(\omega) = 0 \leq t < T_1(\omega)\}}_{\text{required for } R_1^{\theta^*}}, \quad \underbrace{\{T_1(\omega) \leq t < T_2(\omega)\}}_{\text{required for } R_2^{\theta^*}}, \quad \underbrace{\{T_2(\omega) \leq t < T_3(\omega)\}}_{\text{required for } R_3^{\theta^*}}, \quad \dots \tag{2.47}$$

Since for a simple point process, $\lim_{i \to \infty} T_i = \infty$ with probability one, the sets from Equation (2.47) cover the entire positive real line $\mathbb{R}_+$, which is reflected in condition (i). To compute the first I hazard transforms of T, we only have to demand the strict positivity of λ_{θ^*} on the union of the first I of these sets, that is, almost everyone on the interval

$$\{0 \leq t < T_I(\omega)\} = [0, T_I(\omega))\ . \tag{2.48}$$

The importance of Remark 2.21 will be unveiled shortly when the hazard transformation is applied to processes that are subject to a censoring scheme - such as the Basquin load sharing models of this thesis. For now, we turn again to the distributional properties of the hazard transform:

According to Proposition A.35, the distribution of the points $T_i^{(j)}$ is completely determined by the compensator $\Lambda_{\theta^*}^{(j)}$. It is, however, not easy to calculate in general. Moreover, load sharing models with damage accumulation are specifically designed to generate dependencies between these points.

In contrast, considering the hazard transforms of the observed point processes brings - at least at the true parameter - the major advantages of independent observations following a known, simple distribution. Of course, neither the simplicity of the distribution nor the independence of the hazard transforms can be ensured for $\theta \neq \theta^*$. In this case, the hazard transformation still proves useful when the conditional cumulative hazard functions are invertible: On the one hand, this means that the transformation is reversible and thus entails no loss of information. On the other hand, we can then express the conditional (since independence may be lost) distribution of the hazard transforms in terms of the cumulative conditional hazard functions and their inverses.

In particular, by Equation (2.44) we have for $u \in \mathbb{R}$, covariates x_j and $0 < t_1^{(j)} < \ldots < t_{i-1}^{(j)}$:

$$\begin{aligned}
&\mathbb{P}_{\theta^*}\left(R_{j,i}^{\theta} > u \,\middle|\, T_{1:(i-1)}^{(j)} = t_{1:(i-1)}^{(j)}, X_j = x_j\right) \\
&\quad = \mathbb{P}_{\theta^*}\left(H_i^{\theta}\left(T_i^{(j)} \,|\, T_{1:(i-1)}^{(j)}, X_j\right) > u \,\middle|\, T_{1:(i-1)}^{(j)} = t_{1:(i-1)}^{(j)}, X_j = x_j\right)\ .
\end{aligned}$$

By application of the inverse hazard transformation at θ we obtain:

$$\begin{aligned}&\mathbb{P}_{\theta^*}\left(H_i^\theta\left(T_i^{(j)}\,|\,T_{1:(i-1)}^{(j)},X_j\right)>u\,\middle|\,T_{1:(i-1)}^{(j)}=t_{1:(i-1)}^{(j)},X_j=x_j\right)\\&\quad=\mathbb{P}_{\theta^*}\left(T_i^{(j)}>\left(H_i^\theta\right)^{-1}\left(u\,|\,T_{1:(i-1)}^{(j)},X_j\right)\,\middle|\,T_{1:(i-1)}^{(j)}=t_{1:(i-1)}^{(j)},X_j=x_j\right)\,,\end{aligned}$$

and the subsequent application of the hazard transformation at θ^* yields:

$$\begin{aligned}&\mathbb{P}_{\theta^*}\left(T_i^{(j)}>\left(H_i^\theta\right)^{-1}\left(u\,|\,T_{1:(i-1)}^{(j)},X_j\right)\,\middle|\,T_{1:(i-1)}^{(j)}=t_{1:(i-1)}^{(j)},X_j=x_j\right)\\&\quad=\mathbb{P}_{\theta^*}\left(\underbrace{H_i^{\theta^*}\left(T_i^{(j)}\,|\,T_{1:(i-1)}^{(j)},X_j\right)}_{=R_{j,i}^{\theta^*}}>H_i^{\theta^*}\left(\left(H_i^\theta\right)^{-1}\left(u\,|\,T_{1:(i-1)}^{(j)},X_j\right)\,\middle|\,T_{1:(i-1)}^{(j)},X_j\right)\,\middle|\,\cdots\right)\\&\quad=1-\mathbb{P}_{\theta^*}\left(R_{j,i}^{\theta^*}\le H_i^{\theta^*}\left(\left(H_i^\theta\right)^{-1}\left(u\,|\,T_{1:(i-1)}^{(j)},X_j\right)\,\middle|\,T_{1:(i-1)}^{(j)},X_j\right)\,\middle|\,\cdots\right)\,.\end{aligned}$$

Since the distribution of $R_{j,i}^{\theta^*}$ under $\mathbb{P}_{\theta^*}$ is known to be $\mathcal{E}(1)$, we further obtain:

$$\begin{aligned}&1-\mathbb{P}_{\theta^*}\left(R_{j,i}^{\theta^*}\le H_i^{\theta^*}\left(\left(H_i^\theta\right)^{-1}\left(u\,|\,T_{1:(i-1)}^{(j)},X_j\right)\,\middle|\,T_{1:(i-1)}^{(j)},X_j\right)\,\middle|\,\cdots\right)\\&\quad=1-F_{\mathcal{E}(1)}\left(H_i^{\theta^*}\left(\left(H_i^\theta\right)^{-1}\left(u\,|\,t_{1:(i-1)}^{(j)},x_j\right)\,\middle|\,t_{1:(i-1)}^{(j)},x_j\right)\right)\\&\quad=\exp\left(-H_i^{\theta^*}\left(\left(H_i^\theta\right)^{-1}\left(u\,|\,t_{1:(i-1)}^{(j)},x_j\right)\,\middle|\,t_{1:(i-1)}^{(j)},x_j\right)\right)\,.\end{aligned}$$

Combining these equations then provides that

$$\begin{aligned}&\mathbb{P}_{\theta^*}\left(R_{j,i}^\theta>u\,\middle|\,T_{1:(i-1)}^{(j)}=t_{1:(i-1)}^{(j)},X_j=x_j\right)\\&\quad=\exp\left(-H_i^{\theta^*}\left(\left(H_i^\theta\right)^{-1}\left(u\,|\,t_{1:(i-1)}^{(j)},x_j\right)\,\middle|\,t_{1:(i-1)}^{(j)},x_j\right)\right)\,.\end{aligned}\tag{2.49}$$

At $\theta=\theta^*$, the cumulative conditional hazard function and its inverse cancel each other out so that the right-hand side of Equation (2.49) becomes $\exp(-u)$, the survival function of the $\mathcal{E}(1)$ distribution. While $R_{j,i}^{\theta^*}$ is independent of the σ-algebra

$$\mathcal{F}_{T_{i-1}^{(j)}}=\sigma\Big(\Big\{N_{t\wedge T_{i-1}^{(j)}}^{(j)}\,:\,t\ge 0\Big\}\Big)\vee\mathcal{G}_0\,,$$

which contains information about the past of the process up to $T_{i-1}^{(j)}$ as well as the random covariates X_j, Equation (2.49) shows that this does not necessarily apply to $R_{j,i}^\theta$ if $\theta\neq\theta^*$. For the validity of the identity (2.49), it suffices that for each $j\in\{1,\ldots,J\}$ the conditional intensity functions $\lambda_\theta^{(j)}(t)$ and $\lambda_{\theta^*}^{(j)}(t)$ from the model $\mathcal{M}_j$ are non-zero almost everywhere on $\{t<T_i^{(j)}\}$, compare Remark 2.21. However, in the context of load sharing models, this conflicts with our censoring scheme that is implemented through the random covariates of Assumptions 2.3 if either

$$\mathbb{P}_{\theta^*}(C_j<i)>0\tag{2.50}$$

or

$$P_{\theta^*}\left(\tau_j < T_i^{(j)}\right) > 0 \tag{2.51}$$

applies. In both cases, the intensities (for both θ and θ^*) have a positive probability to vanish on a set with positive Lebesgue measure: In the first case, this set is given by $\{T_{C_j}^{(j)} \leq t < T_i^{(j)}\}$, while in the second case it is $\{\tau_j \leq t < T_i^{(j)}\}$. We discuss how to eliminate these complications in the following Remark 2.22.

Remark 2.22 (Compatibility of Hazard Transformation and Censoring Schemes)**.**
If either of Equations (2.50) and (2.51) holds, problems may occur in computing the hazard transform at θ^*. We discuss possible workarounds for the Basquin load sharing model with multiplicative damage accumulation, dealing with these equations one by one.

(i) **Censoring the number of observable failures leads to trivial transforms.** The covariate C_j determines how many component failures can be observed in the jth experiment. Since the intensity functions ${}^{\times\mathrm{D}}\lambda_\theta^{(j)}$ vanish on $\{t \geq T_{C_j}^{(j)}\}$ for all θ, the cumulative intensity functions ${}^{\times\mathrm{D}}\Lambda_\theta^{(j)}$ are constant on $[T_{C_j}^{(j)}, \infty)$. According to Definition 2.19, it follows that

$$R_{j,i}^{\theta} = 0 \quad \text{for all } i > C_j,$$

and hence $R_{j,i}^{\theta^*} \sim \mathcal{E}(1)$ no longer applies. However, with regard to Remark 2.21 and in particular Equation (2.46), the ith hazard transform can still be calculated if always at least i component failures can be observed, that is, $\operatorname{supp}(\mathbb{P}^{C_0}) \cap \{0, 1, \dots, i-1\} = \emptyset$ and thus (recall that C_j is an i.i.d. copy of C_0)

$$\mathbb{P}_{\theta^*}\left(C_j < i\right) = 0\,.$$

In practice, this can be easily ensured by choosing

$$C_1, \dots, C_J \overset{\text{i.i.d.}}{\sim} \delta_{I_c}\,, \quad i \leq I_c \leq I\,,$$

which means that in each experiment exactly I_c component failures can be observed and the censoring is deterministic.

(ii) **Censoring the end of the experiment causes truncated hazard functions.** The covariate τ_j marks the end of the jth experiment. Similar to case (i), the intensity functions ${}^{\times\mathrm{D}}\lambda_\theta^{(j)}$ vanish on $\{t \geq \tau_j\}$ for all θ. Put simply, this has the consequence that the support of the distribution of the hazard transform $R_{j,i}^{\theta^*}$ is bounded at random. At this bound - which depends on the covariate τ_j - the conditional distribution function of $R_{j,i}^{\theta^*}$ jumps to 1 and therefore has a discontinuity. This contradicts Theorem 2.20, see Remark A.52 in the appendix for a more elaborate analysis of censoring in the context of the related Theorem A.46.
Unlike case (i), this problem cannot be overcome by imposing restrictions on the support of τ_0. Instead, we have to require $\tau_0 \equiv \infty$. Therefore, it appears necessary to abolish type I censoring altogether. We might even be tempted to disguise type I censoring as type II censoring by replacing

$$\mathbb{1}_{\left\{N_{t-}^{(j)} < C_j\right\} \cap \left\{t \leq \tau_j\right\}}$$

with

$$\mathbb{1}_{\left\{N^{(j)}_{t^-} < \widetilde{C}_j\right\}}, \quad \text{where} \quad \widetilde{C}_j := \min\left\{C_j, N^{(j)}_{\tau_j^-}\right\}, \tag{2.52}$$

but this entails a dependence of $\widetilde{C}_j$ on the course of the experimental run. This conflicts the premise of $\mathcal{G}_0$-measurable covariates (i.e., the covariates must not depend on the process history), although the $\widetilde{C}_j$ defined in this way remain at least pairwise independent. An actual workaround lies in understanding that the conditional intensity functions with or without type I censoring are indistinguishable as long as $t \leq \tau_j$ holds and the observations are thus uncensored. In practice, this allows us to consider the hazard transforms of only those points not censored by τ_j. Although not mathematically sound, this approach is nevertheless plausible, since an observation horizon up to $\tau = \infty$ is not feasible in real-life applications.

By virtue of Remark 2.22, we need to make a few adjustments in order to provide statistical inference based on the hazard transform. To this end, we further specify the framework from Section 2.1 as well as the covariates from Assumptions 2.3.

Definition 2.23 (Framework for Hazard Transforms in Load Sharing Models).
In the framework of Section 2.1, let $\mathcal{I} = \mathbb{R}_+$. Let $1 \leq I_c \leq I$ be the critical number of component failures, which is the same for all experiments. The covariates C_j and τ_j are presumed to be deterministic with $C_j \equiv I_c$ and $\tau_j \equiv \infty$. To conform with Assumptions 2.3, we set:

$$\mathbb{P}^{C_0} = \delta_{I_c} \quad \textit{and} \quad \mathbb{P}^{\tau_0} = \delta_\infty .$$

The deterministic constant τ can be chosen arbitrarily.

In the framework of Definition 2.23, only one random covariate remains to appear in the (cumulative) conditional hazard function: the random initial stress level s_j. Now that the censoring scheme is no longer subject to randomness, these conditional hazard functions can be conveniently read from the intensity function. As a first lemma, we give the conditional hazard functions and their cumulative versions in the model $^\times\mathrm{D}$, that is, the Basquin load sharing model with multiplicative damage accumulation.

Lemma 2.24 (Conditional Hazard Functions of the Model $^\times\mathrm{D}$; cf. Theorem II.4 of Müller and Meyer 2022, p. 4).
In the framework of Definition 2.23, let $j \in \{1, \ldots, J\}$ and $\theta \in \Theta$. The conditional hazard functions defining the intensity function $^{\times\mathrm{D}}\lambda^{(j)}_\theta$ are given by

$$^{\times\mathrm{D}}h_i^\theta\Big(t \,\big|\, T^{(j)}_{1:(i-1)}, s_j\Big) = \theta_1 B_{j,i}^{\theta_2}\left[\frac{1}{\tau}\left(B_{j,i}\left(t - T^{(j)}_{i-1}\right) + A_{j,i}\right)\right]^{\theta_3}, \quad i \in \{1, \ldots, I_c\}. \tag{2.53}$$

The corresponding cumulative conditional hazard functions for $i \in \{1, \ldots, I_c\}$ are

$$^{\times\mathrm{D}}H_i^\theta\Big(t \,\big|\, T^{(j)}_{1:(i-1)}, s_j\Big) = \frac{\theta_1 B_{j,i}^{\theta_2 - 1}}{\tau^{\theta_3}(\theta_3 + 1)}\left[\left(B_{j,i}\left(t - T^{(j)}_{i-1}\right) + A_{j,i}\right)^{\theta_3+1} - A_{j,i}^{\theta_3+1}\right]. \tag{2.54}$$

In each case, we used the abbreviated notation from Remark 2.7.

Proof. Let $i \in \{1, \ldots, I_c\}$. By Equation (2.5), we have to prove that the conditional hazard function from Equation (2.53) coincides $\mathbb{P}_{\theta^*}$-almost surely with $^{\times\mathrm{D}}\lambda^{(j)}_\theta$ almost

everywhere on the interval $[T^{(j)}_{i-1}, T^{(j)}_i)$.
On this set, $N^{(j)}_{t^-} = i - 1$ holds everywhere except for $t = T^{(j)}_{i-1}$. Therefore,

$$B_j(t) = s_j \frac{I}{I - N^{(j)}_{t^-}} = s_j \frac{I}{I - (i-1)} = B_{j,i} \quad \text{almost everywhere.} \tag{2.55}$$

Moreover, Equation (2.24) shows that

$$\tau A_j(t) = B_{j,i}\Big(t - T^{(j)}_{i-1}\Big) + A_{j,i} \quad \text{almost everywhere.} \tag{2.56}$$

Substituting Equations (2.55) and (2.56) into the intensity process ${}^{\times\mathrm{D}}\lambda^{(j)}_\theta$ then yields:

$$\begin{aligned}
{}^{\times\mathrm{D}}\lambda^{(j)}_\theta(t) &= \theta_1 B_j(t)^{\theta_2} A_j(t)^{\theta_3} \cdot \mathbb{1}_{\{N^{(j)}_{t^-} < C_j\} \cap \{t \le \tau_j\}} \\
&= \theta_1 B^{\theta_2}_{j,i} \left[\frac{1}{\tau}\left(B_{j,i}\left(t - T^{(j)}_{i-1}\right) + A_{j,i}\right)\right]^{\theta_3} \qquad \mathbb{P}_{\theta^*}\text{-a.s. almost everywhere,}
\end{aligned}$$

since on $\{T^{(j)}_{i-1} \le t < T^{(j)}_i\}$, the framework of Definition 2.23 guarantees that

$$\mathbb{1}_{\{N^{(j)}_{t^-} < C_j\} \cap \{t \le \tau_j\}} = \mathbb{1}_{\{N^{(j)}_{t^-} < I_c\}} \cdot \mathbb{1}_{\{t \le \infty\}} = 1 \quad \mathbb{P}_{\theta^*}\text{-almost surely.}$$

This proves Equation (2.53). To obtain the cumulative conditional hazard function from Equation (2.54), we proceed to integrate the conditional hazard function:

$$\begin{aligned}
{}^{\times\mathrm{D}}H^\theta_i\Big(t \,|\, T^{(j)}_{1:(i-1)}, s_j\Big) &= \int_{T^{(j)}_{i-1}}^t {}^{\times\mathrm{D}}h^\theta_i\Big(u \,|\, T^{(j)}_{1:(i-1)}, s_j\Big)\, \mathrm{d}u \\
&= \int_{T^{(j)}_{i-1}}^t \theta_1 B^{\theta_2}_{j,i} \left[\frac{1}{\tau}\left(B_{j,i}\left(u - T^{(j)}_{i-1}\right) + A_{j,i}\right)\right]^{\theta_3} \mathrm{d}u \\
&= \theta_1 B^{\theta_2}_{j,i} \left(\frac{1}{\tau}\right)^{\theta_3} \int_{T^{(j)}_{i-1}}^t \left[B_{j,i}\left(u - T^{(j)}_{i-1}\right) + A_{j,i}\right]^{\theta_3} \mathrm{d}u \\
&= \theta_1 B^{\theta_2}_{j,i} \left(\frac{1}{\tau}\right)^{\theta_3} \left[\frac{1}{B_{j,i}(\theta_3 + 1)}\left(B_{j,i}\left(u - T^{(j)}_{i-1}\right) + A_{j,i}\right)^{\theta_3+1}\right]^t_{u = T^{(j)}_{i-1}} \\
&= \frac{\theta_1 B^{\theta_2 - 1}_{j,i}}{\tau^{\theta_3}(\theta_3 + 1)} \left[\left(B_{j,i}\left(t - T^{(j)}_{i-1}\right) + A_{j,i}\right)^{\theta_3+1} - A^{\theta_3+1}_{j,i}\right],
\end{aligned}$$

which verifies Equation (2.54) and completes the proof. □

As a corollary, we immediately obtain the (cumulative) conditional hazard functions of the model B, that is, the Basquin load sharing model *without* damage accumulation.

Corollary 2.25 (Conditional Hazard Functions of the Model B)**.**
In the situation of Lemma 2.24, the conditional hazard functions of the Basquin load sharing model without damage accumulation are given by

$${}^{\mathrm{B}}h^\theta_i\Big(t \,|\, T^{(j)}_{1:(i-1)}, s_j\Big) = \theta_1 B^{\theta_2}_{j,i}\,, \quad i \in \{1, \ldots, I_c\}\,.$$

The corresponding cumulative conditional hazard functions are

$$ {}^{\mathrm{B}}H_i^{\theta}\Big(t \,|\, T_{1:(i-1)}^{(j)}, s_j\Big) = \theta_1 B_{j,i}^{\theta_2}\Big(t - T_{i-1}^{(j)}\Big)\,, \quad i \in \{1, \ldots, I_c\}\,. $$

Proof. The proof follows immediately from Lemma 2.24 by setting $\theta_3 = 0$, compare Equation (2.27). □

Furthermore, Lemma 2.24 enables us to explicitly state the hazard transforms $R_{j,i}^{\theta}$ of the point process $T^{(j)}$ at any $\theta \in \Theta$.

Corollary 2.26 (Hazard Transforms in the Model $^{\times}$D)**.**
In the framework of Definition 2.23, the hazard transform $R_{j,i}^{\theta}$ of $T_i^{(j)}$ at $\theta \in \Theta$ in the Basquin load sharing model with multiplicative damage accumulation is given by

$$ R_{j,i}^{\theta} = \frac{\theta_1 B_{j,i}^{\theta_2 - 1}}{\tau^{\theta_3}(\theta_3 + 1)} \Big[A_{j,i+1}^{\theta_3+1} - A_{j,i}^{\theta_3+1}\Big]\,, \quad i \in \{1, \ldots, I_c\}, \quad j \in \{1, \ldots, J\}\,. $$

Proof. According to Equation (2.44), the hazard transform $R_{j,i}^{\theta}$ is obtained by substituting $t = T_i^{(j)}$ into the cumulative conditional hazard function of Equation (2.54). □

We have already seen above that an invertible cumulative conditional hazard function is particularly beneficial. Conveniently, we can give the inverses of the cumulative conditional hazard functions from Lemma 2.24 and Corollary 2.25 in closed form.

Lemma 2.27 (Inverse Cumulative Conditional Hazard Functions of the Models $^{\times}$D and B; cf. Theorem II.4 of Müller and Meyer 2022, p. 4)**.**
For $i \in \{1, \ldots, I_c\}$ and $j \in \{1, \ldots, J\}$, the cumulative conditional hazard function of $T_i^{(j)}$ in the Basquin load sharing models $^{\times}$D and B are invertible on $[T_{i-1}^{(j)}, \infty)$. Their inverses are given by

$$ \Big({}^{\times\mathrm{D}}H_i^{\theta}\Big)^{-1}\Big(u \,|\, T_{1:(i-1)}^{(j)}, s_j\Big) = \frac{1}{B_{j,i}}\left[\left(\frac{\tau^{\theta_3}(\theta_3+1)}{\theta_1 B_{j,i}^{\theta_2-1}} u + A_{j,i}^{\theta_3+1}\right)^{\frac{1}{\theta_3+1}} - A_{j,i} + B_{j,i} T_{i-1}^{(j)}\right]\,, $$
$$ \Big({}^{\mathrm{B}}H_j^{\theta}\Big)^{-1}\Big(u \,|\, T_{1:(i-1)}^{(j)}, s_j\Big) = \frac{u}{\theta_1 B_{j,i}^{\theta_2}} + T_{i-1}^{(j)}\,. $$

Proof. For the model $^{\times}$D, the inverse can be easily validated by verifying that

$$ \Big({}^{\times\mathrm{D}}H_i^{\theta}\Big)^{-1}\Big({}^{\times\mathrm{D}}H_i^{\theta}\Big(t \,|\, T_{1:(i-1)}^{(j)}, s_j\Big) \,\Big|\, T_{1:(i-1)}^{(j)}, s_j\Big) = t\,, \quad t \in [T_{i-1}^{(j)}, \infty)\,, $$

and

$$ {}^{\times\mathrm{D}}H_i^{\theta}\Big(\Big({}^{\times\mathrm{D}}H_i^{\theta}\Big)^{-1}\Big(u \,|\, T_{1:(i-1)}^{(j)}, s_j\Big) \,\Big|\, T_{1:(i-1)}^{(j)}, s_j\Big) = u\,, \quad u \in [0, \infty)\,. $$

The inverse for the model B is then again obtained by setting $\theta_3 = 0$ in the model $^{\times}$D. □

By combining Lemmas 2.24 and 2.27, we can compute the probability from Equation (2.49) for both the models $^{\times}$D and B. Although a simple corollary of the above results, we nevertheless formulate this identity as a theorem, because it is of paramount importance to us.

Theorem 2.28 (Cond. Distribution of the Hazard Transforms in the Models $^\times$D and B)**.**
In the framework of Definition 2.23, let $R^{\theta}_{j,i}$ be the hazard transform of $T^{(j)}_i$ at $\theta \in \Theta$ in the Basquin load sharing model with multiplicative damage accumulation. Then,

$$\begin{aligned}
&\mathbb{P}_{\theta^*}\left(R^{\theta}_{j,i} > u \,\middle|\, T^{(j)}_{1:(i-1)}, s_j\right) \\
&= \exp\left(-\frac{\theta_1^* B_{j,i}^{\theta_2^*-1}}{\tau^{\theta_3^*}\left(\theta_3^*+1\right)}\left[\left(\frac{\tau^{\theta_3}\left(\theta_3+1\right)}{\theta_1 B_{j,i}^{\theta_2-1}}u + A_{j,i}^{\theta_3+1}\right)^{\frac{\theta_3^*+1}{\theta_3+1}} - A_{j,i}^{\theta_3^*+1}\right]\right). \qquad (2.57)
\end{aligned}$$

The conditional probability in the Basquin load sharing model without damage accumulation is obtained by setting $\theta_3 = \theta_3^* = 0$ in Equation (2.57)*, so that*

$$\mathbb{P}_{\theta^*}\left(R^{\theta}_{j,i} > u \,\middle|\, T^{(j)}_{1:(i-1)}, s_j\right) = \exp\left(-\frac{\theta_1^* B_{j,i}^{\theta_2^*}}{\theta_1 B_{j,i}^{\theta_2}}u\right). \qquad (2.58)$$

Proof. By Equation (2.49), plugging the inverse cumulative conditional hazard function from Lemma 2.27 at θ into the cumulative conditional hazard function from Lemma 2.24 at θ^* yields the desired result. Notice that upon substituting the inverse cumulative conditional hazard into Equation (2.54), the terms $T^{(j)}_{i-1}$, $B_{j,i}$ and $A_{j,i}$ in the innermost bracket immediately cancel out regardless of the parameters. Since then the remainder

$$\frac{\tau^{\theta_3}\left(\theta_3+1\right)}{\theta_1 B_{j,i}^{\theta_2-1}}u + A_{j,i}^{\theta_3+1}$$

is first raised to the power of $\frac{1}{\theta_3+1}$ and then to the power of θ_3^*+1, a further simplification is only possible in the case $\theta_3 = \theta_3^*$. In that situation, the damage accumulation term $A_{j,i}$ then disappears entirely, and the conditional probability from Equation (2.57) simplifies to Equation (2.58). Obviously, this includes the case $\theta_3 = \theta_3^* = 0$. □

We recall that the median of the $\mathcal{E}(\lambda)$ distribution is known to be $\frac{\ln(2)}{\lambda}$. Therefore, we have $\mathrm{med}_{\theta^*}(R^{\theta^*}_{j,i}) = \ln(2)$ according to Theorem 2.20. A standardized version of the hazard transform can then be introduced by subtracting $\ln(2)$. This accomplishes that the hazard transforms at the true parameter θ^* are median-centred.

Definition 2.29 (Standardized Hazard Transform)**.**
The standardized hazard transform is the process $\tilde{R}^{\theta}_j$ defined by

$$\tilde{R}^{\theta}_{j,i} := R^{\theta}_{j,i} - \ln(2)\,, \quad i \in \{1, \dots, \mathbb{N}\}\,, \quad j \in \{1, \dots, J\}\,, \qquad (2.59)$$

where R^{θ}_j is the hazard transform of the point process $T^{(j)}$.

At the beginning of this section, we established that we need a transformation of the point process for which the probabilities of a positive sign can be quantified. For the standardized hazard transform at θ^*, we have

$$\mathbb{P}_{\theta^*}\left(\tilde{R}^{\theta^*}_{j,i} > 0\right) = \frac{1}{2} = \mathbb{P}_{\theta^*}\left(\tilde{R}^{\theta^*}_{j,i} < 0\right), \qquad (2.60)$$

so that the probability for both a positive and a negative sign is $\frac{1}{2}$. For $\theta \neq \theta^*$ the conditional probability of a positive sign can be specified via Equation (2.57) by setting $u = \ln(2)$:

$$\begin{aligned}\mathbb{P}_{\theta^*}\left(\tilde{R}^{\theta}_{j,i} > 0 \,\middle|\, T^{(j)}_{1:(i-1)}, s_j\right) &= \mathbb{P}_{\theta^*}\left(R^{\theta}_{j,i} > \ln(2) \,\middle|\, T^{(j)}_{1:(i-1)}, s_j\right) \\ &= \exp\left(-\frac{\theta_1^* B_{j,i}^{\theta_2^*-1}}{\tau^{\theta_3^*}\left(\theta_3^*+1\right)}\left[\left(\frac{\tau^{\theta_3}\left(\theta_3+1\right)}{\theta_1 B_{j,i}^{\theta_2-1}}\ln(2) + A_{j,i}^{\theta_3+1}\right)^{\frac{\theta_3^*+1}{\theta_3+1}} - A_{j,i}^{\theta_3^*+1}\right]\right) . \end{aligned} \tag{2.61}$$

Throughout this thesis, we assume that the initial stress level s_j is "known" in advance: this is incorporated in the assumption that s_j be $\mathcal{G}_0$-measurable. In later considerations, we will suppose that s_j is indeed deterministic. Under these circumstances, there are no longer any random covariates. We can hence consider the natural filtration $\mathcal{F}^N$ of the counting process N in place of an intrinsic filtration. Only the previous points $T^{(j)}_{1:(i-1)}$ then appear in the conditions of this section. This also concerns Equation (2.61), from which make a key observation:
Unlike the time $T^{(j)}_i$ of the ith event itself, the probability of a positive standardized hazard transform depends only on the damage accumulation term $A_{j,i}$ and not directly on the past $T^{(j)}_{1:(i-1)}$. We record this conclusion in a remark, where we will conceive of the probability from Equation (2.61) as a $\sigma(A_{j,i})$-measurable random variable.

Remark 2.30 (Relating Damage Accumulation to Standardized Hazard Transforms)**.**
As soon as s_j is considered deterministic, the load sharing term $B_{j,i}$ is also no longer subject to randomness. We can thus restate Equation (2.61) as

$$\mathbb{P}_{\theta^*}\left(\tilde{R}^{\theta}_{j,i} > 0 \,\middle|\, T^{(j)}_{1:(i-1)}\right) = g_{j,i}\left(\theta^*, \theta, A_{j,i}\right) ,$$

where $g_{j,i} : \Theta \times \Theta \times \mathbb{R} \to \mathbb{R}$ is the deterministic function defined by

$$g_{j,i}(\theta, \theta^*, x) := \exp\left(-\frac{\theta_1^* B_{j,i}^{\theta_2^*-1}}{\tau^{\theta_3^*}\left(\theta_3^*+1\right)}\left[\left(\frac{\tau^{\theta_3}\left(\theta_3+1\right)}{\theta_1 B_{j,i}^{\theta_2-1}}\ln(2) + x^{\theta_3+1}\right)^{\frac{\theta_3^*+1}{\theta_3+1}} - x^{\theta_3^*+1}\right]\right) .$$

For each fixed $\theta, \theta^* \in \Theta$, the function $x \mapsto g_{j,i}(\theta, \theta^*, x)$ is continuous. Therefore, the composition $g_{j,i}(\theta, \theta^*, A_{j,i})$ is $\sigma\left(A_{j,i}\right)$-$\mathcal{B}\left(\mathbb{R}\right)$-measurable.
Because $A_{j,i}$ is a weighted sum of the points $T^{(j)}_{1:(i-1)}$ according to Equation (2.23), $\sigma\left(A_{j,i}\right) \subset \sigma(T^{(j)}_{1:(i-1)})$ applies. We can then compute by virtue of the tower property:

$$\begin{aligned}\mathbb{P}_{\theta^*}\left(\tilde{R}^{\theta}_{j,i} > 0 \,\middle|\, A_{j,i}\right) &= \mathbb{E}_{\theta^*}\left(\mathbb{1}_{\left\{\tilde{R}^{\theta}_{j,i} > 0\right\}} \,\middle|\, A_{j,i}\right) = \mathbb{E}_{\theta^*}\left(\mathbb{E}_{\theta^*}\left(\mathbb{1}_{\left\{\tilde{R}^{\theta}_{j,i} > 0\right\}} \,\middle|\, T^{(j)}_{1:(i-1)}\right) \,\middle|\, A_{j,i}\right) \\ &= \mathbb{E}_{\theta^*}\left(\mathbb{P}_{\theta^*}\left(\tilde{R}^{\theta}_{j,i} > 0 \,\middle|\, T^{(j)}_{1:(i-1)}\right) \,\middle|\, A_{j,i}\right) \\ &= \mathbb{E}_{\theta^*}\left(g_{j,i}\left(\theta^*, \theta, A_{j,i}\right) \,\middle|\, A_{j,i}\right) \\ &= g_{j,i}\left(\theta^*, \theta, A_{j,i}\right) = \mathbb{P}_{\theta^*}\left(\tilde{R}^{\theta}_{j,i} > 0 \,\middle|\, T^{(j)}_{1:(i-1)}\right) .\end{aligned}$$

Hence, conditioning on $A_{j,i}$ instead of the entire past $T^{(j)}_{1:(i-1)}$ leads to the same results in terms of the probability to obtain a positive standardized hazard transform.

With Theorem 2.28, we have laid the foundation for statistical methods based on the hazard transformation. Many possible paths branch off from here. For example, the hypothesis $\mathcal{H}_0 : \theta^* = \theta$ can be tested by assessing whether the hazard transforms at θ follow an exponential distribution with parameter 1 (e.g., by the Kolmogorov-Smirnov test). Our goal is not to find the best solution to this testing problem. In Chapter 4, we propose a new, robust approach that demonstrates the applicability of a recent test based on the K-sign depth. Meanwhile, benchmarking with a wider range of methods is considered a task for future research.

Of course, the idea to provide statistical inference by using the hazard transform of a point process is by no means new. For instance, Section 7.4. of Daley and Vere-Jones 2003, pp. 257–267 is devoted to a goodness-of-fit test for point process models with known conditional intensity function. Since a comparison with established methods is postponed, they are only treated superficially within this thesis.

3. Minimum Distance Estimation for Parametric Intensity-Based Models

3.1. Introduction to Self-Exciting Processes and the General Framework

Our first method of inference is set in the context of "self-exciting" point processes: the minimum distance estimator of Kopperschmidt and Stute 2013[3]. The notion "self-exciting" was historically coined by Hawkes (see Hawkes 1971a and Hawkes 1971b) and intended as an extension to the doubly stochastic process (i.e., a counting process N that is conditionally a Poisson process given the random intensity function λ, see Bartlett 1963, p. 269; Chapter 7 of Snyder and Miller 1991 is dedicated to this topic). While for such doubly stochastic processes, the intensity $\lambda(t)$ is determined for all t before the associated counting process N_t is considered (Hawkes 1971a, p. 84), Hawkes allowed that the intensity depends on the past of the process itself and thus "self-excites". Originally, the term was used only for intensity processes of a specific shape - the corresponding processes are nowadays referred to as Hawkes processes (e.g., in Daley and Vere-Jones 2003, pp. 183–185) - but the modern interpretation comprises all adapted counting processes (see Brémaud 1972, p. 46 or Andersen et al. 1993, p. 73). Conveniently, self-exciting counting processes are thus precisely the class of counting processes to which the Doob-Meyer decomposition can be applied and the general self-exciting process can be conceived of as a modified Poisson process in which the intensity is not only a function of time but also the entire past of the counting process (cf. Snyder and Miller 1991, p. 287). Adaptedness (or in other words, "self-excitation") constitutes the essential prerequisite in the study of the minimum distance estimator and is always assumed throughout this thesis. Nevertheless, our framework shows several distinctions from the setting in which the results of Kopperschmidt and Stute 2013 are situated. Although these differences appear to restrict the applicability of this approach, they are in fact imperative to allow comparison with the methods of the following chapters. We point out the main dissimilarities in a first remark.

Remark 3.1 (Differences to the Framework of Kopperschmidt and Stute 2013)**.**

(i) For a stochastic process N to be adapted with respect to a filtration $\{\mathcal{F}_t\}_{t\geq 0}$, it must contain the natural filtration of N, that is:

$$\mathcal{F}_t \supset \mathcal{F}_t^N \,, \quad \text{for all } t \geq 0 \,,$$

see Remark A.12(iv). This inclusion need not be strict, and we primarily settle for equality (i.e., the internal history of N) or at most an intrinsic filtration so that all external information is available in advance. In contrast, Kopperschmidt and Stute 2013 demand that most cases of interest include external random components:

> "One needs to accept intensity processes that combine the internal history of the process with external shocks or impulses to the effect that the model is no longer dominated and straightforward likelihood methods don't exist; allow for jumps [...]; have relevant filtration strictly larger than the internal history of the process." (Kopperschmidt and Stute 2013, p. 1277)

[3]The cited article is based on the dissertation by Kopperschmidt. We refer to Kopperschmidt 2005 whenever we comment on technical details omitted from the concise article.

The disparity in perspective can be attributed to the envisaged field of application: Kopperschmidt and Stute 2013 apply their findings to market research. Specifically, a customer's purchase behavior is represented by a point process, with each observation corresponding to the time of a purchase. Promotional activities such as TV-advertising then influence the customer's behavior by "hopefully creating an impulse leading to an upward jump in the intensity process" (Kopperschmidt and Stute 2013, p. 1276), but are not part of the customer's internal (purchase) history, hence $\mathcal{F}_t \supsetneq \mathcal{F}_t^N$. Our primary interest, on the other hand, lies in experiments (mostly, but not exclusively, from the field of civil engineering) conducted under laboratory conditions. Therefore, we do not want the model to encompass effects that are not part of the intrinsic behavior. Note, though, that even if we are mostly content with the internal history, we still allow for the case of strict inclusion and will not impose any formal restrictions. In this way, unforeseen events can be incorporated into the modelling of an experiment, which in turn enables numerous practical applications. Finally, we technically do not require the model to be dominated either, but we can resort to this assumption in order to compare the minimum distance estimator with straightforward likelihood methods.

(ii) To prove the asymptotic normality of the minimum distance estimator, moment conditions are imposed on the model, see Kopperschmidt 2005, pp. 27–28. While we retain most other requirements, we strengthen this assumption by asking that the model is almost surely uniformly bounded. The reason for this is that the proof presented in Kopperschmidt 2005 contains a fundamental flaw, which we correct by following an alternative (but related) approach. The original proof exploits the posed moment conditions to derive the tightness of the involved processes using a Kolmogorov tightness criterion, whereas we show convergence directly by dominating negligible parts. Consequently, it is understandable that this specific requirement has to be adjusted. In practice, this adjustment is usually unproblematic: According to Remark A.6(iii), in any application modelled by a counting process, the number of observable events in a compact interval is finite. Both Kopperschmidt and Stute 2013 and we assume $\mathcal{I} \subset \mathbb{R}$ to be a fixed compact interval, so increasing information can only stem from sampling i.i.d. copies of a counting process[4]. In simple terms, our condition is then satisfied if that counting process is almost surely bounded by some constant. This assumption is easily justified even in the example given by Kopperschmidt and Stute 2013 themselves, in which a household's purchases of packaged ice cream were considered: Surely the number of purchases is bounded by the total quantity of units produced.

In Remark 3.1(ii) we stated that the proof provided by Kopperschmidt 2005 is defective, and the same then is true for the one published in Kopperschmidt and Stute 2013. The problem lies in a faulty generalization of Kolmogorov's tightness criterion to higher dimensional spaces. We address the errors we detected in Remark 3.27 following our proof. In doing so, we demonstrate that the proof given by Kopperschmidt 2005 is false, but we cannot disprove the correctness of the statement itself. Thus, the original proof remains intact if a proof of the generalized tightness criterion can be found. We elaborate on the common reception of Kolmogorov's and related tightness criteria in the literature. Finally,

[4]Of course, it is also possible to increase the length of $\mathcal{I}$ to infinity in order to gain information about a process observed over $\mathcal{I}$. However, this is in contradiction with our situation where conducted experiments cannot be extended arbitrarily.

we give a heuristic rationale for why the generalization might be incorrect altogether. Having made these preliminary remarks, we turn to the specification of the framework. We recollect Equations (2.1) through (2.3) of Section 2.1 to remind us of the main notations.

Definition 3.2 (Framework for Minimum Distance Estimation; adapted from Kopperschmidt 2005, pp. 21–26 and Kopperschmidt and Stute 2013, pp. 1278–1279)**.**
Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in\mathcal{I}}, \mathbb{P})$ be a filtered probability space, where $\mathcal{I} = [0, \tau]$ is a compact interval with $\tau \in (0, \infty)$. Let $N^{(1)}, \dots, N^{(n)}$, $n \in \mathbb{N}$, denote i.i.d. copies of an adapted counting process $N = (N_t)_{t\in\mathcal{I}}$ with absolutely continuous $(\mathbb{P}, \{\mathcal{F}_t\}_{t\in\mathcal{I}})$-compensator Λ. Let $\theta \in \Theta$ denote the parameter of interest, where $\Theta \subset \mathbb{R}^d$, $d \in \mathbb{N}$, is a bounded open set, so the closure $\overline{\Theta}$ of Θ is compact in $\mathbb{R}^d$. A parametric model is given by a class $\mathcal{M}$ of cumulative intensities, that is,

$$\mathcal{M} = \{\Lambda_\theta \,:\, \theta \in \Theta\} \,.$$

Let μ be a finite measure on the Borel σ-algebra $\mathcal{B}(\mathcal{I})$. For $f, g \in L^2\left(\mathcal{I}, \mathcal{B}(\mathcal{I}), \mu\right)$, we set

$$\langle f, g\rangle_\mu := \int_{\mathcal{I}} f g \,\mathrm{d}\mu$$

with corresponding semi-norm

$$\|f\|_\mu := \sqrt{\langle f, f\rangle_\mu} = \left[\int_{\mathcal{I}} f^2 \,\mathrm{d}\mu\right]^{\frac{1}{2}} \,. \tag{3.1}$$

If $X = (X_t)_{t\in\mathcal{I}}$ is a non-decreasing process with $X_0 = 0$, then for each $\omega \in \Omega$ a Borel measure on $\mathcal{I}$ is determined by

$$\mu_X\left([0, t]\right) := X_t(\omega)\,, \quad t \in \mathcal{I}\,, \tag{3.2}$$

where the dependence on ω is usually suppressed (i.e., we write μ_X instead of $\mu_{X(\omega)}$). For X we primarily consider the processes N and Λ_θ as well as their aggregate versions obtained from

$$\overline{N}^{(n)} := \frac{1}{n}\sum_{j=1}^{n} N^{(j)}\,, \quad \overline{\Lambda}_\theta^{(n)} := \frac{1}{n}\sum_{j=1}^{n} \Lambda_\theta^{(j)}\,, \quad \textit{for } \theta \in \Theta\,.$$

Let again δ_t be the Dirac measure centred on $t \in \mathcal{I}$ and

$$\mathcal{J}_N := \left\{t \,:\, N_t - \lim_{s\uparrow t} N_s \geq 1\right\}$$

denote the set of time points belonging to the jumps of the counting process N. Then, the induced measure μ_N can be expressed explicitly as

$$\mu_N = \sum_{t\in\mathcal{J}_N} \left(N_t - \lim_{s\uparrow t} N_s\right) \delta_t\,.$$

Similarly, we obtain

$$\mu_{\overline{N}^{(n)}} = \frac{1}{n}\sum_{j=1}^{n} \mu_{N^{(j)}} = \sum_{t\in\mathcal{J}_{\overline{N}^{(n)}}} \left(\overline{N}_t^{(n)} - \lim_{s\uparrow t} \overline{N}_s^{(n)}\right) \delta_t\,, \quad \textit{where } \mathcal{J}_{\overline{N}^{(n)}} := \bigcup_{j=1}^{n} \mathcal{J}_{N^{(j)}}\,.$$

In order to simplify the notation, we write $\|\cdot\|_N$ instead of $\|\cdot\|_{\mu_N}$ (and likewise for other induced measures).

The most noticeable difference from the notation in Section 2.1 is that here we use n instead of J for the number of copies of N. This is intentional to stay as close as possible to the notation of Kopperschmidt and Stute 2013.
Before we turn to further assumptions regarding the parametric model $\mathcal{M}$, we would like to point out that due to the randomness of the measures considered (e.g., μ_N), the semi-norms defined in Equation (3.1) may also be probabilistic, even if the integrand f is deterministic. At times, we will use the measure induced by the expected compensator instead, leading to deterministic semi-norms.

Assumption 3.3 (General Requirements for the Parametric Model)**.**
We make several assumptions on the model $\mathcal{M}$ and number them consecutively. For the initial requirements stated here, the letter M is used. Further assumptions are made, but these are tied to the specific results and are therefore deferred for the time being. We will later refer to them by the letters C (for consistency) and A (for asymptotic normality).

(M1) The model includes the compensator Λ. Hence, there is a true parameter $\theta^* \in \Theta$, such that

$$\Lambda = \Lambda_{\theta^*} \, .$$

(M2) Any cumulative intensity contained in $\mathcal{M}$ is $\mathbb{P}_{\theta^*}$-almost surely absolutely continuous: For each $\theta \in \Theta$, there exists a Lebesgue densities λ_θ satisfying

$$\Lambda_\theta(t) = \int_0^t \lambda_\theta(u)\,\mathrm{d}u\,, \quad t \in \mathcal{I}\,.$$

Without loss of generality, we can assume λ_θ to be left-continuous. In the case $\theta = \theta^*$, λ_θ is the $(\mathbb{P}, \{\mathcal{F}_t\}_{t\in\mathcal{I}})$-intensity of N.

It follows from (M1) that, for each $j \in \{1, \dots, n\}$, the innovation martingale $M^{(j)}$ from the Doob-Meyer decomposition of $N^{(j)}$ is given by

$$M^{(j)} = N^{(j)} - \Lambda_{\theta^*}^{(j)} \, .$$

Analogously, the innovation martingale of the aggregate process $\overline{N}^{(n)}$ is obtained:

$$\overline{M}^{(n)} = \overline{N}^{(n)} - \overline{\Lambda}_{\theta^*}^{(n)} \, . \tag{3.3}$$

We have now assembled the framework that enables us to define the minimum distance estimator. For this purpose, we consider the Cramér-von Mises distance (cf. Stute 1986, p. 234) between the aggregated counting process $\overline{N}^{(n)}$ and the aggregated cumulative intensity $\overline{\Lambda}_\theta^{(n)}$, that is:

$$\left\|\overline{N}^{(n)} - \overline{\Lambda}_\theta^{(n)}\right\|_{\overline{N}^{(n)}} = \left[\sum_{t\in\mathcal{J}_{\overline{N}^{(n)}}} \left(\overline{N}_t^{(n)} - \overline{\Lambda}_\theta^{(n)}(t)\right)^2 \left(\overline{N}_t^{(n)} - \lim_{s\uparrow t} \overline{N}_s^{(n)}\right)\right]^{\frac{1}{2}} \, . \tag{3.4}$$

This quantity accumulates the squared distances between the aggregated processes at the jump points of $\overline{N}^{(n)}$ (weighted by the jump size), and hence represents an overall measure of fit of $\overline{\Lambda}_\theta^{(n)}$ to $\overline{N}^{(n)}$ (Kopperschmidt and Stute 2013, p. 1279). As the minimum distance estimator, we then choose the parameter $\theta \in \Theta$ that minimizes the above sum.

Definition 3.4 (Minimum Distance Estimator; Kopperschmidt and Stute 2013, p. 1279)**.** *The minimum distance estimator (MDE) $\hat{\theta}_n$ for θ^* is defined as the element of the parameter space Θ that minimizes the Cramér-von Mises distance between the aggregated counting process $\overline{N}^{(n)}$ and the aggregated cumulative intensity $\overline{\Lambda}_\theta^{(n)}$, that is:*

$$\hat{\theta}_n := \arg\inf_{\theta\in\Theta} \left\| \overline{N}^{(n)} - \overline{\Lambda}_\theta^{(n)} \right\|_{\overline{N}^{(n)}} . \tag{3.5}$$

The remainder of this chapter is devoted to two central properties of this estimator: In Section 3.2 we discuss the (strong) consistency of the estimator, while Section 3.3 revolves around its asymptotic normality. In concluding this introduction, we remark that throughout Chapter 3 we will work with probabilities and expectations that formally depend on the true parameter θ^*. Nevertheless, we will retain the shorter notations $\mathbb{P}(\cdot)$ and $\mathbb{E}(\cdot)$ as opposed to $\mathbb{P}_{\theta^*}(\cdot)$ and $\mathbb{E}_{\theta^*}(\cdot)$.

3.2. Strong Consistency of the Minimum Distance Estimator

We start the section with additional assumptions about the parametric model $\mathcal{M}$ which complement the general requirements (M1) and (M2) that continue to apply.

Assumption 3.5 (Requirements Related to the Strong Consistency of the MDE)**.** Throughout this section, we assume that the following assertions are valid.

(C1) The true parameter θ^* is identifiable: For each $\varepsilon > 0$,

$$\inf_{\theta\,:\,\|\theta-\theta^*\|\geq\varepsilon} \|\mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_\theta\|_{\mathbb{E}\Lambda_{\theta^*}} > 0\,.$$

(C2) With probability one, the mapping

$$\begin{aligned} \Lambda_\cdot(\cdot)\,:\,\mathcal{I}\times\Theta &\longrightarrow \mathbb{R} \\ (t,\theta) &\longmapsto \Lambda_\theta(t) \end{aligned}$$

is continuous on $\mathcal{I}\times\Theta$ and has a continuous extension to the closure $\mathcal{I}\times\overline{\Theta}$ of $\mathcal{I}\times\Theta$. Hence,

$$\Lambda_\cdot(\cdot) \in C^0(\mathcal{I}\times\overline{\Theta}) \quad \mathbb{P}\text{-almost surely.}$$

(C3) There exists a constant $C > 0$, such that $\mathbb{P}$-almost surely holds:

$$N_t \leq C\,, \quad \Lambda_\theta(t) \leq C\,, \quad \text{for all } t\in\mathcal{I}\,, \theta\in\Theta\,.$$

In particular, the model is $\mathbb{P}$-almost surely uniformly bounded. Because of monotonicity, we can equivalently demand boundedness on the right boundary of $\mathcal{I} = [0,\tau]$,

$$N_\tau \leq C\,, \quad \Lambda_\theta(\tau) \leq C\,, \quad \text{for all } \theta\in\Theta\,.$$

In contrast to the uniform boundedness condition (C3), Kopperschmidt 2005 only demands locally uniform integrability of the model. We quote this deviant condition for the sake of completeness; it can be found in Kopperschmidt 2005, pp. 23–24, Equations (3.2.1) and (3.2.6).

($\widetilde{\mathrm{C}}3$) For $r > 0$ and $\theta \in \Theta$, let

$$\mathrm{B}_r(\theta) := \left\{\theta' \in \overline{\Theta} \,:\, \|\theta' - \theta\| < r\right\} \tag{3.6}$$

be the intersection of the ball with radius r around θ with the closure $\overline{\Theta}$ of the parameter space. He assumes that for each $\theta \in \overline{\Theta}$, there exists a positive radius $r_\theta > 0$ such that

$$\mathbb{E}\left(\sup_{\theta' \in \mathrm{B}_{r_\theta}(\theta)} \Lambda_{\theta'}(\tau)^3\right) < \infty\,. \tag{3.7}$$

Additionally, he requires that the third moments of N exist, that is

$$\mathbb{E}\left(N_\tau^3\right) < \infty\,. \tag{3.8}$$

Trivially, (C3) $\Rightarrow$ ($\widetilde{\mathrm{C}}3$), since both expectations are bounded by $C < \infty$. Considering that the balls $\mathrm{B}_{r_\theta}(\theta)$ provide an open cover of the compact set $\overline{\Theta}$, Equation (3.7) remains valid even if we take the supremum over the entire set $\overline{\Theta}$.

Remark 3.6 (Continuity of the Expected Cumulative Intensity)**.**
Combining Conditions (C2) (for the continuity of the cumulative intensity) and (C3) (to obtain an integrable majorant) yields the continuity of the expected cumulative intensity

$$(t, \theta) \mapsto \mathbb{E}\Lambda_\theta(t)$$

as an immediate consequence of the dominated convergence theorem:

$$\lim_{(t,\theta)\to(t',\theta')} \mathbb{E}\Lambda_\theta(t) = \mathbb{E}\left[\lim_{(t,\theta)\to(t',\theta')} \Lambda_\theta(t)\right] = \mathbb{E}\Lambda_{\theta'}(t') \quad \text{ for all } (t',\theta') \in \mathcal{I} \times \overline{\Theta}\,.$$

In proving the consistency of the MDE, we will find that the Cramér-von Mises distance $\left\|\overline{N}^{(n)} - \overline{\Lambda}_\theta^{(n)}\right\|_{\overline{N}^{(n)}}$ converges almost surely to $\|\mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_\theta\|_{\mathbb{E}\Lambda_{\theta^*}}$ as $n \to \infty$. Thus, in order for the MDE $\hat{\theta}_n$ to estimate the true parameter θ^*, different parameter values must be reflected in the quantity $\|\mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_\theta\|_{\mathbb{E}\Lambda_{\theta^*}}$. This justifies the identifiability condition from (C1) (cf. Kopperschmidt 2005, p. 26). Due to the continuity of the expected cumulative intensity demonstrated above, (C1) simplifies considerably: The compactness of $\overline{\Theta} \setminus \mathrm{B}_\varepsilon(\theta^*)$ implies that the infimum is attained, so that the identifiability condition becomes equivalent to

$$\forall\,\theta \neq \theta^*\ \exists\,\mathcal{I}_\theta \subset \mathcal{I} \text{ with } \mu_{\mathbb{E}\Lambda_{\theta^*}}(\mathcal{I}_\theta) > 0\ \forall\,t \in \mathcal{I}_\theta \,:\quad \mathbb{E}\Lambda_{\theta^*}(t) \neq \mathbb{E}\Lambda_\theta(t)\,. \tag{3.9}$$

In other words, this basically means that the expected cumulative intensities differ on some subset of $\mathcal{I}$ where $\mathbb{E}\Lambda_{\theta^*}$ is not constant.

The main theorem of this section states the strong consistency of the MDE.

Theorem 3.7 (Strong Consistency of the Minimum Distance Estimator; Kopperschmidt and Stute 2013, p. 1279)**.**
With the notations introduced in Definition 3.2 and under the assumptions (M1), (M2), (C1), (C2) *and* (C3) *holds:*

$$\hat{\theta}_n \longrightarrow \theta^* \quad (n \to \infty) \qquad \mathbb{P}\text{-}\textit{almost surely.}$$

Here the condition (C3) *can be replaced by* ($\widetilde{\mathrm{C}}3$)*.*

Proof. The proof is - besides the asymptotic normality of the MDE - the main statistical subject of Kopperschmidt's dissertation. The entire Chapter 7, Kopperschmidt 2005, pp. 67–80, is dedicated to this topic. An abridged version can be found in the article Kopperschmidt and Stute 2013, pp. 1284–1288. In addition, the author has spend considerable effort in his master's thesis to flesh out the proof and make it understandable even from an undergraduate's point of view, see Jakubzik 2017, pp. 83–99. A revision adapted to this work is given in the Appendix B.3 of this dissertation. □

3.3. Asymptotic Normality of the Minimum Distance Estimator

Analogous to the previous section, we specify further assumptions on the parametric model $\mathcal{M}$ that complement the ongoing requirements (M1), (M2), (C1), (C2) and (C3). The essential condition here is the threefold continuous differentiability of the cumulative intensity with respect to θ. The first and second order derivatives explicitly appear in the distributional approximation of the MDE (cf. Kopperschmidt and Stute 2013, p. 1280), whereas the third order derivatives are required to dominate its negligible parts. Throughout this section, the notations $\frac{\mathrm{d}}{\mathrm{d}\theta}$ and D_θ are used interchangeably to indicate the total derivative with respect to θ: The former is preferred to pick up the notation found in Kopperschmidt and Stute 2013 while maintaining distinctness from the partial derivatives $\frac{\partial}{\partial\theta_j}$, whereas the latter permits tighter formulas, especially in conjunction with further derivatives or whenever we seek to highlight the matrix structure of the total derivative. Likewise, we also denote the pth total differential as either $\frac{\mathrm{d}^p}{\mathrm{d}\theta^p}$ or D_θ^p depending on the context. Finally, we will abbreviate the notation for derivatives wherever appropriate by writing, for instance,

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta^0}(t) \quad \text{instead of} \quad \frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta(t)\bigg|_{\theta=\theta^0}.$$

Assumption 3.8 (Requirements Related to the Asymptotic Normality of the MDE)**.**
In order to obtain the asymptotic normality of the minimum distance estimator, the following conditions are henceforth assumed to be valid.

(A1) There exists an open neighbourhood $\mathrm{B}_\varepsilon(\theta^*) \subset \Theta$ of θ^*, such that for each fixed $t \in \mathcal{I}$, the mapping

$$\begin{aligned} \Lambda_\cdot(t) \,:\, \Theta &\longrightarrow \mathbb{R} \\ \theta &\longmapsto \Lambda_\theta(t) \end{aligned}$$

is $\mathbb{P}$-almost surely three times continuously differentiable with respect to θ, that is,

$$\Lambda_\cdot(t) \in C^3\left(\mathrm{B}_\varepsilon(\theta^*)\right) \quad \mathbb{P}\text{-almost surely.}$$

(A2) The partial derivatives up to the third order are $\mathbb{P}$-almost surely continuous on $\mathcal{I} \times \mathrm{B}_\varepsilon(\theta^*)$ (as functions of both t and θ). Hence, with probability 1,

$$\begin{aligned}
\frac{\partial}{\partial\theta_j}\Lambda_\theta(\cdot)\Big|_{\theta=\cdot} &\in C^0\left(\mathcal{I} \times \mathrm{B}_\varepsilon(\theta^*)\right), \quad \text{for all } j \in \{1,\dots,d\},\\
\frac{\partial^2}{\partial\theta_j\partial\theta_k}\Lambda_\theta(\cdot)\Big|_{\theta=\cdot} &\in C^0\left(\mathcal{I} \times \mathrm{B}_\varepsilon(\theta^*)\right), \quad \text{for all } j,k \in \{1,\dots,d\},\\
\frac{\partial^3}{\partial\theta_j\partial\theta_k\partial\theta_l}\Lambda_\theta(\cdot)\Big|_{\theta=\cdot} &\in C^0\left(\mathcal{I} \times \mathrm{B}_\varepsilon(\theta^*)\right), \quad \text{for all } j,k,l \in \{1,\dots,d\}.
\end{aligned}$$

(A3) There exists a constant $C > 0$, such that the third-order partial derivatives with respect to θ are $\mathbb{P}$-almost surely uniformly bounded by C, that is,

$$\left|\frac{\partial^3}{\partial\theta_j\partial\theta_k\partial\theta_l}\Lambda_\theta(t)\right| \le C\,, \quad \text{for all } t \in \mathcal{I},\ \theta \in \mathrm{B}_\varepsilon(\theta^*) \text{ and } j,k,l \in \{1,\dots,d\}.$$

(A4) For each $v \in \mathbb{R}^d \setminus \{0\}$ there exists a Borel set $B_v \subset \mathbb{R}$ with $\mu_{\mathbb{E}\Lambda_{\theta^*}}(B_v) > 0$ such that

$$v^\top \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta^*}(t)^\top \neq 0\,, \quad \text{for all } t \in B_v.$$

Since by Theorem 3.7 the minimum distance estimator is strongly consistent for θ^*, it holds with probability 1 that $\hat{\theta}_n \in \mathrm{B}_\varepsilon(\theta^*)$ for almost all $n \in \mathbb{N}$. As we seek to differentiate at $\hat{\theta}_n$, we will hereafter always assume n to be chosen sufficiently large. Moreover, it then holds for these n that $\hat{\theta}_n$ is an interior point of Θ.

Remark 3.9 (Boundedness of First and Second Order Derivatives of the Cumulative Intensity)**.**
Condition (A3) also extends to the partial derivatives of the first and second order, so that these can also be assumed to be $\mathbb{P}$-almost surely uniformly bounded by some (properly adjusted) constant C. We demonstrate this exemplarily only for the second order derivatives, since the proof for the first order proceeds completely analogous. Fix an arbitrary $\theta^0 \in \Theta$. For any $\theta \in \Theta$, the mean value theorem for functions of several variables then yields:

$$\frac{\partial^2}{\partial\theta_j\partial\theta_k}\Lambda_\theta(t) - \frac{\partial^2}{\partial\theta_j\partial\theta_k}\Lambda_{\theta^0}(t) = D_\theta \frac{\partial^2}{\partial\theta_j\partial\theta_k}\Lambda_\theta(t)\Big|_{\theta=\tilde{\theta}} \left(\theta - \theta^0\right), \tag{3.10}$$

for some $\tilde{\theta} \in \{\theta^0 + s(\theta - \theta^0) \,:\, s \in [0,1]\}$, i.e., on the line segment adjoining θ^0 and θ in $\mathbb{R}^d$. If $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^d$, we obtain from condition (A3):

$$\begin{aligned}
\left\|D_\theta \frac{\partial^2}{\partial\theta_j\partial\theta_k}\Lambda_\theta(t)\Big|_{\theta=\tilde{\theta}}\right\|^2 &= \left\|\left(\frac{\partial^3}{\partial\theta_j\partial\theta_k\partial\theta_l}\Lambda_\theta(t)\Big|_{\theta=\tilde{\theta}}\right)_{l=1,\dots,d}\right\|^2\\
&= \sum_{l=1}^d \underbrace{\left|\frac{\partial^3}{\partial\theta_j\partial\theta_k\partial\theta_l}\Lambda_\theta(t)\Big|_{\theta=\tilde{\theta}}\right|^2}_{\le C} \le dC^2
\end{aligned}$$

so that we can use the Cauchy-Schwarz inequality to infer the Lipschitz continuity of the second order partial derivatives from Equation (3.10),

$$\begin{aligned}\left|\frac{\partial^2}{\partial\theta_j\partial\theta_k}\Lambda_\theta(t)-\frac{\partial^2}{\partial\theta_j\partial\theta_k}\Lambda_{\theta^0}(t)\right| &\leq \left\|D_\theta\frac{\partial^2}{\partial\theta_j\partial\theta_k}\Lambda_\theta(t)\Big|_{\theta=\tilde{\theta}}\right\|\left\|\theta-\theta^0\right\| \\ &\leq \sqrt{d}C\left\|\theta-\theta^0\right\| . \end{aligned} \tag{3.11}$$

From here, the triangle inequality enables us to separate the variables t and θ as follows:

$$\begin{aligned}\left|\frac{\partial^2}{\partial\theta_j\partial\theta_k}\Lambda_\theta(t)\right| &\leq \left|\frac{\partial^2}{\partial\theta_j\partial\theta_k}\Lambda_\theta(t)-\frac{\partial^2}{\partial\theta_j\partial\theta_k}\Lambda_{\theta^0}(t)\right|+\left|\frac{\partial^2}{\partial\theta_j\partial\theta_k}\Lambda_{\theta^0}(t)\right| \\ &\leq \sqrt{d}C\sup_{\theta\in\Theta}\left\|\theta-\theta^0\right\|+\sup_{t\in\mathcal{I}}\left|\frac{\partial^2}{\partial\theta_j\partial\theta_k}\Lambda_{\theta^0}(t)\right| . \end{aligned} \tag{3.12}$$

The right-hand side of Equation (3.12) then provides the adjusted constant, since both summands are finite and no longer depend on θ or t: For the first one we utilize the boundedness of Θ, while for the second one condition (A2) is needed in conjunction with the compactness of $\mathcal{I}$.
We can thus formulate an extended condition ($\widetilde{\text{A}3}$), which is redundant in the context of Assumption 3.8 since it can be inferred from conditions (A1) to (A3). However, this allows us to later refer to ($\widetilde{\text{A}3}$) instead of "(A2) and (A3) in conjunction with Remark 3.9".

($\widetilde{\text{A}3}$) There exists a constant $C > 0$, such that the partial derivatives with respect to θ are $\mathbb{P}$-almost surely uniformly bounded by C up to the third order, that is,

$$\begin{aligned}&\left|\frac{\partial}{\partial\theta_j}\Lambda_\theta(t)\right|\leq C\,, \qquad \left|\frac{\partial^2}{\partial\theta_j\partial\theta_k}\Lambda_\theta(t)\right|\leq C\,, \\ &\left|\frac{\partial^3}{\partial\theta_j\partial\theta_k\partial\theta_l}\Lambda_\theta(t)\right|\leq C\,, \quad \text{for all } t\in\mathcal{I},\ \theta\in \mathrm{B}_\varepsilon(\theta^*) \text{ and } j,k,l\in\{1,\dots,d\}.\end{aligned}$$

Before comparing our premises with those of Kopperschmidt and Stute as in Section 3.2, we would like to state an elementary but crucial lemma: It forms the starting point for the upcoming proof and allows us to both sketch the idea of proof and outline the differences with Kopperschmidt's approach.

Lemma 3.10 (Derivative of the Cramér-von Mises Distance; cf. Lemma 7 of Kopperschmidt and Stute 2013, p. 1288)**.**
Under Assumption 3.8, we have[5]*:*

$$\begin{aligned}&\sqrt{n}\int_{\mathcal{I}}\left[\overline{\Lambda}_{\hat{\theta}_n}^{(n)}-\overline{\Lambda}_{\theta^*}^{(n)}\right]\frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)}\Big|_{\theta=\hat{\theta}_n}\mathrm{d}\overline{M}^{(n)}+\sqrt{n}\int_{\mathcal{I}}\left[\overline{\Lambda}_{\hat{\theta}_n}^{(n)}-\overline{\Lambda}_{\theta^*}^{(n)}\right]\frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)}\Big|_{\theta=\hat{\theta}_n}\mathrm{d}\overline{\Lambda}_{\theta^*}^{(n)} \\ &\quad=\sqrt{n}\int_{\mathcal{I}}\overline{M}^{(n)}\frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)}\Big|_{\theta=\hat{\theta}_n}\mathrm{d}\overline{M}^{(n)}+\sqrt{n}\int_{\mathcal{I}}\overline{M}^{(n)}\frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)}\Big|_{\theta=\hat{\theta}_n}\mathrm{d}\overline{\Lambda}_{\theta^*}^{(n)} . \end{aligned} \tag{3.13}$$

[5]Reminder: Even in the presence of non-predictable integrands, the existence of the stochastic integrals involved can always be seen by rewriting the individual martingale $M^{(m)}$ as the difference of the counting process $N^{(m)}$ and its compensator $\Lambda_{\theta^*}^{(m)}$, which are both non-decreasing.

Proof. By definition, the minimum distance estimator $\hat{\theta}_n$ minimizes the Cramér-von Mises distance between $\overline{N}^{(n)}$ and $\overline{\Lambda}_\theta^{(n)}$. Moreover, $\hat{\theta}_n$ lies in $\mathrm{B}_\varepsilon(\theta^*)$ and thus in the interior of Θ according to Assumption 3.8. Since by (A1) the cumulative intensity is continuously differentiable as a function of θ for each fixed t with probability 1, this is also true for the Cramér-von Mises distance, as can be seen from Equation (3.4). Hence, there is a critical point at $\hat{\theta}_n$, meaning that

$$\begin{aligned}
0 &= \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}\theta}\left\|\overline{N}^{(n)} - \overline{\Lambda}_\theta^{(n)}\right\|^2_{\overline{N}^{(n)}}\Bigg|_{\theta=\hat{\theta}_n} = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}\theta}\int_{\mathcal{I}}\left[\overline{N}^{(n)} - \overline{\Lambda}_\theta^{(n)}\right]^2 \mathrm{d}\overline{N}^{(n)}\Bigg|_{\theta=\hat{\theta}_n} \\
&= \frac{1}{2}\int_{\mathcal{I}}\frac{\mathrm{d}}{\mathrm{d}\theta}\left[\overline{N}^{(n)} - \overline{\Lambda}_\theta^{(n)}\right]^2\Bigg|_{\theta=\hat{\theta}_n}\mathrm{d}\overline{N}^{(n)} = \int_{\mathcal{I}}\left[\overline{N}^{(n)} - \overline{\Lambda}_{\hat{\theta}_n}^{(n)}\right]\frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)}\Bigg|_{\theta=\hat{\theta}_n}\mathrm{d}\overline{N}^{(n)} \\
&= \int_{\mathcal{I}}\left[\overline{N}^{(n)} - \overline{\Lambda}_{\theta^*}^{(n)}\right]\frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)}\Bigg|_{\theta=\hat{\theta}_n}\mathrm{d}\overline{N}^{(n)} + \int_{\mathcal{I}}\left[\overline{\Lambda}_{\theta^*}^{(n)} - \overline{\Lambda}_{\hat{\theta}_n}^{(n)}\right]\frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)}\Bigg|_{\theta=\hat{\theta}_n}\mathrm{d}\overline{N}^{(n)} \\
&= \int_{\mathcal{I}}\overline{M}^{(n)}\frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)}\Bigg|_{\theta=\hat{\theta}_n}\mathrm{d}\overline{N}^{(n)} + \int_{\mathcal{I}}\left[\overline{\Lambda}_{\theta^*}^{(n)} - \overline{\Lambda}_{\hat{\theta}_n}^{(n)}\right]\frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)}\Bigg|_{\theta=\hat{\theta}_n}\mathrm{d}\overline{N}^{(n)}, \qquad (3.14)
\end{aligned}$$

where we used the identity $\overline{M}_n = \overline{N}^{(n)} - \overline{\Lambda}_{\theta^*}^{(n)}$ from Equation (3.3). Integration and differentiation can be interchanged here due to the Leibniz integral rule, as the derivatives of the integrand with respect to θ are $\mathbb{P}$-almost surely uniformly bounded according to conditions (C3) and ($\widetilde{\mathrm{A}}$3) and thus trivially integrable[6]. By subtracting the second summand of Equation (3.14) on both sides, exploiting the above identity once again and multiplying by $\sqrt{n}$ one then obtains Equation (3.13). □

We introduce auxiliary parametric processes which enable a condensed representation of Lemma 3.10. The notation is adapted from Kopperschmidt and Stute 2013 to facilitate comparison.

Definition 3.11 (Auxiliary Parametric Processes)**.**
We define the following parametric processes, each with index set $B_\varepsilon(\theta^*) \subset \Theta$, where as usual the dependence on the realization ω is suppressed:

$$\begin{aligned}
\alpha_n(\theta) &:= \sqrt{n}\int_{\mathcal{I}}\left[\overline{\Lambda}_\theta^{(n)} - \overline{\Lambda}_{\theta^*}^{(n)}\right]\frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)}\,\mathrm{d}\overline{M}^{(n)}, \\
\beta_n(\theta) &:= \sqrt{n}\int_{\mathcal{I}}\overline{M}^{(n)}\frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)}\,\mathrm{d}\overline{M}^{(n)}, \\
\gamma_n(\theta) &:= \sqrt{n}\int_{\mathcal{I}}\overline{M}^{(n)}\frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)}\,\mathrm{d}\overline{\Lambda}_{\theta*}^{(n)}, \\
\Phi_n(\theta) &:= \frac{\mathrm{d}}{\mathrm{d}\theta}\int_{\mathcal{I}}\left[\overline{\Lambda}_\theta^{(n)} - \overline{\Lambda}_{\theta^*}^{(n)}\right]\frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)}\,\mathrm{d}\overline{\Lambda}_{\theta^*}^{(n)}, \\
\Phi_0(\theta) &:= \frac{\mathrm{d}}{\mathrm{d}\theta}\int_{\mathcal{I}}\left[\mathbb{E}\Lambda_\theta - \mathbb{E}\Lambda_{\theta^*}\right]\mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta\,\mathrm{d}\mathbb{E}\Lambda_{\theta^*}, \\
\Psi_n(\theta) &:= \int_0^1 \Phi_n\left(\theta^* + s\left(\theta - \theta^*\right)\right)\mathrm{d}s\,.
\end{aligned}$$

By definition, $\Psi_n(\theta)$ can be understood as the average of the matrix-valued process Φ_n along $\{\theta^* + s(\theta - \theta^*)\ :\ s \in [0,1]\}$, i.e., the line connecting θ^* and θ in $\mathrm{B}_\varepsilon(\theta^*) \subset \mathbb{R}^d$.

[6] This reasoning will be illustrated in more detail later using the parametric process Φ_n, see page 52.

Moreover, we will later encounter Φ_0 as the uniform limit of Φ_n. Each of the other three processes can directly be associated with a summand from Lemma 3.10. With Definition 3.11 in mind, Equation (3.13) becomes:

$$\alpha_n\left(\hat{\theta}_n\right) + \sqrt{n}\int_{\mathcal{I}}\left[\overline{\Lambda}^{(n)}_{\hat{\theta}_n} - \overline{\Lambda}^{(n)}_{\theta^*}\right]\frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}^{(n)}_{\theta}\bigg|_{\theta=\hat{\theta}_n}\mathrm{d}\overline{\Lambda}^{(n)}_{\theta^*} = \beta_n\left(\hat{\theta}_n\right) + \gamma_n\left(\hat{\theta}_n\right). \tag{3.15}$$

For the remaining integral, an application of the mean value theorem for vector-valued functions[7] (see Forster 2017, pp. 84–85) yields:

$$\begin{aligned}
&\int_{\mathcal{I}}\left[\overline{\Lambda}^{(n)}_{\hat{\theta}_n} - \overline{\Lambda}^{(n)}_{\theta^*}\right]\frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}^{(n)}_{\theta}\bigg|_{\theta=\hat{\theta}_n}\mathrm{d}\overline{\Lambda}^{(n)}_{\theta^*}\\
&= \int_{\mathcal{I}}\left[\overline{\Lambda}^{(n)}_{\hat{\theta}_n} - \overline{\Lambda}^{(n)}_{\theta^*}\right]\frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}^{(n)}_{\theta}\bigg|_{\theta=\hat{\theta}_n}\mathrm{d}\overline{\Lambda}^{(n)}_{\theta^*} - \underbrace{\int_{\mathcal{I}}\left[\overline{\Lambda}^{(n)}_{\theta^*} - \overline{\Lambda}^{(n)}_{\theta^*}\right]\frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}^{(n)}_{\theta}\bigg|_{\theta=\theta^*}\mathrm{d}\overline{\Lambda}^{(n)}_{\theta^*}}_{=0}\\
&= \left(\int_0^1 \frac{\mathrm{d}}{\mathrm{d}\theta}\int_{\mathcal{I}}\left[\overline{\Lambda}^{(n)}_{\theta} - \overline{\Lambda}^{(n)}_{\theta^*}\right]\frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}^{(n)}_{\theta}\,\mathrm{d}\overline{\Lambda}^{(n)}_{\theta^*}\bigg|_{\theta=\theta^*+s\left(\hat{\theta}_n-\theta^*\right)}\mathrm{d}s\right)\left(\hat{\theta}_n-\theta^*\right)\\
&= \left(\int_0^1 \Phi_n\left(\theta^* + s\left(\hat{\theta}_n-\theta^*\right)\right)\mathrm{d}s\right)\left(\hat{\theta}_n-\theta^*\right) = \Psi_n\left(\hat{\theta}_n\right)\left(\hat{\theta}_n-\theta^*\right),
\end{aligned} \tag{3.16}$$

so that by substituting Equation (3.16) into Equation (3.15) we can restate Lemma 3.10 in terms of Definition 3.11.

Lemma 3.10 (Abbreviated Form)**.**
Under Assumption 3.8 and with the notation from Definition 3.11, it holds:

$$\alpha_n\left(\hat{\theta}_n\right) + \sqrt{n}\Psi_n\left(\hat{\theta}_n\right)\left(\hat{\theta}_n-\theta^*\right) = \beta_n\left(\hat{\theta}_n\right) + \gamma_n\left(\hat{\theta}_n\right). \tag{3.17}$$

Starting from Lemma 3.10, the proof of Kopperschmidt and Stute and the approach taken here begin to diverge. Therefore, this point lends itself to a discussion of the differences. We outline and summarize the proof steps followed by Kopperschmidt and Stute and then highlight our deviations. We refer to Kopperschmidt only in the following, since the complete proof given in his dissertation (see Kopperschmidt 2005) is better suited for distinguishing its individual parts. As a final remark, we note that instead of considering the averaged process Ψ_n, Kopperschmidt and Stute evaluate the process Φ_n at some "appropriate intermediate point" (Kopperschmidt and Stute 2013, p. 1294). However, such an intermediate point does not necessarily exist as soon as the dimension d of the parameter space Θ - and thus the dimension of both the domain (which is d) and codomain (which is d^2) of Φ_n - exceeds 1. To avoid confusion, below we present the proof as if Kopperschmidt had considered Ψ_n instead of Φ_n.

(i) Kopperschmidt uses the differentiability of the cumulative intensities (see condition (A1)) to obtain local Lipschitz-constants. These Lipschitz constants are random variables by themselves, so he imposes additional moment conditions on them. Furthermore, the existing condition ($\widetilde{\mathrm{C}}3$) is strengthened (cf. Kopperschmidt 2005, pp. 27–28).

[7] Its applicability is once again due to conditions (A1) through (A3). The technical details are deferred here and will be discussed in subsequent proofs.

(ii) Kopperschmidt continues by further decomposing the auxiliary parametric process of Definition 3.11. He employs Kolmogorov's tightness criterion (see Theorem B.2.6) to show the tightness of the decomposed processes, for which the local Lipschitz constants are needed. As sums of tight processes are in turn tight (a direct consequence of Corollary B.2.5), the tightness of $(\alpha_n(\cdot))_{n\in\mathbb{N}}$, $(\beta_n(\cdot))_{n\in\mathbb{N}}$, $(\gamma_n(\cdot))_{n\in\mathbb{N}}$ and $(\Psi_n(\cdot))_{n\in\mathbb{N}}$ ensues.

(iii) An essential problem is that the stochastic processes in Lemma 3.10 are evaluated at the MDE, which adds a secondary source of randomness. Due to the strong consistency of the minimum distance estimator, Kopperschmidt can utilize the tightness of the involved processes to conclude that Equation (3.17) is asymptotically equivalent to

$$\alpha_n(\theta^*) + \sqrt{n}\Psi_n(\theta^*)\left(\hat{\theta}_n - \theta^*\right) = \beta_n(\theta^*) + \gamma_n(\theta^*)$$

by virtue of Lemma B.2.7. Since $\alpha_n(\theta^*) = 0$ and because Kopperschmidt demonstrates that $\beta_n(\theta^*) = o_{\mathbb{P}}(1)$ using L^2 techniques (cf. Kopperschmidt 2005, pp. 139–143), this further simplifies to

$$\sqrt{n}\Psi_n(\theta^*)\left(\hat{\theta}_n - \theta^*\right) = \gamma_n(\theta^*)\,. \tag{3.18}$$

While the processes in Equation (3.18) are still stochastic, their arguments are now deterministic (albeit unknown).

(iv) For any fixed θ, Kopperschmidt shows that the process γ_n admits the representation

$$\gamma_n(\theta) = \sqrt{n}\int_{\mathcal{I}} \overline{M}^{(n)}\mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta \,\mathrm{d}\mathbb{E}\Lambda_{\theta^*} + o_{\mathbb{P}}(1)\,.$$

Using Fubini's theorem, he rearranges the integral to obtain[8]:

$$\sqrt{n}\int_{\mathcal{I}} \overline{M}^{(n)}\mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta \,\mathrm{d}\mathbb{E}\Lambda_{\theta^*} = \sqrt{n}\sum_{i=1}^{n}\int_{\mathcal{I}} \varphi_\theta \,\mathrm{d}M^{(i)} \tag{3.19}$$

for some deterministic function φ_θ. By Lemma A.40, on the right-hand side of Equation (3.19) there is a sum of centred i.i.d. random variables. Specifically, for $\theta = \theta^*$, the central limit theorem yields:

$$\gamma_n(\theta^*) \xrightarrow{\mathrm{d}} \mathcal{N}_d\left(0, \Sigma\right) \quad (n \to \infty)\,.$$

(v) In the penultimate step, Kopperschmidt shows that $\Psi_n(\theta)$ converges in probability to a deterministic standardizing matrix $\Phi_0(\theta)$ for any fixed θ. Given the condition (A4), he then verifies that this matrix is positive definite and therefore invertible. This allows the application of Corollary B.4.2, and by Slutzky's theorem the asymptotic distribution does not change when $\Psi_n(\theta^*)$ is replaced by $\Phi_0(\theta^*)$ in Equation (3.18).

[8] The exact procedure is described later in the proof of Lemma 3.24. It is mainly needed to derive a convenient expression for the asymptotic covariance matrix.

(vi) Combining the findings of (iv) and (v), Kopperschmidt concludes that

$$\sqrt{n}\Phi_0(\theta^*)\left(\hat{\theta}_n - \theta^*\right) \xrightarrow{\mathrm{d}} \mathcal{N}_d(0, \Sigma)\,,$$

or, equivalently, since $\Phi_0(\theta^*)$ and its inverse are symmetric,

$$\sqrt{n}\left(\hat{\theta}_n - \theta^*\right) \xrightarrow{\mathrm{d}} \mathcal{N}_d\left(0, \Phi_0(\theta^*)^{-1}\,\Sigma\,\Phi_0(\theta^*)^{-1}\right)\,.$$

We pointed out earlier in Remark 3.1 that we strengthened the assumptions made in Kopperschmidt and Stute 2013 to dispense with the Kolmogorov tightness criterion because it was found not to be applicable as intended by Kopperschmidt and Stute. Clearly, proving tightness in another way would be just as purposeful to simplify the identity (3.17) to that from Equation (3.18), so that finding an appropriate tightness criterion would be sufficient to correct the proof. However, we bypass this intermediate step altogether and infer the desired convergence directly from Equation (3.17) by rearranging superfluous terms. Essentially, we use assumption ($\widetilde{\mathrm{A}}3$) to reason based on the mean value theorem in lieu of tightness criteria. Instead of steps (i) through (iii) above, we derive convenient representations of the auxiliary processes α_n, β_n and γ_n to identify their negligible parts and use the preliminary work done in steps (iv) through (vi) to our advantage. The proof is carried out in the following steps, the order of which can be partially permuted. To simplify the comparison to the proof of Kopperschmidt and Stute, a similar order as above is chosen here.

(i) We apply the mean value theorem to the auxiliary parametric processes α_n, β_n and γ_n to get

$$\begin{aligned}\alpha_n(\hat{\theta}_n) &= \alpha_n(\theta^*) + \sqrt{n}A_n(\hat{\theta}_n - \theta^*)\,,\\ \beta_n(\hat{\theta}_n) &= \beta_n(\theta^*) + \sqrt{n}B_n(\hat{\theta}_n - \theta^*)\,,\\ \gamma_n(\hat{\theta}_n) &= \gamma_n(\theta^*) + \sqrt{n}C_n(\hat{\theta}_n - \theta^*)\,.\end{aligned}$$

Here - let us ignore the scaling factor $\sqrt{n}$ for the moment - the random matrix A_n (and, similarly, B_n and C_n) cannot simply be chosen as $D_\theta\alpha_n$ evaluated at a suitable intermediate point $\tilde{\theta}_n$ on the line connecting θ^* and $\hat{\theta}_n$, since such a point does not exist in general. Instead, the matrix is composed of the vectors obtained by applying the mean value theorem for functions of several variables to the individual components $\alpha_{n,j}$ of α_n. Specifically, the j-th row of A_n consists of $D_\theta\alpha_{n,j}(\tilde{\theta}_{n,j})$, where the respective parameter $\tilde{\theta}_{n,j}$ depends on the component under consideration.

(ii) We use the uniform boundedness assumption on our model to derive that the matrices in the above representation are asymptotically insignificant, that is, $A_n = o_{\mathbb{P}}(1)$, $B_n = o_{\mathbb{P}}(1)$ and $C_n = o_{\mathbb{P}}(1)$ (and thereby avoid the application of tightness criteria). This is equivalent to showing that the rows of the above matrices each converge to 0 in probability. To this end, we again use the mean value theorem for vector-valued functions, or equivalently, and to be consistent with the following interpretation, a first-order Taylor approximation at θ^*. We then proceed to prove that the constant part is negligible, while the slope of the linear part is bounded according to our assumptions. The asymptotic behavior is therefore determined by the differences $\tilde{\theta}_{n,j} - \theta^*$, where $\tilde{\theta}_{n,j}$ is located between the true parameter θ^* and the MDE $\hat{\theta}_n$.

Convergence to 0 then follows from the strong consistency of the MDE.

(iii) We recall that $\alpha_n(\theta^*) = 0$ and $\beta_n(\theta^*) = o_{\mathbb{P}}(1)$ (this was already shown by Kopperschmidt). The identity (3.17) then becomes asymptotically equivalent to

$$\begin{aligned} & \sqrt{n}A_n(\hat{\theta}_n - \theta^*) + \sqrt{n}\Psi_n(\hat{\theta}_n)(\hat{\theta}_n - \theta^*) \\ & \quad = \sqrt{n}B_n(\hat{\theta}_n - \theta^*) + \gamma_n(\theta^*) + \sqrt{n}C_n(\hat{\theta}_n - \theta^*) \\ \iff \quad & \sqrt{n}\left(\Psi_n(\hat{\theta}_n) + A_n - B_n - C_n\right)(\hat{\theta}_n - \theta^*) = \gamma_n(\theta^*)\,, \end{aligned} \tag{3.20}$$

which resembles Equation (3.18). Note that here, in contrast to Kopperschmidt's proof, the MDE still appears in the argument of the standardizing matrix Ψ_n.

(iv) With arguments analogous to Kopperschmidt's proof, we obtain that

$$\gamma_n(\theta^*) \xrightarrow{\mathrm{d}} \mathcal{N}_d(0, \Sigma) \quad (n \to \infty)$$

by the central limit theorem.

(v) Other than in Kopperschmidt's proof, we cannot assume the tightness of the sequence $(\Psi_n(\cdot))_{n\in\mathbb{N}}$ in this step. We therefore first show the uniform convergence of Φ_n to Φ_0 in probability. We then recall that $\Psi_n(\hat{\theta}_n)$ is the average of the parametric process Φ_n along the line connecting θ^* and $\hat{\theta}_n$. Since the strong consistency of the MDE implies that this line contracts to θ^*, the convergence of $\Psi_n(\hat{\theta}_n)$ to $\Phi_0(\theta^*)$ follows. Slutzky's theorem hence provides that

$$\Psi_n(\hat{\theta}_n) + A_n - B_n - C_n \xrightarrow{\mathbb{P}} \Phi_0(\theta^*)\,.$$

(vi) Replacing the standardizing matrix on the left-hand side of Equation (3.20) by $\Phi_0(\theta^*)$, we can conclude, as before, that

$$\sqrt{n}\Phi_0(\theta^*)\left(\hat{\theta}_n - \theta^*\right) \xrightarrow{\mathrm{d}} \mathcal{N}_d(0, \Sigma)\,.$$

The above order of steps closely follows the structure of Kopperschmidt and Stute's approach. As we prefer an order adapted to our proof, we proceed as follows:
First, we examine the standardizing matrix from step (v) in Paragraph 3.3.1 "Asymptotics of the Standardizing Matrix Ψ_n". Along the way, we study the limits of average cumulative intensities and their derivatives, which later give an intuition for the asymptotics of the other auxiliary processes, but are not directly relevant to their representation theorems. Next, we prove the representations from steps (i) and (ii) separately for the remaining three auxiliary parametric processes α_n, β_n and γ_n. Of particular interest are the leading terms $\beta_n(\theta^*)$ and $\gamma_n(\theta^*)$, the former vanishing for $n \to \infty$ as per step (iii) and the latter contributing the essential part of the asymptotic behavior according to step (iv). Together, these steps form the Paragraph 3.3.2 "Asymptotics of the Auxiliary Parametric Processes α_n, β_n and γ_n". Finally, we combine our findings to prove the main theorem of this section, which we will state here in advance, in Paragraph 3.3.3 "Proof of the Asymptotic Normality".

Theorem 3.12 (Asymptotic Normality of the Minimum Distance Estimator; cf. Kopperschmidt and Stute 2013, p. 1281)**.**
Under assumptions (A1), (A2), (A3) *and* (A4) *together with the assumptions from Section 3.2 holds:*

$$\sqrt{n}\Phi_0(\theta^*)\left(\hat{\theta}_n - \theta^*\right) \xrightarrow{\mathrm{d}} \mathcal{N}_d\left(0, \Sigma(\theta^*)\right) \quad (n \to \infty)\,,$$

where $\Sigma(\theta^*)$ *is a* $d \times d$ *matrix with entries*

$$\begin{aligned} \Sigma_{ij}(\theta^*) &:= \int_{\mathcal{I}} \varphi_i(t)\varphi_j(t)\, \mathrm{d}\mathbb{E}\Lambda_{\theta^*}(t)\,, && 1 \le i, j \le d\,, && (3.21) \\ \varphi(t) = (\varphi_1(t), \dots, \varphi_d(t)) &:= \int_{[t,\tau]} \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta^*}\, \mathrm{d}\mathbb{E}\Lambda_{\theta^*}\,, && t \in \mathcal{I}\,. \end{aligned}$$

3.3.1. Asymptotics of the Standardizing Matrix Ψ_n

We first turn to the matrix-valued auxiliary processes from Definition 3.11. Our preliminary considerations here concern the process Φ_n and, as we shall prove, its uniform limit Φ_0.

Proposition 3.13 (Uniform Convergence of Φ_n on Compact Subsets of $\mathrm{B}_\varepsilon(\theta^*)$; cf. Kopperschmidt and Stute 2013, p. 1293)**.**
Let $\|\cdot\|$ *denote any matrix norm on* $\mathbb{R}^{d\times d}$ *and let* K *be an arbitrary compact subset of the open set* $B_\varepsilon(\theta^*)$ *given by Assumption 3.8. Then,*

$$\sup_{\theta\in K} \|\Phi_n(\theta) - \Phi_0(\theta)\| \xrightarrow{\mathbb{P}} 0 \quad \textit{as } n \to \infty.$$

Before proceeding to the proof of this proposition, we briefly explain the matrix norm used hereafter and refer to the appendix on why considering "more benign" (i.e., sub-multiplicative) norms is not of interest to us.

Remark 3.14 (On the Choice of Matrix Norm)**.**
Without loss of generality, the equivalence of norms on finite dimensional spaces (in our case, on the d^2-dimensional $\mathbb{R}$-vector field $\mathbb{R}^{d\times d}$) allows us to use the max norm $\|\cdot\|_{\max}$ for matrices. The norm $\|A\|_{\max}$ of any matrix $A \in \mathbb{R}^{k\times l}$ amounts to the maximum absolute element of A, that is,

$$\|A\|_{\max} := \max_{\substack{i=1,\dots,k \\ j=1,\dots,l}} |a_{ij}|\,, \quad A \in \mathbb{R}^{k\times l}\,.$$

The matrix norm defined in this way is not sub-multiplicative, although this can be remedied by rescaling with the factor $\sqrt{kl}$. We address this issue in Remark B.4.8 in the appendix. We also demonstrate there why the lack of sub-multiplicity does not entail any functional disadvantages and why the simplicity of the max norm prevails as the decisive advantage.

Let us now consider the parametric processes Φ_n and Φ_0 from Definition 3.11,

$$\begin{aligned} \Phi_n(\theta) &:= \frac{\mathrm{d}}{\mathrm{d}\theta}\int_{\mathcal{I}} \left[\overline{\Lambda}_\theta^{(n)} - \overline{\Lambda}_{\theta^*}^{(n)}\right] \frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)}\, \mathrm{d}\overline{\Lambda}_{\theta^*}^{(n)}\,, \\ \Phi_0(\theta) &:= \frac{\mathrm{d}}{\mathrm{d}\theta}\int_{\mathcal{I}} \left[\mathbb{E}\Lambda_\theta - \mathbb{E}\Lambda_{\theta^*}\right] \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta\, \mathrm{d}\mathbb{E}\Lambda_{\theta^*}\,. \end{aligned}$$

We are faced here with differentiating a parameter integral. Differentiation and integration can be interchanged here, as we demonstrate with the example of Φ_n: The integrand is at least three times continuously differentiable on $\mathrm{B}_\varepsilon(\theta^*)$ w.r.t. θ for almost every $(t,\omega) \in \mathcal{I} \times \Omega$ by condition (A1), and we obtain by the product rule that

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\left[\overline{\Lambda}_\theta^{(n)} - \overline{\Lambda}_{\theta^*}^{(n)}\right] \frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)}\right) = \frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)\top} \frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)} + \left[\overline{\Lambda}_\theta^{(n)} - \overline{\Lambda}_{\theta^*}^{(n)}\right] \frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\overline{\Lambda}_\theta^{(n)}. \tag{3.22}$$

We can differentiate under the integral sign according to the measure theoretic version of the well known Leibniz integral rule (cf. Lemma 16.2 of Bauer 2001, p. 89) if integrable majorants for the partial derivatives occurring in Equation (3.22) can be found. In fact, due to conditions (C3) and ($\widetilde{\mathrm{A}}3$), these partial derivatives are $\mathbb{P}$-almost surely uniformly bounded, which implies integrability, since the measure induced by $\overline{\Lambda}_{\theta^*}^{(n)}$ is finite on $\mathcal{I}$ with probability 1. Hence,

$$\Phi_n(\theta) = \int_{\mathcal{I}} \frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)\top} \frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)} + \left[\overline{\Lambda}_\theta^{(n)} - \overline{\Lambda}_{\theta^*}^{(n)}\right] \frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\overline{\Lambda}_\theta^{(n)}\,\mathrm{d}\overline{\Lambda}_{\theta^*}^{(n)}, \tag{3.23}$$

and, as the same arguments apply to Φ_0 instead of Φ_n,

$$\Phi_0(\theta) = \int_{\mathcal{I}} \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta^\top \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta + \left[\mathbb{E}\Lambda_\theta - \mathbb{E}\Lambda_{\theta^*}\right] \mathbb{E}\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\Lambda_\theta\,\mathrm{d}\mathbb{E}\Lambda_{\theta^*}, \tag{3.24}$$

where also the expectation and the derivative w.r.t. θ were interchanged with analogous justification. The proof of Proposition 3.13 can now be carried out in two steps:
The first step of the proof involves showing the uniform convergence on $\mathcal{I} \times K$ of both the integrand (in probability) and the integrator ($\mathbb{P}$-almost surely). In the second step, this convergence is then transferred to the integral.
To make the proof easier to follow, we split its first part into two lemmas.

Lemma 3.15 (Uniform Convergence of Average Cumulative Intensities)**.**
Let K be an arbitrary compact subset of $B_\varepsilon(\theta^*)$, where again the conditions from Assumption 3.8 are assumed to hold. Then, with probability 1,

$$\overline{\Lambda}_\theta^{(n)} \longrightarrow \mathbb{E}\Lambda_\theta \quad \textit{uniformly on } \mathcal{I} \times K \textit{ as } n \to \infty.$$

Proof. We recall that the Glivenko-Cantelli theorem can also be formulated for cumulative intensities (see Lemma B.1.2 from Appendix B.1). This already implies that - for fixed $\theta \in \Theta$ - the average cumulative intensity $\overline{\Lambda}_{\theta^*}^{(n)}$ (i.e., the integrator of Φ_n) $\mathbb{P}$-almost surely converges uniformly on $\mathcal{I}$ to $\mathbb{E}\Lambda_{\theta^*}$ (i.e., the integrator of Φ_0). This easily extends to hold uniformly on $\mathcal{I} \times K$ by virtue of ($\widetilde{\mathrm{A}}3$). To this end, we note that the cumulative intensities $\Lambda_\theta^{(i)}$ are i.i.d. copies of Λ_θ, and use the mean value theorem to infer for any $\theta, \theta' \in \mathrm{B}_\varepsilon(\theta^*)$ and $t \in \mathcal{I}$:

$$\begin{aligned}\left|\overline{\Lambda}_\theta^{(n)}(t) - \overline{\Lambda}_{\theta'}^{(n)}(t)\right| &\le \frac{1}{n}\sum_{i=1}^n \left|\Lambda_\theta^{(i)}(t) - \Lambda_{\theta'}^{(i)}(t)\right| \\ &\le \frac{1}{n}\sum_{i=1}^n \left(\sup_{\theta \in \mathrm{B}_\varepsilon(\theta^*)} \underbrace{\left\|\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta^{(i)}(t)\right\|}_{\le\sqrt{d}C}\right)\|\theta - \theta'\| \\ &\le \sqrt{d}C\,\|\theta - \theta'\| \qquad \mathbb{P}\text{-almost surely.}\end{aligned} \tag{3.25}$$

For any fixed $\theta' \in \mathrm{B}_\varepsilon(\theta^*)$, we can thus compute:

$$\sup_{(t,\theta)\in\mathcal{I}\times\mathrm{B}_\delta(\theta')} \left|\overline{\Lambda}_\theta^{(n)}(t) - \overline{\Lambda}_{\theta'}^{(n)}(t)\right| \leq 2\sqrt{d}C\delta\,. \tag{3.26}$$

As per Remark 3.6, the continuity of the cumulative intensity carries over to its expectation. On the compact set $\mathcal{I}\times K$, this implies uniform continuity of the expected cumulative intensity. For an arbitrary $\nu > 0$, we can then find $0 < \delta < \frac{\nu}{2\sqrt{d}C}$, such that

$$\sup_{(t,\theta)\in\mathcal{I}\times\mathrm{B}_\delta(\theta')} \left|\mathbb{E}\Lambda_\theta(t) - \mathbb{E}\Lambda_{\theta'}(t)\right| < \nu\,, \tag{3.27}$$

where the upper bound on δ secures that the right-hand side of Equation (3.26) is smaller than ν. We proceed to prove the $\mathbb{P}$-almost sure uniform convergence on $\mathcal{I}\times\mathrm{B}_\delta(\theta')$, and for this decompose the associated event (without lim sup) using the triangle inequality:

$$\begin{aligned}
&\left\{\sup_{(t,\theta)\in\mathcal{I}\times\mathrm{B}_\delta(\theta')} \left|\overline{\Lambda}_\theta^{(n)}(t) - \mathbb{E}\Lambda_\theta(t)\right| > 3\nu\right\}\\
&\quad\subset \left\{\sup_{(t,\theta)\in\mathcal{I}\times\mathrm{B}_\delta(\theta')} \left|\overline{\Lambda}_\theta^{(n)}(t) - \overline{\Lambda}_{\theta'}^{(n)}(t)\right| > \nu\right\} \cup \left\{\sup_{(t,\theta)\in\mathcal{I}\times\mathrm{B}_\delta(\theta')} \underbrace{\left|\overline{\Lambda}_{\theta'}^{(n)}(t) - \mathbb{E}\Lambda_{\theta'}(t)\right|}_{\text{does not depend on }\theta} > \nu\right\}\\
&\qquad\cup \left\{\sup_{(t,\theta)\in\mathcal{I}\times\mathrm{B}_\delta(\theta')} \left|\mathbb{E}\Lambda_\theta(t) - \mathbb{E}\Lambda_{\theta'}(t)\right| > \nu\right\}.
\end{aligned} \tag{3.28}$$

By the choice of δ, the first and last set of Equation (3.28) are empty. We obtain:

$$\left\{\sup_{(t,\theta)\in\mathcal{I}\times\mathrm{B}_\delta(\theta')} \left|\overline{\Lambda}_\theta^{(n)}(t) - \mathbb{E}\Lambda_\theta(t)\right| > 3\nu\right\} \subset \left\{\sup_{t\in\mathcal{I}} \left|\overline{\Lambda}_{\theta'}^{(n)}(t) - \mathbb{E}\Lambda_{\theta'}(t)\right| > \nu\right\}.$$

As $\bigcup_{\theta'\in K}\mathrm{B}_\delta(\theta')$ is an open cover of the compact set K, there exists a finite subcover $K \subset \mathrm{B}_\delta(\theta^{(1)}) \cup \ldots \cup \mathrm{B}_\delta(\theta^{(L)})$. We then observe:

$$\begin{aligned}
\left\{\sup_{(t,\theta)\in\mathcal{I}\times K} \left|\overline{\Lambda}_\theta^{(n)}(t) - \mathbb{E}\Lambda_\theta(t)\right| > 3\nu\right\} &= \bigcup_{l=1}^{L}\left\{\sup_{(t,\theta)\in\mathcal{I}\times\mathrm{B}_\delta(\theta^{(l)})} \left|\overline{\Lambda}_\theta^{(n)}(t) - \mathbb{E}\Lambda_\theta(t)\right| > 3\nu\right\}\\
&\subset \bigcup_{l=1}^{L}\left\{\sup_{t\in\mathcal{I}} \left|\overline{\Lambda}_{\theta^{(l)}}^{(n)}(t) - \mathbb{E}\Lambda_{\theta^{(l)}}(t)\right| > \nu\right\}.
\end{aligned}$$

In this way, the $\mathbb{P}$-almost sure convergence on the sets $\mathcal{I}\times\mathrm{B}_\delta(\theta')$ directly expands to $\mathcal{I}\times K$, as combining Lemma 4.18 with the Glivenko-Cantelli theorem for cumulative intensities gives the desired result:

$$\begin{aligned}
&\mathbb{P}\left(\limsup_{n\to\infty}\left\{\sup_{(t,\theta)\in\mathcal{I}\times K} \left|\overline{\Lambda}_\theta^{(n)}(t) - \mathbb{E}\Lambda_\theta(t)\right| > 3\nu\right\}\right)\\
&\quad\leq \mathbb{P}\left(\limsup_{n\to\infty}\bigcup_{l=1}^{L}\left\{\sup_{t\in\mathcal{I}} \left|\overline{\Lambda}_{\theta^{(l)}}^{(n)}(t) - \mathbb{E}\Lambda_{\theta^{(l)}}(t)\right| > \nu\right\}\right) = 0\,.
\end{aligned}$$

$\square$

The result of Lemma 3.15 still applies as soon as derivatives with respect to θ are involved, but at the cost of receiving only the weaker stochastic convergence. We first point out that although the Equations (3.26) and (3.27) remain valid even if partial derivatives with respect to θ are considered (in the first case we need condition ($\widetilde{\text{A}}3$), in the second case condition (A2)), unlike in the above reasoning we cannot resort to the Glivenko-Cantelli theorem because the required monotonicity is missing. We therefore have to localize with respect to t. This is exactly the point where the almost sure convergence gets lost, because here increments in t have to be considered and cannot be dismissed for monotonicity reasons.

Lemma 3.16 (Uniform Convergence of Averages of Derivative Cumulative Intensities)**.** *Let K be an arbitrary compact subset of $B_\varepsilon(\theta^*)$, where again the conditions from Assumption 3.8 are assumed to hold. Then, for $p \in \{1, 2\}$,*

$$\frac{\mathrm{d}^p}{\mathrm{d}\theta^p}\overline{\Lambda}_\theta^{(n)} \xrightarrow{\mathbb{P}} \mathbb{E}\frac{\mathrm{d}^p}{\mathrm{d}\theta^p}\Lambda_\theta \quad \textit{uniformly on } \mathcal{I} \times K \textit{ as } n \to \infty.$$

Proof. Instead of the Lipschitz continuity shown in Equation (3.26), we will obtain the stochastic uniform equicontinuity of the aggregate cumulative intensity. For this, we choose arbitrary $\nu > 0$ and $\delta > 0$. Since the cumulative intensities $\Lambda_\theta^{(i)}$ are i.i.d., so are their derivatives with respect to θ, and the Markov inequality implies for all $n \in \mathbb{N}$:

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\|(t,\theta)-(t',\theta')\|<\delta}\left\|\frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)}(t) - \frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_{\theta'}^{(n)}(t')\right\| > \nu\right) \\
&\quad\le \frac{1}{\nu}\mathbb{E}\left(\sup_{\|(t,\theta)-(t',\theta')\|<\delta}\left\|\frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)}(t) - \frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_{\theta'}^{(n)}(t')\right\|\right) \\
&\quad\le \frac{1}{\nu}\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\sup_{\|(t,\theta)-(t',\theta')\|<\delta}\left\|\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta^{(i)}(t) - \frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta'}^{(i)}(t')\right\|\right) \\
&\quad= \frac{1}{\nu}\mathbb{E}\left(\sup_{\|(t,\theta)-(t',\theta')\|<\delta}\left\|\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta(t) - \frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta'}(t')\right\|\right),
\end{aligned} \tag{3.29}$$

where the supremum includes only those $(t,\theta), (t',\theta') \in \mathcal{I} \times K$. By condition (A2), the partial derivatives of the cumulative intensity are continuous on $\mathcal{I} \times \Theta$ and hence uniformly continuous on the compact set $\mathcal{I} \times K$. This transfers to the total differential. Thus, with probability 1,

$$\sup_{\|(t,\theta)-(t',\theta')\|<\delta}\left\|\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta(t) - \frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta'}(t')\right\| \to 0 \quad (\delta \to 0),$$

and by the dominated convergence theorem (one uses the bounds offered by condition ($\widetilde{\text{A}}3$)) the convergence of the expectation in Equation (3.29) ensues. Consequently, for every $\eta > 0$ there exists a $\delta > 0$ such that

$$\mathbb{P}\left(\sup_{\|(t,\theta)-(t',\theta')\|<\delta}\left\|\frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)}(t) - \frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_{\theta'}^{(n)}(t')\right\| > \nu\right) < \eta, \quad \text{for all } n \in \mathbb{N}. \tag{3.30}$$

We briefly note that technically we just established the tightness of the differentiated process (see Corollary B.2.5 in Appendix B.2). This comes as no surprise, since tightness is essentially characterized by stochastic uniform equicontinuity, and it therefore serves as the stochastic stand-in for the almost sure uniform convergence in the first part of the proof. More details on the indicated relationships are available in the dedicated article Newey 1991, from which the following proof step draws its inspiration (see Theorem 2.1, Newey 1991, p. 1162). We can again cover the compact set $\mathcal{I} \times K$ with open balls of radius δ and find a finite subcover,

$$\mathcal{I} \times K \subset \bigcup_{l=1}^{L} \underbrace{\mathrm{B}_\delta((t^{(l)}, \theta^{(l)}))}_{=:\mathrm{B}_\delta^{(l)}} .$$

By the triangle inequality, we have:

$$\begin{aligned}
&\sup_{(t,\theta)\in \mathrm{B}_\delta^{(l)}} \left\| \frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)}(t) - \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta(t) \right\| \\
&= \sup_{(t,\theta)\in \mathrm{B}_\delta^{(l)}} \left\| \frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)}(t) - \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta(t) - \left(\frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_{\theta^{(l)}}^{(n)}(t^{(l)}) - \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta^{(l)}}(t^{(l)}) \right) \right. \\
&\qquad\qquad \left. + \left(\frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_{\theta^{(l)}}^{(n)}(t^{(l)}) - \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta^{(l)}}(t^{(l)}) \right) \right\| \\
&\leq \sup_{(t,\theta)\in \mathrm{B}_\delta^{(l)}} \left\| \frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)}(t) - \frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_{\theta^{(l)}}^{(n)}(t^{(l)}) \right\| + \sup_{(t,\theta)\in \mathrm{B}_\delta^{(l)}} \left\| \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta(t) - \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta^{(l)}}(t^{(l)}) \right\| \\
&\qquad\qquad + \left\| \frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_{\theta^{(l)}}^{(n)}(t^{(l)}) - \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta^{(l)}}(t^{(l)}) \right\| .
\end{aligned}$$

We can thus compute:

$$\begin{aligned}
&\mathbb{P}\left(\sup_{(t,\theta)\in\mathcal{I}\times K} \left\| \frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)}(t) - \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta(t) \right\| > 3\nu \right) \\
&= \mathbb{P}\left(\max_{l=1,\ldots,L} \sup_{(t,\theta)\in \mathrm{B}_\delta^{(l)}} \left\| \frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)}(t) - \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta(t) \right\| > 3\nu \right) \\
&\leq \mathbb{P}\left(\max_{l=1,\ldots,L} \sup_{(t,\theta)\in \mathrm{B}_\delta^{(l)}} \left\| \frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)}(t) - \frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_{\theta^{(l)}}^{(n)}(t^{(l)}) \right\| > \nu \right) \\
&\quad + \mathbb{P}\left(\max_{l=1,\ldots,L} \sup_{(t,\theta)\in \mathrm{B}_\delta^{(l)}} \left\| \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta(t) - \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta^{(l)}}(t^{(l)}) \right\| > \nu \right) \\
&\quad + \mathbb{P}\left(\max_{l=1,\ldots,L} \left\| \frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_{\theta^{(l)}}^{(n)}(t^{(l)}) - \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta^{(l)}}(t^{(l)}) \right\| > \nu \right) . \qquad (3.31)
\end{aligned}$$

We deal with the summands of Equation (3.31) individually. For the first one, note that

$$\max_{l=1,\ldots,L} \sup_{(t,\theta)\in \mathrm{B}_\delta^{(l)}} \left\| \frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)}(t) - \frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_{\theta^{(l)}}^{(n)}(t^{(l)}) \right\| \leq \sup_{\|(t,\theta)-(t',\theta')\|<\delta} \left\| \frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)}(t) - \frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_{\theta'}^{(n)}(t') \right\| ,$$

and therefore the probability of the associated event is bounded by η according to Equation

(3.30). The event belonging to the second summand is deterministic, so its probability is either 0 or 1 depending on δ. Since the uniform continuity of the differentiated cumulative intensity transfers to its expectation, the corresponding probability once again vanishes as $\delta \to 0$. Consequently, and analogous to the first part of the proof, we only need to choose δ sufficiently small in advance. The remaining summand is subject to the law of large numbers (and, technically, another application of Lemma 4.18). However, the weak version of this law already suffices to infer convergence of the last summand to 0 for $n \to \infty$ (remarkably, this is the only term that depends on n in a significant way). Hence, there exists $n_0 \in \mathbb{N}$ such that

$$\mathbb{P}\left(\sup_{(t,\theta)\in\mathcal{I}\times K}\left\|\frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)}(t) - \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta(t)\right\| > 3\nu\right) < \eta\,, \quad \text{for all } n \geq n_0, \tag{3.32}$$

and this is stochastic uniform convergence on $\mathcal{I} \times K$. This proof can be transferred one to one to derivatives of higher order, as long as conditions (A2) and ($\widetilde{\text{A}}$3) are still in force. This is especially true for the second total derivative with respect to θ, where matrix norms must be considered, but nothing else changes. □

We now return to the proof of Proposition 3.13.

Proof of Proposition 3.13. From Lemma 3.15 we obtain that the integrator of $\Phi_n(\theta)$ converges uniformly[9] on $\mathcal{I} \times K$ to the integrator of $\Phi_0(\theta)$ with probability 1. We can conclude by combining Lemma 3.15 and Lemma 3.16 that the integrand of $\Phi_n(\theta)$ also converges uniformly on $\mathcal{I} \times K$ (albeit only in probability) by a Slutzky-type argument. Therefore, in the following we deal with integrals of the shape

$$\int_{\mathcal{I}} X_\theta^{(n)}\,\mathrm{d}Y^{(n)}\,,$$

where

(a) $\left(X_\cdot^{(n)}\right)_{n\in\mathbb{N}}$ is a sequence of matrices whose entries are continuous stochastic processes with index set $\mathcal{I} \times K$. Furthermore, there exists a deterministic matrix-valued mapping $X_\cdot$ which is continuous on $\mathcal{I} \times K$ such that $X_\cdot^{(n)} \xrightarrow{\mathbb{P}} X_\cdot$ uniformly on $\mathcal{I} \times K$.

(b) $\left(Y^{(n)}\right)_{n\in\mathbb{N}}$ is a sequence of continuous stochastic processes with index set $\mathcal{I}$ and $Y^{(n)}(0) = 0$ ($n \in \mathbb{N}$) that are $\mathbb{P}$-almost surely non-decreasing and uniformly bounded by some constant[10] $C > 0$. Additionally, there exists a deterministic continuous function Y, such that $Y^{(n)} \to Y$ uniformly on $\mathcal{I}$ with probability 1. In particular, Y satisfies $Y(0) = 0$, is non-decreasing and uniformly bounded by C as well.

In this setting, we want to prove the following convergence:

$$\sup_{\theta\in K}\left\|\int_{\mathcal{I}} X_\theta^{(n)}\,\mathrm{d}Y^{(n)} - \int_{\mathcal{I}} X_\theta\,\mathrm{d}Y\right\| \xrightarrow{\mathbb{P}} 0 \quad (n \to \infty)\,. \tag{3.33}$$

[9] For the integrator, $\theta^* \in K$ is fixed, so Lemma B.1.2 from Appendix B.1 could also be applied. But since the averaged cumulative intensity also appears in the integrand, we still need the statement of Lemma 3.15, even though convergence in probability would suffice here.

[10] As such, they can be understood as generalized distribution functions.

We can split the occurring difference in Equation (3.33) to receive:

$$\int_{\mathcal{I}} X_\theta^{(n)} \,\mathrm{d}Y^{(n)} - \int_{\mathcal{I}} X_\theta \,\mathrm{d}Y = \int_I \left(X_\theta^{(n)} - X_\theta\right) \,\mathrm{d}Y^{(n)} + \left(\int_I X_\theta \,\mathrm{d}Y^{(n)} - \int_I X_\theta \,\mathrm{d}Y\right) ,$$

and because of the triangle inequality we can treat the two resulting differences separately. Since we want to proof convergence in probability, we choose an arbitrary $\nu > 0$. For the first difference, note that $\mu_{Y^{(n)}}(\mathcal{I}) = Y^{(n)}(\tau) - Y^{(n)}(0)$ and $Y^{(n)}(0) = 0$, so we can exploit the boundedness of $Y^{(n)}$ to get:

$$\begin{aligned}
\mathbb{P}\left(\sup_{\theta\in K}\left\|\int_I \left(X_\theta^{(n)} - X_\theta\right) \,\mathrm{d}Y^{(n)}\right\| > \nu\right) &\leq \mathbb{P}\left(\sup_{(t,\theta)\in\mathcal{I}\times K}\left\|X_\theta^{(n)}(t) - X_\theta(t)\right\| \cdot \underbrace{Y^{(n)}(\tau)}_{\leq C} > \nu\right) \\
&\leq \mathbb{P}\left(\sup_{(t,\theta)\in\mathcal{I}\times K}\left\|X_\theta^{(n)}(t) - X_\theta(t)\right\| > \frac{\nu}{C}\right) \\
&\longrightarrow 0 \quad (n\to\infty) ,
\end{aligned}$$

since $X_\cdot^{(n)} \xrightarrow{\mathbb{P}} X_\cdot$ uniformly on $\mathcal{I}\times K$. Hence, the first difference convergences uniformly on K in probability. For the second difference, the $\mathbb{P}$-almost sure uniform convergence of the integrator implies the weak convergence of the induced probability measures on $\mathcal{B}(\mathcal{I})$. By Helly-Bray's theorem[11], this implies that with probability 1 holds for all continuous bounded functions h:

$$\int_{\mathcal{I}} h \,\mathrm{d}Y^{(n)} \to \int_{\mathcal{I}} h \,\mathrm{d}Y . \tag{3.34}$$

Specifically, this convergence holds for $h = X_\theta$, $\theta\in K$, since $X_\cdot$ is continuous and thus bounded on the compact set $\mathcal{I}\times K$. However, this equates only to $\mathbb{P}$-almost sure pointwise convergence on K of the integrals

$$Z^{(n)}(\theta) := \int_{\mathcal{I}} X_\theta \,\mathrm{d}Y^{(n)} .$$

In order to infer uniform convergence from pointwise convergence, we need to prove that the sequence $\left(Z^{(n)}\right)_{n\in\mathbb{N}}$ is equicontinuous. For this, recall that $X_\cdot$ is by assumption continuous on the compact set $\mathcal{I}\times K$ and hence uniformly continuous. Moreover, $X_\cdot$ is deterministic. For any given $\eta > 0$, there then exists $\delta > 0$ such that

$$\sup_{\|(t,\theta)-(t',\theta')\|<\delta} \|X_\theta(t) - X_{\theta'}(t')\| < \frac{\eta}{C} ,$$

where the constant C from the above calculations has been included for aesthetic reasons. We can utilize that $Z^{(n)}(\theta)$ depends on n only by virtue of the uniformly bounded integrator $Y^{(n)}$, and conclude for any $\theta, \theta' \in K$ with $\|(0, \theta-\theta')\| < \delta$ (remember that $\|(0,\cdot)\|$ defines a norm on K that is equivalent to any other predefined norm):

$$\begin{aligned}
\left\|Z^{(n)}(\theta) - Z^{(n)}(\theta')\right\| &= \left\|\int_{\mathcal{I}} (X_\theta - X_{\theta'}) \,\mathrm{d}Y^{(n)}\right\| \\
&\leq \sup_{t\in\mathcal{I}} \|X_\theta(t) - X_{\theta'}(t)\| \cdot Y^{(n)}(\tau)
\end{aligned}$$

[11] Occasionally, Equation (3.34) is used to define weak convergence in the first place. See Definition 13.12 and Theorem 13.23 of Klenke 2020, p. 286 for reference.

$$\leq \sup_{\|(t,\theta)-(t',\theta')\|<\delta} \left\| X_\theta(t) - X_{\theta'}(t') \right\| \cdot C < \eta \,,$$

and this is the desired equicontinuity. Therefore, the second difference converges $\mathbb{P}$-almost surely uniformly on K, which implies convergence in probability. Combining our findings, the convergence in Equation (3.33) then follows. □

Having dealt with the asymptotic behavior of the parametric process Φ_n, we now transfer our results to the related process Ψ_n. As explained earlier, $\Psi_n(\hat{\theta}_n)$ can be considered the average of the parametric process Φ_n along the line segment adjoining θ^* and $\hat{\theta}_n$ in $\mathbb{R}^d$. This observation forms the basis for the following corollary.

Corollary 3.17 (Convergence of $\Psi_n(\hat{\theta}_n)$)**.**
In the situation of Proposition 3.13, it holds:

$$\Psi_n(\hat{\theta}_n) \xrightarrow{\mathbb{P}} \Phi_0\left(\theta^*\right) \quad \text{as } n \to \infty.$$

Proof. We first note that the function Φ_0 defined on $\mathrm{B}_\varepsilon\left(\theta^*\right)$ is deterministic. By conditions (C2) and (A2) as well as Remark 3.6 (which extends to derivatives with respect to θ), Φ_0 is continuous: sequential continuity can be directly proved by interchanging integration and limit (again by the dominated convergence theorem applicable due to conditions (C3) and $(\widetilde{\mathrm{A}}3)$). For any $\nu > 0$, there thus exists $0 < \delta < \varepsilon$ (so that $\overline{\mathrm{B}_\delta\left(\theta^*\right)} \subset \mathrm{B}_\varepsilon\left(\theta^*\right)$), for which

$$\|\theta - \theta^*\| < \delta \quad \Longrightarrow \quad \|\Phi_0(\theta) - \Phi_0(\theta^*)\| < \nu \,,$$

and hence

$$\sup_{\theta \in \mathrm{B}_\delta(\theta^*)} \|\Phi_0(\theta) - \Phi_0(\theta^*)\| \leq \nu \,. \tag{3.35}$$

Now let $\eta > 0$ be given. The $\mathbb{P}$-almost sure convergence of $\left(\hat{\theta}_n\right)_{n\in\mathbb{N}}$ to θ^* implies that there exists $n_1 \in \mathbb{N}$ with

$$\mathbb{P}\left(\|\hat{\theta}_n - \theta^*\| < \delta \text{ for all } n \geq n_1\right) \geq 1 - \frac{\eta}{2} \,.$$

Therefore, in the above event, the condition $\hat{\theta}_n \in \mathrm{B}_\delta\left(\theta^*\right)$ holds simultaneously for all $n \geq n_1$. In particular, for each individual $n \geq n_1$ we obtain as well:

$$\mathbb{P}\left(\hat{\theta}_n \in \mathrm{B}_\delta\left(\theta^*\right)\right) \geq 1 - \frac{\eta}{2} \,. \tag{3.36}$$

By Proposition 3.13 (setting $K := \overline{\mathrm{B}_\delta\left(\theta^*\right)}$), there further exists $n_2 \in \mathbb{N}$ such that

$$\mathbb{P}\left(\sup_{\theta\in K} \|\Phi_n(\theta) - \Phi_0(\theta)\| > \nu\right) < \frac{\eta}{2} \quad \text{for all } n \geq n_2. \tag{3.37}$$

Combining the above equations, we then obtain for each $n \geq \max\{n_1, n_2\}$:

$$\begin{aligned}
&\mathbb{P}\left(\|\Psi_n(\hat{\theta}_n) - \Phi_0(\theta^*)\| > 2\nu\right) \\
&\quad = \mathbb{P}\left(\left\|\int_0^1 \Phi_n(\theta^* + s(\hat{\theta}_n - \theta^*))\,\mathrm{d}s - \Phi_0(\theta^*)\right\| > 2\nu\right) \\
&\quad = \mathbb{P}\left(\left\|\int_0^1 \left(\Phi_n(\theta^* + s(\hat{\theta}_n - \theta^*)) - \Phi_0(\theta^*)\right)\mathrm{d}s\right\| > 2\nu\right)
\end{aligned}$$

$$
\begin{aligned}
&\leq \mathbb{P}\left(\int_0^1 \left\|\Phi_n(\theta^* + s(\hat{\theta}_n - \theta^*)) - \Phi_0(\theta^*)\right\| \,\mathrm{d}s > 2\nu\right) \\
&\leq \mathbb{P}\left(\left\{\int_0^1 \left\| \Phi_n(\underbrace{\theta^* + s(\hat{\theta}_n - \theta^*)}_{\in \mathrm{B}_\delta(\theta^*)}) - \Phi_0(\theta^*)\right\| \,\mathrm{d}s > 2\nu\right\} \cap \left\{\hat{\theta}_n \in \mathrm{B}_\delta(\theta^*)\right\}\right) \\
&\qquad + \underbrace{\mathbb{P}\left(\hat{\theta}_n \notin \mathrm{B}_\delta(\theta^*)\right)}_{\leq \frac{\eta}{2} \text{ by Eq. (3.36)}} \\
&\leq \mathbb{P}\left(\sup_{\theta \in \mathrm{B}_\delta(\theta^*)} \|\Phi_n(\theta) - \Phi_0(\theta^*)\| > 2\nu\right) + \frac{\eta}{2} \\
&\leq \mathbb{P}\left(\sup_{\theta \in \mathrm{B}_\delta(\theta^*)} (\|\Phi_n(\theta) - \Phi_0(\theta)\| + \|\Phi_0(\theta) - \Phi_0(\theta^*)\|) > 2\nu\right) + \frac{\eta}{2} \\
&\leq \underbrace{\mathbb{P}\left(\sup_{\theta \in K} \|\Phi_n(\theta) - \Phi_0(\theta)\| > \nu\right)}_{\leq \frac{\eta}{2} \text{ by Eq. (3.37)}} + \underbrace{\mathbb{P}\left(\sup_{\theta \in \mathrm{B}_\delta(\theta^*)} \|\Phi_0(\theta) - \Phi_0(\theta^*)\| > \nu\right)}_{=0 \text{ by Eq. (3.35)}} + \frac{\eta}{2} \\
&\leq \eta\,.
\end{aligned}
$$

Since $\nu > 0$ and $\eta > 0$ were chosen arbitrarily, the assertion then follows. □

To conclude the discussion of the standardizing matrix and its asymptotics, we show how the positive definiteness of $\Phi_0(\theta^*)$ can be derived from condition (A4) of Assumption 3.8.

Lemma 3.18 (Positive Definiteness of the Standardizing Matrix $\Phi_0(\theta^*)$; cf. Kopperschmidt 2005, p. 84)**.**
Under Assumptions 3.8, the asymptotic standardizing matrix $\Phi_0(\theta^*)$ *is positive definite.*

Proof. At the true parameter θ^*, the standardizing matrix from Equation (3.24) simplifies to

$$
\Phi_0(\theta^*) = \int_{\mathcal{I}} \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta^*}^\top \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta^*} \,\mathrm{d}\mathbb{E}\Lambda_{\theta^*}\,. \tag{3.38}
$$

The integrand is positive semidefinite as a dyadic product of the d-dimensional vector $\mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta^*}^\top$ with itself, and due to the monotonicity of the integral, the semidefiniteness carries over to $\Phi_0(\theta^*)$. Now let $v \in \mathbb{R}^d \setminus \{0\}$ be arbitrary and $B_v \subset \mathbb{R}$ as in condition (A4). As a finite Borel measure, $\mu_{\mathbb{E}\Lambda_{\theta^*}}$ is inner regular (see Lemma 26.2, Bauer 2001, p. 158). Consequently, there exists a compact set $K_v \subset B_v$ with $\mu_{\mathbb{E}\Lambda_{\theta^*}}(K_v) > 0$. Since the function

$$
t \mapsto v^\top \left[\mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta^*}(t)^\top \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta^*}(t)\right] v
$$

is continuous on $\mathcal{I}$, it attains a minimum on $K_v \subset \mathcal{I}$. The linearity of the integral yields:

$$
\begin{aligned}
v^\top \Phi_0(\theta^*) v &= \int_{\mathcal{I}} \underbrace{v^\top \left[\mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta^*}^\top \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta^*}\right] v}_{=\left(v^\top \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta^*}^\top\right)^2 \geq 0} \,\mathrm{d}\mathbb{E}\Lambda_{\theta^*} \\
&\geq \int_{K_v} v^\top \left[\mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta^*}^\top \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta^*}\right] v \,\mathrm{d}\mathbb{E}\Lambda_{\theta^*}
\end{aligned}
$$

$$\geq \min_{t\in K_v} \left\{ v^\top \left[\mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta^*}(t)^\top \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta^*}(t) \right] v \right\} \cdot \mu_{\mathbb{E}\Lambda_{\theta^*}}(K_v) > 0\,,$$

and hence positive definiteness of $\Phi_0(\theta^*)$ follows, because v was chosen arbitrarily. □

3.3.2. Asymptotics of the Auxiliary Parametric Processes α_n, β_n and γ_n

Having dealt with the asymptotics of the standardizing matrix, we now proceed to steps (i) through (iii) of our previous proof sketch. The first step is identical for all the auxiliary processes; for the second step, we also use essentially the same technique adapted to each process. It is only in the third step that the methods differ more significantly, and we consider the processes in ascending complexity of their leading terms: first α_n (where the leading term is 0), then β_n (where the leading term is asymptotically negligible), and finally γ_n (where the leading term determines the overall asymptotic distribution). Beforehand, we give a fairly simple lemma that extends the earlier Corollary 2.13 and makes the following proofs more convenient.

Lemma 3.19 (Exploiting Martingale Bounds for Integration)**.**
In the situation of Corollary 2.13, let $f = (f_t)_{t\in\mathcal{I}}$ be an $\{\mathcal{F}_t\}_{t\in\mathcal{I}}$-adapted stochastic process which is integrable with respect to the martingale M. Then,

$$\left\| \int_{\mathcal{I}} f_t \, dM_t \right\| \leq (N_\tau + \Lambda_\tau) \cdot \sup_{t\in\mathcal{I}} \|f_t\| \,.$$

Proof. We use the decomposition $M_t = N_t - \Lambda_t$, so the triangle inequality directly yields:

$$\begin{aligned}\left\| \int_{\mathcal{I}} f_t \, \mathrm{d}M_t \right\| &= \left\| \int_{\mathcal{I}} f_t \, \mathrm{d}N_t - \int_{\mathcal{I}} f_t \, \mathrm{d}\Lambda_t \right\| \leq \left\| \int_{\mathcal{I}} f_t \, \mathrm{d}N_t \right\| + \left\| \int_{\mathcal{I}} f_t \, \mathrm{d}\Lambda_t \right\| \\ &\leq \int_{\mathcal{I}} \|f_t\| \, \mathrm{d}N_t + \int_{\mathcal{I}} \|f_t\| \, \mathrm{d}\Lambda_t \leq (N_\tau + \Lambda_\tau) \cdot \sup_{t\in\mathcal{I}} \|f_t\| \,,\end{aligned}$$

which is the assertion. □

In our case, the martingale M in Lemma 3.19 is of bounded variation as the difference of a counting process N and its compensator Λ (the class of functions of bounded variation is characterized as the difference of two increasing functions, see Szőkefalvi-Nagy 1965, pp. 93–95). The requirement of integrability with respect to M can therefore be reduced to the existence of the Lebesgue-Stieltjes integrals with respect to N and Λ. Moreover, the stochastic process f here is always the product of (derivative) cumulative intensities and the martingale M, which again can be decomposed into a cumulative intensity (the compensator) and a step function (the counting process). As a result, we never have to be concerned about the existence of the integrals involved.

Representation Theorem for the Process α_n

We study the process α_n to familiarize ourselves with the methods also used for β_n and γ_n. By Definition 3.11, α_n is the parametric process with index set $\mathrm{B}_\varepsilon(\theta^*)$ given via

$$\alpha_n(\theta) = \sqrt{n} \int_{\mathcal{I}} \left[\overline{\Lambda}_\theta^{(n)} - \overline{\Lambda}_{\theta^*}^{(n)} \right] \frac{\mathrm{d}}{\mathrm{d}\theta} \overline{\Lambda}_\theta^{(n)} \, \mathrm{d}\overline{M}^{(n)} \,.$$

Theorem 3.20 (Representation Theorem for the Auxiliary Process α_n)**.**
Under Assumption 3.8, the process α_n evaluated at the MDE $\hat{\theta}_n$ admits the representation

$$\alpha_n(\hat{\theta}_n) = \sqrt{n} A_n (\hat{\theta}_n - \theta^*) \,, \qquad \textit{where } A_n \xrightarrow{\mathbb{P}} 0 \quad (n \to \infty). \tag{3.39}$$

Proof. We start by noting that the process disappears at the true parameter θ^*, that is, $\alpha_n(\theta^*) = 0$. An application of the mean value theorem for vector-valued functions then directly yields:

$$\begin{aligned}
\alpha_n(\hat{\theta}_n) &= \alpha_n(\theta^*) + \left[\int_0^1 D_\theta\, \alpha_n(\theta^* + s(\hat{\theta}_n - \theta^*))\, \mathrm{d}s\right] (\hat{\theta}_n - \theta^*) \\
&= \sqrt{n} \left[\int_0^1 D_\theta \int_{\mathcal{I}} \left(\left[\overline{\Lambda}_\theta^{(n)} - \overline{\Lambda}_{\theta^*}^{(n)}\right] \frac{\mathrm{d}}{\mathrm{d}\theta} \overline{\Lambda}_\theta^{(n)} \right) \mathrm{d}\overline{M}^{(n)} \Big|_{\theta = \theta^* + s\left(\hat{\theta}_n - \theta^*\right)} \mathrm{d}s\right] (\hat{\theta}_n - \theta^*) \,,
\end{aligned} \tag{3.40}$$

which already resembles Equation (3.39). Unfortunately, this matrix representation is barely suitable for further calculation. However, only the convergence of the matrix towards 0 needs to be shown, so that we can study the individual components of $\alpha_n(\hat{\theta}_n)$ to infer said convergence. Herein, the jth component of $\alpha_n(\theta)$ is given by:

$$\alpha_{n,j}(\theta) := \sqrt{n} \int_{\mathcal{I}} \left[\overline{\Lambda}_\theta^{(n)} - \overline{\Lambda}_{\theta^*}^{(n)}\right] \frac{\partial}{\partial \theta_j} \overline{\Lambda}_\theta^{(n)}\, \mathrm{d}\overline{M}^{(n)} \,, \quad j \in \{1, \ldots, d\} \,. \tag{3.41}$$

By applying the mean value theorem for functions of several variables and the Leibniz integral rule we obtain:

$$\alpha_{n,j}(\hat{\theta}_n) = \sqrt{n} \underbrace{\left[\int_{\mathcal{I}} D_\theta \left(\left[\overline{\Lambda}_\theta^{(n)} - \overline{\Lambda}_{\theta^*}^{(n)}\right] \frac{\partial}{\partial \theta_j} \overline{\Lambda}_\theta^{(n)} \right) \Big|_{\theta = \tilde{\theta}_{n,j}} \mathrm{d}\overline{M}^{(n)} \right]}_{=: D_\theta \tilde{\alpha}_{n,j}(\tilde{\theta}_{n,j})} (\hat{\theta}_n - \theta^*) \,,$$

for some $\tilde{\theta}_{n,j} \in \{\theta^* + s(\hat{\theta}_n - \theta^*) \,:\, s \in [0,1]\}$, i.e., on the line segment adjoining θ^* and $\hat{\theta}_n$ in $\mathbb{R}^d$. Note that $\tilde{\theta}_{n,j}$ is therefore itself a d-dimensional vector, with the subscripted j indicating that the particular intermediate point may differ depending on the component under consideration - which is why the matrix representation of Equation (3.40) became necessary in the first place. The $1 \times d$ matrix $D_\theta \tilde{\alpha}_{n,j}(\tilde{\theta}_{n,j})$ now contributes the jth row of the $d \times d$ matrix A_n, that is,

$$A_n := \begin{pmatrix} D_\theta \tilde{\alpha}_{n,1}(\tilde{\theta}_{n,1}) \\ \vdots \\ D_\theta \tilde{\alpha}_{n,d}(\tilde{\theta}_{n,d}) \end{pmatrix} .$$

For $A_n \xrightarrow{\mathbb{P}} 0$ to hold, it then suffices to show that all rows of A_n converge to 0 in probability. In the following, we thus study these rows in more detail. Note that by the differentiability of the parameter integral, the identity $\sqrt{n} D_\theta \tilde{\alpha}_{n,j}(\tilde{\theta}_{n,j}) = D_\theta \alpha_{n,j}(\tilde{\theta}_{n,j})$ holds, which justifies the suggestive choice of notion. We observe by performing similar steps as before:

$$D_\theta \tilde{\alpha}_{n,j}(\tilde{\theta}_{n,j}) = D_\theta \tilde{\alpha}_{n,j}(\theta^*) + \left[\int_0^1 D_\theta^2\, \tilde{\alpha}_{n,j}(\theta^* + s(\tilde{\theta}_{n,j} - \theta^*))\,\mathrm{d}s\right](\tilde{\theta}_{n,j} - \theta^*)\,. \tag{3.42}$$

The proof concludes by proving convergence to 0 in Equation (3.42). To do this, we examine both summands individually, starting with the second one. Since $\tilde{\theta}_{n,j}$ lies between θ^* and $\hat{\theta}_n$ and the MDE is strongly consistent, $\tilde{\theta}_{n,j}$ converges almost surely (and hence stochastically) to θ^* for $n \to \infty$. The second summand of Equation (3.42) therefore converges to 0 if the occurring integrand (and hence the entire integral) is almost surely bounded. Again, the integral expression could be avoided by looking at individual components, but here the matrix proves benign for our purposes. Once more interchanging integration and differentiation by means of the Leibniz integral rule, we obtain for each $\theta \in \mathrm{B}_\varepsilon(\theta^*)$:

$$\begin{aligned}
D_\theta^2\, \tilde{\alpha}_{n,j}(\theta) &= D_\theta \int_{\mathcal{I}} D_\theta \left(\left[\overline{\Lambda}_\theta^{(n)} - \overline{\Lambda}_{\theta^*}^{(n)}\right] \frac{\partial}{\partial\theta_j}\overline{\Lambda}_\theta^{(n)}\right) \mathrm{d}\overline{M}^{(n)} \\
&= \int_{\mathcal{I}} D_\theta^2 \left(\left[\overline{\Lambda}_\theta^{(n)} - \overline{\Lambda}_{\theta^*}^{(n)}\right] \frac{\partial}{\partial\theta_j}\overline{\Lambda}_\theta^{(n)}\right) \mathrm{d}\overline{M}^{(n)} \\
&= \int_{\mathcal{I}} D_\theta^2\, \overline{\Lambda}_\theta^{(n)} \frac{\partial}{\partial\theta_j}\overline{\Lambda}_\theta^{(n)} + \left(D_\theta\, \overline{\Lambda}_\theta^{(n)}\right)^\top D_\theta \frac{\partial}{\partial\theta_j}\overline{\Lambda}_\theta^{(n)} \\
&\quad + \left(D_\theta \frac{\partial}{\partial\theta_j}\overline{\Lambda}_\theta^{(n)}\right)^\top D_\theta\, \overline{\Lambda}_\theta^{(n)} + \left[\overline{\Lambda}_\theta^{(n)} - \overline{\Lambda}_{\theta^*}^{(n)}\right] D_\theta^2 \frac{\partial}{\partial\theta_j}\overline{\Lambda}_\theta^{(n)}\, \mathrm{d}\overline{M}^{(n)}\,.
\end{aligned} \tag{3.43}$$

Due to the boundedness condition (C3), it suffices according to Lemma 3.19 to find uniform bounds for the integrand of Equation (3.43). As the choice of matrix norm is irrelevant (cf. Remark 3.14), we can again opt for the max norm and study the components individually. The (k,l)th component of the matrix-valued integrand from Equation (3.43) is given by

$$\begin{aligned}
&\frac{\partial^2}{\partial\theta_k\partial\theta_l}\overline{\Lambda}_\theta^{(n)} \cdot \frac{\partial}{\partial\theta_j}\overline{\Lambda}_\theta^{(n)} + \frac{\partial}{\partial\theta_k}\overline{\Lambda}_\theta^{(n)} \cdot \frac{\partial^2}{\partial\theta_l\partial\theta_j}\overline{\Lambda}_\theta^{(n)} \\
&\quad + \frac{\partial^2}{\partial\theta_k\partial\theta_j}\overline{\Lambda}_\theta^{(n)} \cdot \frac{\partial}{\partial\theta_l}\overline{\Lambda}_\theta^{(n)} + \left[\overline{\Lambda}_\theta^{(n)} - \overline{\Lambda}_{\theta^*}^{(n)}\right] \cdot \frac{\partial^3}{\partial\theta_k\partial\theta_l\partial\theta_j}\overline{\Lambda}_\theta^{(n)}\,,
\end{aligned}$$

which is $\mathbb{P}$-almost surely uniformly bounded on $\mathcal{I} \times \mathrm{B}_\varepsilon(\theta^*)$ by virtue of conditions (C3) and ($\widetilde{\mathrm{A}}$3). Hence, as $\tilde{\theta}_{n,j} - \theta^* \to 0$ holds $\mathbb{P}$-almost surely, the same applies to the second summand of Equation (3.42), namely

$$\left[\int_0^1 D_\theta^2\, \tilde{\alpha}_{n,j}(\theta^* + s(\tilde{\theta}_{n,j} - \theta^*))\,\mathrm{d}s\right](\tilde{\theta}_{n,j} - \theta^*) \longrightarrow 0\,,$$

so Equation (3.42) can be restated as

$$D_\theta \tilde{\alpha}_{n,j}(\tilde{\theta}_{n,j}) = D_\theta \tilde{\alpha}_{n,j}(\theta^*) + o_{\mathbb{P}}(1)\,.$$

We turn to the remaining summand, which is now easily dealt with because the argument as the true parameter θ^* is no longer random or dependent on j or n. As before, differentiating under the integral sign yields by the product rule:

$$
\begin{aligned}
D_\theta \tilde{\alpha}_{n,j}\left(\theta^*\right) &= \int_{\mathcal{I}} D_\theta \left(\left[\overline{\Lambda}_\theta^{(n)} - \overline{\Lambda}_{\theta^*}^{(n)} \right] \frac{\partial}{\partial \theta_j} \overline{\Lambda}_\theta^{(n)} \right) \Big|_{\theta=\theta^*} \mathrm{d}\overline{M}^{(n)} \\
&= \int_{\mathcal{I}} \frac{\mathrm{d}}{\mathrm{d}\theta} \left[\overline{\Lambda}_\theta^{(n)} - \overline{\Lambda}_{\theta^*}^{(n)} \right] \Big|_{\theta=\theta^*} \frac{\partial}{\partial \theta_j} \overline{\Lambda}_{\theta^*}^{(n)} \,\mathrm{d}\overline{M}^{(n)} \\
&\quad + \int_{\mathcal{I}} \underbrace{\left[\overline{\Lambda}_{\theta^*}^{(n)} - \overline{\Lambda}_{\theta^*}^{(n)} \right]}_{=0} \frac{\mathrm{d}}{\mathrm{d}\theta} \frac{\partial}{\partial \theta_j} \overline{\Lambda}_\theta^{(n)} \Big|_{\theta=\theta^*} \mathrm{d}\overline{M}^{(n)} \\
&= \int_{\mathcal{I}} \frac{\mathrm{d}}{\mathrm{d}\theta} \overline{\Lambda}_{\theta^*}^{(n)} \frac{\partial}{\partial \theta_j} \overline{\Lambda}_{\theta^*}^{(n)} \,\mathrm{d}\overline{M}^{(n)} \,. \qquad (3.44)
\end{aligned}
$$

We finish the proof by showing that this integral also converges to 0 as $n \to \infty$. This is intuitively plausible, since the integrator is the mean $\overline{M}^{(n)}$ of the independent centered martingales $M^{(1)}, \ldots, M^{(n)}$, which converges $\mathbb{P}$-almost surely to 0 everywhere on $\mathcal{I}$ by the strong law of large numbers. However, this effect could be nullified as a result of stochastic dependencies. For this reason, we first ensured that the true parameter θ^* appears instead of the estimator $\hat{\theta}_n$ or some intermediate point. We now define

$$
U_{i_1,i_2,i_3} := \int_{\mathcal{I}} \frac{\mathrm{d}}{\mathrm{d}\theta} \Lambda_{\theta^*}^{(i_1)} \frac{\partial}{\partial \theta_j} \Lambda_{\theta^*}^{(i_2)} \,\mathrm{d}M^{(i_3)} \,, \qquad (3.45)
$$

so that Equation (3.44) can be written as

$$
\int_{\mathcal{I}} \frac{\mathrm{d}}{\mathrm{d}\theta} \overline{\Lambda}_{\theta^*}^{(n)} \frac{\partial}{\partial \theta_j} \overline{\Lambda}_{\theta^*}^{(n)} \,\mathrm{d}\overline{M}^{(n)} = \frac{1}{n^3} \sum_{i_1,i_2,i_3=1}^{n} U_{i_1,i_2,i_3} \,.
$$

The random vectors U_{i_1,i_2,i_3} are almost surely bounded and thus uniformly square-integrable. This is shown in a similar fashion as above, where starting from Equation (3.43) we used Lemma 3.19 and the boundedness conditions (C3) and ($\widetilde{\mathrm{A}}$3). Next, we note that the conditions (i) to (iv) of Lemma B.3.1 from Appendix B.3 are satisfied if we set $X = \frac{\mathrm{d}}{\mathrm{d}\theta} \Lambda_{\theta^*}^{(i_1)}$, $Y = \frac{\partial}{\partial \theta_j} \Lambda_{\theta^*}^{(i_2)}$ and $Z = M^{(i_3)}$ for pairwise distinct indices i_1, i_2, i_3:
The martingale $M^{(i_3)}$ can be decomposed into the counting process $Z^+ = N^{(i_3)}$ and its compensator $Z^- = \Lambda_{\theta^*}^{(i_3)}$, both of which are almost surely non-negative, right-continuous and non-decreasing (hence, condition (i) is fulfilled). Moreover, X and Y are both continuous by condition (A2), meaning that $\mathcal{J}_X = \emptyset = \mathcal{J}_Y$ (which in turn implies that both conditions (ii) and (iii) as well as Equation (B.35) hold). The existence of the occurring expectations (and thus the validity of condition (iv) and the remaining presumptions) follows as usual from conditions (C3) and ($\widetilde{\mathrm{A}}$3). Since the indices were chosen to be distinct, Z_t is independent of $\sigma\left(\{X_s, Y_s \,:\, s \in \mathcal{I}\,, s \leq t\}\right)$ for each $t \in \mathcal{I}$, so the application of Lemma B.3.1(ii) - which is due to Kopperschmidt 2005 - yields[12]:

$$
\mathbb{E}\left(U_{i_1,i_2,i_3}\right) = \mathbb{E}\left[\int_{\mathcal{I}} XY \,\mathrm{d}Z\right] = \mathbb{E}\left[\int_{\mathcal{I}} XY \,\mathrm{d}\mathbb{E}Z\right]
$$

[12]Lemma A.40 here provides the same result, because the occurring derivatives of the cumulative intensity Λ_θ are predictable as limits of predictable functions and the martingale X from Equation (A.39) is centred. However, we will frequently use Lemma B.3.1 in treating the other auxiliary processes, so we prefer the approach taken here for the sake of consistency.

$$= \mathbb{E}\left[\int_{\mathcal{I}} \frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta^*}^{(i_1)} \frac{\partial}{\partial\theta_j}\Lambda_{\theta^*}^{(i_2)}\, \mathrm{d}\mathbb{E}M^{(i_3)}\right] = 0$$

because of $\mathbb{E}M^{(i_3)} \equiv 0$, as martingales have constant expectation and $M_0^{(i_3)} = N_0^{(i_3)} - \Lambda_{\theta^*}^{(i_3)}(0) = 0$ with probability one (by construction, both a counting process and its compensator start in the origin at time 0). Additionally, U_{i_1,i_2,i_3} and U_{j_1,j_2,j_3} are stochastically indepedent if all the indices differ, resulting in

$$\mathbb{E}\left(U_{i_1,i_2,i_3}^{\top} U_{j_1,j_2,j_3}\right) = \mathbb{E}\left(U_{i_1,i_2,i_3}\right)^{\top} \mathbb{E}\left(U_{j_1,j_2,j_3}\right) = 0\,.$$

The L^2-convergence of the above integral to 0 therefore follows from Lemma B.4.7, which in turn implies stochastic convergence. Overall, we have:

$$D_\theta \tilde{\alpha}_{n,j}(\tilde{\theta}_{n,j}) = o_{\mathbb{P}}(1)\,, \quad j \in \{1, \ldots, d\}\,,$$

and hence the rows of A_n all converge to 0 in probability, thereby finishing the proof. □

Representation Theorem for the Process β_n

Next, we turn to the process β_n. While we can treat β_n largely analogously to α_n, here the leading term $\beta_n(\theta^*)$ does not vanish and must therefore be discussed separately. Recall that, by Definition 3.11, β_n is the parametric process with index set $\mathrm{B}_\varepsilon(\theta^*)$ given via

$$\beta_n(\theta) = \sqrt{n} \int_{\mathcal{I}} \overline{M}^{(n)} \frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)}\, \mathrm{d}\overline{M}^{(n)}\,.$$

Lemma 3.21 (Asymptotics of the Leading Term $\beta_n(\theta^*)$; cf. Lemma 12 of Kopperschmidt and Stute 2013, p. 1291)**.**
Under Assumption 3.8, for each $\theta \in B_\varepsilon(\theta^*)$ holds:

$$\beta_n(\theta) \xrightarrow{L^2} 0 \quad \text{as } n \to \infty.$$

In particular, $\beta_n(\theta^*) = o_{\mathbb{P}}(1)$, that is, $\beta_n(\theta^*) \xrightarrow{\mathbb{P}} 0$ as $n \to \infty$.

Proof. The proof is functionally identical to that given in Kopperschmidt and Stute 2013, p. 1292. We start by defining the d-variate random vectors

$$U_{pki} := \int_{\mathcal{I}} M^{(k)} \frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta^{(p)}\, \mathrm{d}M^{(i)}\,, \tag{3.46}$$

so that

$$\begin{aligned}
\beta_n(\theta) &= \sqrt{n} \int_{\mathcal{I}} \overline{M}^{(n)} \frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_\theta^{(n)}\, \mathrm{d}\overline{M}^{(n)} \\
&= \frac{\sqrt{n}}{n^3} \sum_{p,k,i=1}^{n} \int_{\mathcal{I}} M^{(k)} \frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta^{(p)}\, \mathrm{d}M^{(i)} \\
&= n^{-\frac{5}{2}} \sum_{p,k,i=1}^{n} U_{pki}\,,
\end{aligned}$$

and seek to apply Lemma B.4.6 later. Because the random vectors U_{pki} are almost surely bounded,

$$\begin{aligned}
\|U_{pki}\| &= \left\| \int_{\mathcal{I}} M^{(k)} \frac{\mathrm{d}}{\mathrm{d}\theta} \Lambda_\theta^{(p)} \, \mathrm{d}M^{(i)} \right\| && \Big| \text{apply Lemma 3.19} \\
&\leq \underbrace{\left(N_\tau^{(i)} + \Lambda_\theta^{(i)}(\tau) \right)}_{\leq 2C \text{ by (C3)}} \cdot \sup_{t \in \mathcal{I}} \left\| M_t^{(k)} \frac{\mathrm{d}}{\mathrm{d}\theta} \Lambda_\theta^{(p)}(t) \right\| \\
&\leq 2C \cdot \underbrace{\sup_{t \in \mathcal{I}} \left\| M_t^{(k)} \right\|}_{\leq 2C \text{ by Cor. 2.13}} \cdot \underbrace{\sup_{t \in \mathcal{I}} \left\| \frac{\mathrm{d}}{\mathrm{d}\theta} \Lambda_\theta^{(p)}(t) \right\|}_{\leq C \text{ by } (\widetilde{\mathrm{A}}3)} \leq 4C^3 < \infty \,, \qquad (3.47)
\end{aligned}$$

they are square-integrable. We can split the sum over the U_{pki} into a sub-sum where all indices are distinct and a sub-sum over partially matching indices. Since there are $\mathcal{O}(n^2)$ summands with partially matching indices, for the corresponding sub-sum we have:

$$\left\| n^{-\frac{5}{2}} \sum U_{pki} \right\| \leq n^{-\frac{5}{2}} \sum \underbrace{\|U_{pki}\|}_{\leq 4C^3} \leq 4C^3 n^{-\frac{5}{2}} \cdot \mathcal{O}(n^2) \longrightarrow 0 \quad \text{as } n \to \infty. \qquad (3.48)$$

As the convergence in Equation (3.48) carries over to the expectation of the square, the L^2-convergence of this sub-sum follows. So, in the following, it suffices to consider the sub-sum over differing indices. The above approach does not work here (the $\mathcal{O}(n^3)$ terms are no longer dominated by the factor $n^{-\frac{5}{2}}$), which is why we need to apply Lemma B.4.6 in order to obtain the desired convergence. We have to verify condition (B.83), that is,

$$\mathbb{E}\left[U_{pki} U_{qlj}^\top \right] = 0 \qquad \text{whenever } k,\, i,\, l \text{ or } j \text{ differs from the rest.}$$

Note that we have swapped the transpose to comply with the dimensions, since we take U_{pki} to be a row vector and not a column vector. For symmetry reasons, we need only consider the cases where k or i is different from all other indices. Let us first assume that k differs from the rest. Then,

$$\begin{aligned}
\mathbb{E}\left[U_{pki} U_{qlj}^\top \right] &= \mathbb{E}\left[\left(\int_{\mathcal{I}} M_s^{(k)} \frac{\mathrm{d}}{\mathrm{d}\theta} \Lambda_\theta^{(p)}(s) \, \mathrm{d}M_s^{(i)} \right) \left(\int_{\mathcal{I}} M_t^{(l)} \frac{\mathrm{d}}{\mathrm{d}\theta} \Lambda_\theta^{(q)}(t) \, \mathrm{d}M_t^{(j)} \right)^\top \right] \\
&= \mathbb{E}\left[\int_{\mathcal{I}} \underbrace{M_s^{(k)}}_{=:X_s} \underbrace{\int_{\mathcal{I}} M_t^{(l)} \frac{\mathrm{d}}{\mathrm{d}\theta} \Lambda_\theta^{(p)}(s) \frac{\mathrm{d}}{\mathrm{d}\theta} \Lambda_\theta^{(q)}(t)^\top \, \mathrm{d}M_t^{(j)}}_{=:Y_s} \, \mathrm{d} \underbrace{M_s^{(i)}}_{=:Z_s} \right] \\
&= \mathbb{E}\left[\int_{\mathcal{I}} XY \, \mathrm{d}Z \right] , \qquad (3.49)
\end{aligned}$$

where X_t is independent of $\sigma\left(\{Y_s, Z_s \,:\, s \in \mathcal{I}\,, s \leq t\}\right)$ for each $t \in \mathcal{I}$ according to the assumption on k. We can thus apply Lemma B.3.1(i) from Appendix B.3 to achieve that

$$\mathbb{E}\left[\int_{\mathcal{I}} XY \, \mathrm{d}Z \right] = \mathbb{E}\left[\int_{\mathcal{I}} \underbrace{\mathbb{E}(X)}_{=0} Y \, \mathrm{d}Z \right] = 0 \,,$$

because $X = M^{(k)}$ is a centred martingale (see again the proof of Theorem 3.20)[13]. If, on the other hand, i is different from all other indices, then we can apply part (ii) of Lemma B.3.1 instead to conclude:

$$\mathbb{E}\left[\int_{\mathcal{I}} XY \,\mathrm{d}Z\right] = \mathbb{E}\left[\int_{\mathcal{I}} XY \,\mathrm{d}\underbrace{\mathbb{E}Z}_{=0}\right] = 0\,,$$

again because $Z = M^{(i)}$ is a centred martingale. Consequently, condition (B.83) applies, so Lemma B.4.6 provides:

$$\mathbb{E}\left\| \sum_{\substack{p,k,i=1\\ p\neq k\neq i\neq p}}^{n} U_{pki} \right\|^2 \leq 32 \sum \mathbb{E}\left[U_{pki}U_{qki}^{\top}\right]\,,$$

where the summation takes place over $\mathcal{O}(n^4)$ summands. The specific index combinations are given in Lemma B.4.6, but are of no interest to us. Overall, Hölder's inequality yields:

$$\begin{aligned}
\mathbb{E}\left\| n^{-\frac{5}{2}} \sum_{\substack{p,k,i=1\\ p\neq k\neq i\neq p}}^{n} U_{pki} \right\|^2 &\leq 32n^{-5} \sum \mathbb{E}\left[U_{pki}U_{qki}^{\top}\right] \\
&\leq \frac{32}{n^5} \sum \underbrace{\sqrt{\mathbb{E}\left\|U_{pki}\right\|^2 \mathbb{E}\left\|U_{qki}\right\|^2}}_{\leq 16C^6 \text{ by Eq. (3.47)}} \\
&\leq \frac{512C^6}{n^5}\mathcal{O}(n^4) \longrightarrow 0 \quad \text{as } n \to \infty.
\end{aligned}$$

Finally, since both sub-sums that make up $\beta_n(\theta)$ converge to 0 in the quadratic mean, $\beta_n(\theta) \xrightarrow{\mathrm{L}^2} 0$ follows. ☐

Having shown that the leading term is asymptotically negligible, the proof of the representation theorem for β_n is now analogous to that for α_n.

Theorem 3.22 (Representation Theorem for the Auxiliary Process β_n)**.**
Under Assumption 3.8, the process β_n evaluated at the MDE $\hat{\theta}_n$ admits the representation

$$\beta_n(\hat{\theta}_n) = \beta_n(\theta^*) + \sqrt{n}B_n(\hat{\theta}_n - \theta^*)\,, \quad \textit{where } \beta_n(\theta^*) \xrightarrow{\mathbb{P}} 0 \\ \textit{and } B_n \xrightarrow{\mathbb{P}} 0 \quad (n \to \infty). \tag{3.50}$$

Proof. As in the proof of Theorem 3.20, we apply the mean value theorem for functions of several variables to the individual components of β_n, and obtain

$$\beta_{n,j}(\hat{\theta}_n) = \beta_{n,j}(\theta^*) + \sqrt{n}\underbrace{\left[\int_{\mathcal{I}} \overline{M}^{(n)} D_\theta \frac{\partial}{\partial\theta_j}\overline{\Lambda}_\theta^{(n)}\bigg|_{\theta=\tilde{\theta}_{n,j}} \mathrm{d}\overline{M}^{(n)}\right]}_{=:D_\theta\tilde{\beta}_{n,j}(\tilde{\theta}_{n,j})}(\hat{\theta}_n - \theta^*)\,,$$

[13] In order to apply Lemma B.3.1(i), we must formally ensure that the integrator and integrands do not share common discontinuities: Due to $M^{(k)} = N^{(k)} - \Lambda_{\theta^*}^{(k)}$, the discontinuities of $M^{(k)}$ correspond to the jumps of $N^{(k)}$, whereas the discontinuities of $M^{(i)}$ are determined by $N^{(i)}$. Therefore, by Lemma A.37, $M^{(k)}$ and $M^{(i)}$ (and hence X and Z) have no common discontinuities with probability 1. Moreover, Y inherits the continuity from $\Lambda_\theta^{(p)}$ by virtue of condition (A2).

where again $\tilde{\theta}_{n,j} \in \{\theta^* + s(\hat{\theta}_n - \theta^*) \,:\, s \in [0,1]\,\}$ is a d-dimensional vector. The stochastic convergence of the leading term follows immediately from Lemma 3.21. Moreover, the matrix B_n can then be constructed similarly to the matrix A_n before, that is,

$$B_n := \begin{pmatrix} D_\theta \tilde{\beta}_{n,1}(\tilde{\theta}_{n,1}) \\ \vdots \\ D_\theta \tilde{\beta}_{n,d}(\tilde{\theta}_{n,d}) \end{pmatrix}.$$

We proceed as in the proof of the previous representation theorem to derive that $B_n \xrightarrow{\mathbb{P}} 0$ as $n \to \infty$. For this, we observe by virtue of the mean value theorem for vector-valued functions:

$$D_\theta \tilde{\beta}_{n,j}(\tilde{\theta}_{n,j}) = D_\theta \tilde{\beta}_{n,j}(\theta^*) + \left[\int_0^1 D_\theta^2\, \tilde{\beta}_{n,j}(\theta^* + s(\tilde{\theta}_{n,j} - \theta^*))\,\mathrm{d}s\right](\tilde{\theta}_{n,j} - \theta^*)\,. \tag{3.51}$$

To show that the second summand converges to 0 in probability, it once more suffices to prove that the integrand is $\mathbb{P}$-almost surely uniformly bounded on $\mathrm{B}_\varepsilon(\theta^*)$. Here, the integrand is given by

$$D_\theta^2\, \tilde{\beta}_{n,j}(\theta) = \int_{\mathcal{I}} \overline{M}^{(n)} D_\theta^2 \frac{\partial}{\partial \theta_j} \overline{\Lambda}_\theta^{(n)}\, \mathrm{d}\overline{M}^{(n)}\,,$$

which is $\mathbb{P}$-almost surely bounded according to Lemma 3.19 in conjunction with conditions (C3) and ($\widetilde{\mathrm{A}}3$), compare the proof of Theorem 3.20 for more details. Consequently, the second summand is also dominated here by the $\mathbb{P}$-almost sure convergence of the difference $\tilde{\theta}_{n,j} - \theta^*$ toward 0. For the remaining summand, we have

$$\begin{aligned} D_\theta \tilde{\beta}_{n,j}(\theta^*) &= \int_{\mathcal{I}} \overline{M}^{(n)} D_\theta \frac{\partial}{\partial \theta_j} \overline{\Lambda}_\theta^{(n)}\Big|_{\theta=\theta^*} \mathrm{d}\overline{M}^{(n)} \\ &= \int_{\mathcal{I}} \overline{M}^{(n)} \frac{\mathrm{d}}{\mathrm{d}\theta} \frac{\partial}{\partial \theta_j} \overline{\Lambda}_{\theta^*}^{(n)}\, \mathrm{d}\overline{M}^{(n)} \\ &= \frac{1}{n^3} \sum_{i_1,i_2,i_3=1}^{n} U_{i_1,i_2,i_3}\,, \end{aligned}$$

where we defined analogously to the proof of Theorem 3.20:

$$U_{i_1,i_2,i_3} := \int_{\mathcal{I}} M^{(i_1)} \frac{\mathrm{d}}{\mathrm{d}\theta} \frac{\partial}{\partial \theta_j} \Lambda_{\theta^*}^{(i_2)}\, \mathrm{d}M^{(i_3)}\,.$$

We can proceed with the random vectors U_{i_1,i_2,i_3} in a similar way as with their counterparts from Equation (3.45). Again, we only need to verify that $\mathbb{E}(U_{i_1,i_2,i_3}) = 0$ holds as soon as the indices i_1, i_2, i_3 are distinct. For this, we can easily adopt the technique from Theorem 3.20 by applying Lemma B.3.1 and exploiting that the involved martingales are centred. Lemma B.4.7 readily yields

$$\frac{1}{n^3} \sum_{i_1,i_2,i_3=1}^{n} U_{i_1,i_2,i_3} \xrightarrow{\mathrm{L}^2} 0 \quad \text{as } n \to \infty,$$

and from here $D_\theta \tilde{\beta}_{n,j}(\tilde{\theta}_{n,j}) = o_{\mathbb{P}}(1)$ follows as before. □

Representation Theorem for the Process γ_n

In the final step of dealing with the asymptotics of the auxiliary parametric processes, we treat the process γ_n. Again, the major effort lies in the study of the leading term $\gamma_n(\theta^*)$, even more so than for the process β_n, where it was found to be asymptotically negligible: As it turns out, $\gamma_n(\theta^*)$ contributes the essential part to the asymptotic distribution. We once more recollect Definition 3.11, where γ_n was defined to be the parametric process with index set $\mathrm{B}_\varepsilon(\theta^*)$ given via

$$\gamma_n(\theta) = \sqrt{n} \int_{\mathcal{I}} \overline{M}^{(n)} \frac{\mathrm{d}}{\mathrm{d}\theta} \overline{\Lambda}_\theta^{(n)} \, \mathrm{d}\overline{\Lambda}_{\theta^*}^{(n)} \, .$$

Lemma 3.23 (Asymptotics of the Leading Term $\gamma_n(\theta^*)$; cf. Lemma 13 of Kopperschmidt and Stute 2013, p. 1292)**.**
Under Assumption 3.8, for each $\theta \in B_\varepsilon(\theta^*)$ holds:

$$\gamma_n(\theta) = \sqrt{n} \int_{\mathcal{I}} \overline{M}^{(n)} \mathbb{E} \frac{\mathrm{d}}{\mathrm{d}\theta} \Lambda_\theta \, \mathrm{d}\mathbb{E}\Lambda_{\theta^*} + o_{\mathbb{P}}(1) \, . \tag{3.52}$$

Proof. Even though the statement itself differs slightly[14] from Lemma 13 of Kopperschmidt and Stute 2013, the essential idea of the proof can once more be adopted. The proof is similar to that of Lemma 3.21 in this respect, and again relies on the second moment bounds from Appendix B.4.2. In particular, we will find that the convergence of the remainder holds not only stochastically, but also in the quadratic mean. For each $\theta \in \mathrm{B}_\varepsilon(\theta^*)$, we have to show that

$$\sqrt{n} \int_{\mathcal{I}} \overline{M}^{(n)} \frac{\mathrm{d}}{\mathrm{d}\theta} \overline{\Lambda}_\theta^{(n)} \, \mathrm{d}\overline{\Lambda}_{\theta^*}^{(n)} - \underbrace{\sqrt{n} \int_{\mathcal{I}} \overline{M}^{(n)} \mathbb{E} \frac{\mathrm{d}}{\mathrm{d}\theta} \Lambda_\theta \, \mathrm{d}\mathbb{E}\Lambda_{\theta^*}}_{=:\tilde{\gamma}_n(\theta)} = \gamma_n(\theta) - \tilde{\gamma}_n(\theta) = o_{\mathbb{P}}(1) \, . \tag{3.53}$$

This is intuitively plausible, since according to the Lemmas 3.15 and 3.16 it holds uniformly on $\mathcal{I} \times K$, for any compact $K \subset \mathrm{B}_\varepsilon(\theta^*)$, that

$$\overline{\Lambda}_\theta^{(n)} \xrightarrow{\mathbb{P}} \mathbb{E}\Lambda_\theta \, , \qquad \frac{\mathrm{d}}{\mathrm{d}\theta} \overline{\Lambda}_\theta^{(n)} \xrightarrow{\mathbb{P}} \mathbb{E} \frac{\mathrm{d}}{\mathrm{d}\theta} \Lambda_\theta \, .$$

In this intuition, however, the counteracting factor $\sqrt{n}$ is neglected, and the subsidiary lemmas from the appendix are needed to control the inflation it causes. For this, we rearrange the terms of Equation (3.53) to obtain (this decomposition is due to Kopperschmidt and Stute 2013, pp. 1292–1293):

$$\begin{aligned} \gamma_n(\theta) - \tilde{\gamma}_n(\theta) = {} & \sqrt{n} \int_{\mathcal{I}} \overline{M}^{(n)} \frac{\mathrm{d}}{\mathrm{d}\theta} \overline{\Lambda}_\theta^{(n)} \, \mathrm{d}\overline{\Lambda}_{\theta^*}^{(n)} - \sqrt{n} \int_{\mathcal{I}} \overline{M}^{(n)} \frac{\mathrm{d}}{\mathrm{d}\theta} \overline{\Lambda}_\theta^{(n)} \, \mathrm{d}\mathbb{E}\Lambda_{\theta^*} \\ & + \sqrt{n} \int_{\mathcal{I}} \overline{M}^{(n)} \frac{\mathrm{d}}{\mathrm{d}\theta} \overline{\Lambda}_\theta^{(n)} \, \mathrm{d}\mathbb{E}\Lambda_{\theta^*} - \sqrt{n} \int_{\mathcal{I}} \overline{M}^{(n)} \mathbb{E} \frac{\mathrm{d}}{\mathrm{d}\theta} \Lambda_\theta \, \mathrm{d}\mathbb{E}\Lambda_{\theta^*} \end{aligned}$$

[14] The corresponding Lemma in Kopperschmidt and Stute 2013, p. 1292 contains a stronger property. It states that Equation (3.52) holds not only pointwise for each θ but uniformly on compacta $K \subset \mathrm{B}_\varepsilon(\theta^*)$, and that the leading term is tight. Accordingly, only parts of the proof are relevant to us. This is due to the fact that here we only deal with step (iv) of the proof sketch, while the original proof also covers parts of step (ii).

$$
\begin{aligned}
&= \underbrace{\sqrt{n}\int_{\mathcal{I}} \overline{M}^{(n)} \frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_{\theta}^{(n)} \,\mathrm{d}\left(\overline{\Lambda}_{\theta^*}^{(n)} - \mathbb{E}\Lambda_{\theta^*}\right)}_{=:\tilde{\gamma}_n^{(1)}(\theta)} \\
&\quad + \underbrace{\sqrt{n}\int_{\mathcal{I}} \overline{M}^{(n)} \left[\frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_{\theta}^{(n)} - \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta}\right] \mathrm{d}\mathbb{E}\Lambda_{\theta^*}}_{=:\tilde{\gamma}_n^{(2)}(\theta)} \,.
\end{aligned}
$$

Equation (3.53) consequently holds if both $\tilde{\gamma}_n^{(1)}(\theta)$ and $\tilde{\gamma}_n^{(2)}(\theta)$ converge to 0 in probability. We proceed to study these processes individually, again using the auxiliary lemmas from Appendix B.4.2 given by Kopperschmidt and Stute 2013, pp. 1295–1297. The process $\tilde{\gamma}_n^{(1)}$ is handled analogously to β_n in the proof of Lemma 3.21. We need to apply Lemma B.4.6, so we define the d-variate random vectors

$$
U_{pki} := \int_{\mathcal{I}} M^{(k)} \frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta}^{(p)} \,\mathrm{d}\left(\Lambda_{\theta^*}^{(i)} - \mathbb{E}\Lambda_{\theta^*}\right) \,,
$$

which are square-integrable (compare Equation (3.47)). Repeating the arguments we applied to $\beta_n(\theta)$, it is sufficient to prove condition (B.83), that is,

$$
\mathbb{E}\left[U_{pki}U_{qlj}^{\top}\right] = 0 \qquad \text{whenever } k,\, i,\, l \text{ or } j \text{ differs from the rest,}
$$

in order to show that $\tilde{\gamma}_n^{(1)}(\theta) \xrightarrow{\mathrm{L}^2} 0$. Here, the reasoning can be adopted almost verbatim, since β_n and $\tilde{\gamma}_n^{(1)}$ differ only in terms of the integrator: $Z = \Lambda_{\theta^*}^{(i)} - \mathbb{E}\Lambda_{\theta^*}$ thereby replaces the martingale $M^{(i)}$ in Equation (3.49) (the fact that the integrator of Y also changes does not interfere with the proof). Since we have only exploited throughout the proof that the process $M^{(i)}$ is centred (and not the martingale property), the statements continue to hold because

$$
\mathbb{E}Z = \mathbb{E}\Lambda_{\theta^*}^{(i)} - \mathbb{E}\Lambda_{\theta^*} \equiv 0 \,,
$$

as $\Lambda_{\theta^*}^{(i)}$ is an i.i.d. copy of Λ_{θ^*} by definition. Hence, $\tilde{\gamma}_n^{(1)}(\theta) \xrightarrow{\mathrm{L}^2} 0$ and thus also $\tilde{\gamma}_n^{(1)}(\theta) = o_{\mathbb{P}}(1)$ follows from Lemma B.4.6.
For the process $\tilde{\gamma}_n^{(2)}$, we similarly define square-integrable d-variate random vectors via

$$
U_{pk} := \int_{\mathcal{I}} M^{(k)} \left[\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta}^{(p)} - \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta}\right] \mathrm{d}\mathbb{E}\Lambda_{\theta^*} \,,
$$

so that

$$
\begin{aligned}
\tilde{\gamma}_n^{(2)}(\theta) &= \sqrt{n}\int_{\mathcal{I}} \overline{M}^{(n)} \left[\frac{\mathrm{d}}{\mathrm{d}\theta}\overline{\Lambda}_{\theta}^{(n)} - \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta}\right] \mathrm{d}\mathbb{E}\Lambda_{\theta^*} \\
&= \frac{\sqrt{n}}{n^2}\sum_{p,k=1}^{n} \int_{\mathcal{I}} M^{(k)} \left[\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta}^{(p)} - \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta}\right] \mathrm{d}\mathbb{E}\Lambda_{\theta^*} \\
&= n^{-\frac{3}{2}} \sum_{p,k=1}^{n} U_{pk} \,. \qquad (3.54)
\end{aligned}
$$

While we will apply Lemma B.4.5 instead of Lemma B.4.6, the principal argument remains the same: In Equation (3.54), the sub-sum over $p = k$ is of negligible order (compare

Equation (3.48)), so it suffices to consider the summands where $p \neq k$. From here, we only need to check condition (B.80) and use Lemma B.4.5 to conclude that

$$\begin{aligned}\mathbb{E}\Big\| n^{-\frac{3}{2}} \sum_{p\neq k} U_{pk} \Big\|^2 &\leq 2n^{-3} \sum_{p\neq k} \mathbb{E}\, \|U_{pk}\|^2 \\ &= \frac{2}{n^3}\mathcal{O}(n^2) \longrightarrow 0 \quad \text{as } n \to \infty,\end{aligned}$$

which in turn implies that $\tilde{\gamma}_n^{(2)}(\theta) = o_{\mathbb{P}}(1)$. In order to verify condition (B.80), we have to show that

$$\mathbb{E}\left[U_{pk} U_{ql}^\top\right] = 0 \qquad \text{whenever one index differs from the rest.}$$

As usual, we need to examine only the cases where p or k is different from the other indices. If we assume that k differs from the rest, we obtain:

$$\begin{aligned}\mathbb{E}\left[U_{pk} U_{ql}^\top\right] &= \mathbb{E}\Bigg[\left(\int_{\mathcal{I}} M_s^{(k)} \left[\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta^{(p)}(s) - \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta(s)\right] \mathrm{d}\mathbb{E}\Lambda_{\theta^*}(s)\right) \\ &\qquad \left(\int_{\mathcal{I}} M_t^{(l)} \left[\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta^{(q)}(t) - \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta(t)\right] \mathrm{d}\mathbb{E}\Lambda_{\theta^*}(t)\right)^\top\Bigg] \\ &= \mathbb{E}\Bigg[\int_{\mathcal{I}} \underbrace{M_s^{(k)}}_{=:X_s} \underbrace{\int_{\mathcal{I}} M_t^{(l)} \left[\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta^{(p)}(s) - \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta(s)\right] \left[\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta^{(q)}(t) - \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta(t)\right]^\top \mathrm{d}\mathbb{E}\Lambda_{\theta^*}(t)}_{=:Y_s} \,\mathrm{d}\, \underbrace{\mathbb{E}\Lambda_{\theta^*}(s)}_{=:Z_s}\Bigg] \\ &= \mathbb{E}\left[\int_{\mathcal{I}} XY \,\mathrm{d}Z\right].\end{aligned}$$

Since k is distinct from the other indices, X_t is independent of $\sigma\left(\{Y_s, Z_s \,:\, s \in \mathcal{I}, s \leq t\}\right)$ for each $t \in \mathcal{I}$, so Lemma B.3.1(i) yields

$$\mathbb{E}\left[U_{pk} U_{ql}^\top\right] = \mathbb{E}\left[\int_{\mathcal{I}} \mathbb{E}(X) Y \,\mathrm{d}Z\right] = 0\,,$$

where once more we exploited that $X = M^{(k)}$ is a centred martingale by construction. If instead p is different from the remaining indices, then we simply need to swap the roles of $M_s^{(k)}$ and $\left[\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta^{(p)}(s) - \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta(s)\right]$ in the above chain of equations to achieve the exact same result. The only difference here is that $\left[\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta^{(p)}(s) - \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta(s)\right]$ - while also centred - is not a martingale, but that is irrelevant in this context. Hence, condition (B.80) applies, which concludes the proof. □

Lemma 3.23 shows that $\gamma_n(\theta)$ and $\tilde{\gamma}_n(\theta)$ are asymptotically equivalent (see Equation (3.53)), where

$$\tilde{\gamma}_n(\theta) := \sqrt{n} \int_{\mathcal{I}} \overline{M}^{(n)} \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta \,\mathrm{d}\mathbb{E}\Lambda_{\theta^*}\,.$$

We are therefore interested in the limiting behavior of $\tilde{\gamma}_n(\theta)$, which is the subject of the following lemma. The advantage in considering $\tilde{\gamma}_n(\theta)$ is that instead of the random aggregated cumulative intensity $\overline{\Lambda}_\theta^{(n)}$, its deterministic expectation $\mathbb{E}\Lambda_\theta$ occurs, which is furthermore independent of n.

Lemma 3.24 (Application of the Central Limit Theorem to $\tilde{\gamma}_n(\theta)$; cf. Kopperschmidt and Stute 2013, p. 1294)**.**
Under Assumption 3.8 holds for each $\theta \in B_\varepsilon(\theta^*)$:

$$\tilde{\gamma}_n(\theta) \xrightarrow{\mathrm{d}} \mathcal{N}_d(0, \Sigma(\theta)) \quad (n \to \infty)\,,$$

where $\Sigma(\theta)$ is a $d \times d$ matrix with entries

$$\begin{aligned} \Sigma_{ij}(\theta) &:= \int_{\mathcal{I}} \varphi_i(t,\theta)\varphi_j(t,\theta)\,\mathrm{d}\mathbb{E}\Lambda_{\theta^*}(t)\,, && 1 \le i, j \le d\,, \\ \varphi(t,\theta) = (\varphi_1(t,\theta), \ldots, \varphi_d(t,\theta)) &:= \int_{[t,\tau]} \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta\,\mathrm{d}\mathbb{E}\Lambda_{\theta^*}\,, && t \in \mathcal{I}\,. \end{aligned}$$

The connection to the main Theorem 3.12 is evident, since at $\theta = \theta^*$ the matrix $\Sigma(\theta)$ coincides with the matrix $\Sigma(\theta^*)$ from Equation (3.21). This lemma is therefore central to the asymptotic distribution of the MDE. Even though the proof of this statement is straightforward, we would like to emphasize that the major effort lies in the definition and discussion of the auxiliary processes, which is based on the work of Kopperschmidt and Stute.

Proof of Lemma 3.24. We only need to verify the specific shape of the asymptotic covariance matrix, since normality immediately follows from the central limit theorem by writing

$$\tilde{\gamma}_n(\theta) = \frac{\sqrt{n}}{n}\sum_{i=1}^{n} \underbrace{\int_{\mathcal{I}} M^{(i)}\mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta\,\mathrm{d}\mathbb{E}\Lambda_{\theta^*}}_{=:X_i} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\,,$$

where the $X_1, X_2, \ldots$ form a sequence of centred i.i.d. random vectors by construction. The proof for this (i.e., the application of Fubini's theorem and the Itô-Isometry) essentially parallels the calculations in Kopperschmidt 2005, pp. 110–111, although we can avoid to reapply Lemma B.3.1 and simplify the proof to some extent.
Note that because of $M_0^{(i)} = 0$, we always have

$$M_t^{(i)} = M_t^{(i)} - M_0^{(i)} = \int_{[0,t]} \mathrm{d}M^{(i)}\,.$$

Therefore, using Fubini's theorem, we can rearrange X_i to achieve:

$$\begin{aligned} X_i &= \int_{\mathcal{I}} M_t^{(i)}\mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta(t)\,\mathrm{d}\mathbb{E}\Lambda_{\theta^*}(t) \\ &= \int_{\mathcal{I}}\left(\int_{[0,t]} \mathrm{d}M_s^{(i)}\right)\mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta(t)\,\mathrm{d}\mathbb{E}\Lambda_{\theta^*}(t) \\ &= \int_{\mathcal{I}}\int_{[0,t]} \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta(t)\,\mathrm{d}M_s^{(i)}\,\mathrm{d}\mathbb{E}\Lambda_{\theta^*}(t) && \Big|\text{apply Fubini's theorem} \end{aligned}$$

$$= \int_{\mathcal{I}} \underbrace{\int_{[s,\tau]} \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_\theta(t)\,\mathrm{d}\mathbb{E}\Lambda_{\theta^*}(t)}_{=\varphi(s,\theta)}\,\mathrm{d}M_s^{(i)}$$
$$= \int_{\mathcal{I}} \varphi(s,\theta)\,\mathrm{d}M_s^{(i)} .$$

The $M^{(i)}$ are independent copies of M, so the covariance matrix $\Sigma(\theta)$ can be computed by

$$\Sigma(\theta) = \mathbb{E}\left[\left(\int_{\mathcal{I}} \varphi(\cdot,\theta)\,\mathrm{d}M\right)\left(\int_{\mathcal{I}} \varphi(\cdot,\theta)\,\mathrm{d}M\right)^{\top}\right]$$
$$= \mathbb{E}\left[\left(\int_{\mathcal{I}} \varphi_i(\cdot,\theta)\,\mathrm{d}M\right)\left(\int_{\mathcal{I}} \varphi_j(\cdot,\theta)\,\mathrm{d}M\right)\right]_{1\le i,j\le d} .$$

The martingale M is square-integrable by condition (C3), while $\varphi(\cdot,\theta)$ is deterministic and hence predictable. For the (i,j)-th entry of this matrix, we thus obtain by virtue of the Itô-Isometry for square-integrable martingales from Theorem A.41 and the identity $4ab = (a+b)^2 - (a-b)^2$:

$$\Sigma_{ij}(\theta) = \mathbb{E}\left[\left(\int_{\mathcal{I}} \varphi_i(\cdot,\theta)\,\mathrm{d}M\right)\left(\int_{\mathcal{I}} \varphi_j(\cdot,\theta)\,\mathrm{d}M\right)\right]$$
$$= \frac{1}{4}\mathbb{E}\left[\left(\int_{\mathcal{I}} \varphi_i(\cdot,\theta) + \varphi_j(\cdot,\theta)\,\mathrm{d}M\right)^2 - \left(\int_{\mathcal{I}} \varphi_i(\cdot,\theta) - \varphi_j(\cdot,\theta)\,\mathrm{d}M\right)^2\right]$$
$$= \frac{1}{4}\mathbb{E}\left[\int_{\mathcal{I}} \left(\varphi_i(\cdot,\theta) + \varphi_j(\cdot,\theta)\right)^2\,\mathrm{d}\Lambda_{\theta^*} - \int_{\mathcal{I}} \left(\varphi_i(\cdot,\theta) - \varphi_j(\cdot,\theta)\right)^2\,\mathrm{d}\Lambda_{\theta^*}\right]$$
$$= \mathbb{E}\left[\int_{\mathcal{I}} \varphi_i(\cdot,\theta)\varphi_j(\cdot,\theta)\,\mathrm{d}\Lambda_{\theta^*}\right]$$
$$= \int_{\mathcal{I}} \varphi_i(\cdot,\theta)\varphi_j(\cdot,\theta)\,\mathrm{d}\mathbb{E}\Lambda_{\theta^*} ,$$

where the last equation holds since $\varphi(\cdot,\theta)$ is deterministic (this can be seen as another application of Lemma B.3.1, because both the integrand and the integrator are continuous, but the statement applies more generally too). This proves the asserted shape of the covariance matrix. □

Similar to the representation theorems formulated for α_n and β_n, we close this paragraph with the corresponding counterpart for γ_n.

Theorem 3.25 (Representation Theorem for the Auxiliary Process γ_n)**.**
Under Assumption 3.8, the process γ_n evaluated at the MDE $\hat{\theta}_n$ admits the representation

$$\gamma_n(\hat{\theta}_n) = \gamma_n(\theta^*) + \sqrt{n}C_n(\hat{\theta}_n - \theta^*)\,, \quad \textit{where } \gamma_n(\theta^*) \xrightarrow{\mathrm{d}} \mathcal{N}_d\left(0, \Sigma(\theta^*)\right)$$
$$\textit{and } C_n \xrightarrow{\mathbb{P}} 0 \quad (n\to\infty). \tag{3.55}$$

Proof. The stochastic of the leading term is due to Lemmas 3.23 and 3.24, since by virtue of Slutzky's theorem we have

$$\gamma_n(\theta^*) = \tilde{\gamma}_n(\theta^*) + \underbrace{(\gamma_n(\theta^*) - \tilde{\gamma}_n(\theta^*))}_{=o_{\mathbb{P}}(1)} \xrightarrow{d} \mathcal{N}_d\left(0, \Sigma(\theta^*)\right) . \tag{3.56}$$

The rest of the proof is then in complete agreement with that of Theorem 3.22, the representation theorem for β_n. Here,

$$\gamma_{n,j}\left(\hat{\theta}_n\right) = \gamma_{n,j}\left(\theta^*\right) + \sqrt{n}\left[\int_{\mathcal{I}} \overline{M}^{(n)} D_\theta \frac{\partial}{\partial \theta_j} \overline{\Lambda}_\theta^{(n)}\bigg|_{\theta=\tilde{\theta}_{n,j}} \mathrm{d}\overline{\Lambda}_{\theta^*}^{(n)}\right]\left(\hat{\theta}_n - \theta^*\right) ,$$

which allows us to construct the matrix C_n as we did with A_n and B_n (we omit further specification to avoid duplicate use of the notation $\tilde{\gamma}_n$). The convergence $C_n \xrightarrow{\mathbb{P}} 0$ can then be shown exactly as for B_n, because β_n and γ_n differ only in terms of the integrator ($\overline{M}^{(n)}$ instead of $\overline{\Lambda}_{\theta^*}^{(n)}$). At first glance, this may seem problematic, as we exploited $\mathbb{E}\overline{M}^{(n)} \equiv 0$ in the application of Lemma B.3.1(ii). However, we can simply use statement (i) of that lemma here, since the process $\overline{M}^{(n)}$ occurs a second time as part of the integrand. The assertion then follows as before from applying Lemma B.4.7. □

3.3.3. Proof of the Asymptotic Normality

We have now completed the preliminary work to merge the proved asymptotics of Ψ_n and the auxiliary processes α_n, β_n and γ_n into a proof of Theorem 3.12. Our main contribution consisted of the asymptotic analysis of the process Ψ_n (which was neglected by Kopperschmidt and Stute) and the new representation Theorems 3.20, 3.22 and 3.25 for the auxiliary processes, while the asymptotics of their leading terms were already discussed by Kopperschmidt 2005 and Kopperschmidt and Stute 2013 and we only had to properly adjust the existing proofs. In particular, by modifying the assumptions, we were able to circumvent proving the tightness property that has been essential in the original proof. For ease of reading, we restate the main theorem one more time.

Theorem 3.12 (Asymptotic Normality of the Minimum Distance Estimator; cf. Kopperschmidt and Stute 2013, p. 1281)**.**
Under assumptions (A1), (A2), (A3) *and* (A4) *together with the assumptions from Section 3.2 holds:*

$$\sqrt{n}\Phi_0(\theta^*)\left(\hat{\theta}_n - \theta^*\right) \xrightarrow{\mathrm{d}} \mathcal{N}_d\left(0, \Sigma(\theta^*)\right) \quad (n \to \infty)\,,$$

where $\Sigma(\theta^*)$ *is a* $d \times d$ *matrix with entries*

$$\Sigma_{ij}(\theta^*) := \int_{\mathcal{I}} \varphi_i(t)\varphi_j(t)\,\mathrm{d}\mathbb{E}\Lambda_{\theta^*}(t)\,, \qquad 1 \le i,j \le d\,, \tag{3.21}$$

$$\varphi(t) = (\varphi_1(t), \ldots, \varphi_d(t)) := \int_{[t,\tau]} \mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta}\Lambda_{\theta^*}\,\mathrm{d}\mathbb{E}\Lambda_{\theta^*}\,, \qquad t \in \mathcal{I}\,.$$

Proof. We start from Lemma 3.10, which in abbreviated form reads as follows:

$$\alpha_n(\hat{\theta}_n) + \sqrt{n}\Psi_n(\hat{\theta}_n)(\hat{\theta}_n - \theta^*) = \beta_n(\hat{\theta}_n) + \gamma_n(\hat{\theta}_n)\,.$$

Applying the representation Theorems 3.20 (for α_n), 3.22 (for β_n) and 3.25 (for γ_n) yields:

$$\begin{aligned} & \sqrt{n}A_n(\hat{\theta}_n - \theta^*) + \sqrt{n}\Psi_n(\hat{\theta}_n)(\hat{\theta}_n - \theta^*) \\ & \quad = \underbrace{\beta_n(\theta^*)}_{=o_{\mathbb{P}}(1)} + \sqrt{n}B_n(\hat{\theta}_n - \theta^*) + \gamma_n(\theta^*) + \sqrt{n}C_n(\hat{\theta}_n - \theta^*) \\ \Longleftrightarrow \quad & \gamma_n(\theta^*) + o_{\mathbb{P}}(1) = \sqrt{n}\Big(\Psi_n(\hat{\theta}_n) + A_n - B_n - C_n\Big)(\hat{\theta}_n - \theta^*)\,. \end{aligned} \tag{3.57}$$

Since the left-hand side of Equation (3.57) converges in distribution to the required normal distribution according to Slutzky's theorem and Equation (3.56), the same is true for the right-hand side, so we obtain:

$$\underbrace{\sqrt{n}\left(\Psi_n(\hat{\theta}_n) + A_n - B_n - C_n\right)}_{=:\Sigma_n}(\hat{\theta}_n - \theta^*) \xrightarrow{\mathrm{d}} \mathcal{N}_d\left(0, \Sigma(\theta^*)\right) \quad (n \to \infty)\,. \tag{3.58}$$

To complete the proof, it remains only to show that this asymptotic distribution is preserved when Σ_n is replaced by $\Phi_0(\theta^*)$. By Corollary 3.17 and because $A_n \xrightarrow{\mathbb{P}} 0$, $B_n \xrightarrow{\mathbb{P}} 0$, $C_n \xrightarrow{\mathbb{P}} 0$, the continuous mapping theorem provides:

$$\Sigma_n \xrightarrow{\mathbb{P}} \Phi_0(\theta^*) \quad (n \to \infty)\,.$$

According to Lemma 3.18, $\Phi_0(\theta^*)$ is positive definite and thus invertible. In particular, $\det \Phi_0(\theta^*) > 0$. We can then infer from Corollary B.4.2 on the limit of an inverse matrix sequence that

$$\Sigma_n^{-1} \cdot \mathbb{1}_{\{\Sigma_n \text{ is invertible.}\}} \xrightarrow{\mathbb{P}} \Phi_0(\theta^*)^{-1} \quad (n \to \infty)\,. \tag{3.59}$$

Moreover, the continuous mapping theorem (in the proof of Corollary B.4.2 we observed that the determinant mapping is continuous) once again implies that

$$\det \Sigma_n \xrightarrow{\mathbb{P}} \det \Phi_0(\theta^*) \quad (n \to \infty)\,,$$

which allows us to conclude for all $0 < \varepsilon < \det \Phi_0(\theta^*)$:

$$\begin{aligned} \mathbb{P}\left(\{\Sigma_n \text{ is invertible.}\}^{\complement}\right) &= \mathbb{P}\left(\det \Sigma_n = 0\right) \\ &\leq \mathbb{P}\left(|\det \Sigma_n - \det \Phi_0(\theta^*)| > \varepsilon\right) \longrightarrow 0 \quad (n \to \infty)\,. \end{aligned} \tag{3.60}$$

Therefore,

$$\sqrt{n}\Phi_0(\theta^*)(\hat{\theta}_n - \theta^*) - \sqrt{n}\Phi_0(\theta^*)(\hat{\theta}_n - \theta^*) \cdot \mathbb{1}_{\{\Sigma_n \text{ is invertible.}\}} \xrightarrow{\mathbb{P}} 0 \quad (n \to \infty)\,, \tag{3.61}$$

since by Equation (3.60) we have

$$\begin{aligned} & \mathbb{P}\Big(\Big\| \underbrace{\sqrt{n}\Phi_0(\theta^*)(\hat{\theta}_n - \theta^*) - \sqrt{n}\Phi_0(\theta^*)(\hat{\theta}_n - \theta^*) \cdot \mathbb{1}_{\{\Sigma_n \text{ is invertible.}\}}}_{= 0 \text{ as long as } \Sigma_n \text{ is invertible.}} \Big\| > \varepsilon\Big) \\ & \leq \mathbb{P}\left(\{\Sigma_n \text{ is invertible.}\}^{\complement}\right) \longrightarrow 0 \quad (n \to \infty)\,. \end{aligned}$$

Additionally,

$$\begin{aligned}
&\sqrt{n}\Phi_0(\theta^*)(\hat{\theta}_n - \theta^*) \cdot \mathbb{1}_{\{\Sigma_n \text{ is invertible.}\}} \\
&\quad = \sqrt{n}\Phi_0(\theta^*)\Sigma_n^{-1}\Sigma_n(\hat{\theta}_n - \theta^*) \cdot \mathbb{1}_{\{\Sigma_n \text{ is invertible.}\}} \\
&\quad = \sqrt{n}V_n\Sigma_n(\hat{\theta}_n - \theta^*) \,,
\end{aligned}$$

where, again by the continuous mapping theorem and Equation (3.59),

$$V_n := \Phi_0(\theta^*)\Sigma_n^{-1} \cdot \mathbb{1}_{\{\Sigma_n \text{ is invertible.}\}} \xrightarrow{\mathbb{P}} \Phi_0(\theta^*)\Phi_0(\theta^*)^{-1} = \mathbb{I}_{d\times d} \,,$$

and hence

$$\begin{aligned}
&\sqrt{n}\Phi_0(\theta^*)(\hat{\theta}_n - \theta^*) \cdot \mathbb{1}_{\{\Sigma_n \text{ is invertible.}\}} - \sqrt{n}\Sigma_n(\hat{\theta}_n - \theta^*) \\
&\quad = \sqrt{n}V_n\Sigma_n(\hat{\theta}_n - \theta^*) - \sqrt{n}\mathbb{I}_{d\times d}\Sigma_n(\hat{\theta}_n - \theta^*) \\
&\quad = \underbrace{(V_n - \mathbb{I}_{d\times d})}_{=o_{\mathbb{P}}(1)}\sqrt{n}\Sigma_n(\hat{\theta}_n - \theta^*) \xrightarrow{\mathbb{P}} 0 \quad (n \to \infty)
\end{aligned} \tag{3.62}$$

due to Slutzky's theorem and Equation (3.58). With the same arguments then follows:

$$\begin{aligned}
&\sqrt{n}\Phi_0(\theta^*)(\hat{\theta}_n - \theta^*) \\
&\quad = \underbrace{\sqrt{n}\Phi_0(\theta^*)(\hat{\theta}_n - \theta^*) - \sqrt{n}\Phi_0(\theta^*)(\hat{\theta}_n - \theta^*) \cdot \mathbb{1}_{\{\Sigma_n \text{ is invertible.}\}}}_{=\, o_{\mathbb{P}}(1) \text{ by Eq. (3.61)}} \\
&\quad\quad + \underbrace{\sqrt{n}\Phi_0(\theta^*)(\hat{\theta}_n - \theta^*) \cdot \mathbb{1}_{\{\Sigma_n \text{ is invertible.}\}} - \sqrt{n}\Sigma_n(\hat{\theta}_n - \theta^*)}_{=\, o_{\mathbb{P}}(1) \text{ by Eq. (3.62)}} \\
&\quad\quad + \sqrt{n}\Sigma_n(\hat{\theta}_n - \theta^*) \xrightarrow{\mathrm{d}} \mathcal{N}_d\left(0, \Sigma(\theta^*)\right) \quad (n \to \infty) \,,
\end{aligned}$$

which completes the proof. □

Lemma 3.18 on the positive definiteness of $\Phi_0(\theta^*)$ furthermore allows us to give the following corollary:

Corollary 3.26 (Asymptotic Normality of the Minimum Distance Estimator; cf. Kopperschmidt and Stute 2013, p. 1281)**.**
In the situation of Theorem 3.12, we have:

$$\sqrt{n}\left(\hat{\theta}_n - \theta^*\right) \xrightarrow{\mathrm{d}} \mathcal{N}_d\left(0, \Phi_0(\theta^*)^{-1}\Sigma(\theta^*)\Phi_0(\theta^*)^{-1}\right) \quad (n \to \infty) \,.$$

Proof. This is an immediate consequence of Theorem 3.12, since according to Lemma 3.18 the matrix $\Phi_0(\theta^*)$ is positive definite and thus invertible. □

We would like to round out this paragraph by pointing out the errors we found in the original proof given by Kopperschmidt 2005 and Kopperschmidt and Stute 2013 that led to the strengthened requirements of Assumption 3.8.

Remark 3.27 (Errors in the Original Proof of the Asymptotic Normality)**.**
The original proof of the MDE's asymptotic normal distribution - both the detailed version in Kopperschmidt 2005 and its abbreviated version in Kopperschmidt and Stute 2013 - is

largely stringent and comprehensible, with the exception of minor inaccuracies. These flaws mostly turn out to be insignificant: we considered the averaged process Ψ_n when Φ_n could not be evaluated at a suitable intermediate point, and while Glivenko-Cantelli arguments do *not* yield the almost sure uniform convergence of averages of derivative cumulative intensities as claimed in Kopperschmidt and Stute 2013, p. 1293, we were able to devise an alternative proof for Lemma 3.16. Nevertheless, there remained an error that we were not able to fix: the generalization of Kolmogorov's tightness criterion, Theorem 12.3 of Billingsley 1968, p. 95, to multiparameter processes. A proper version of this criterion adjusted to our situation is given in Theorem B.2.6 in the appendix, but Kopperschmidt and Stute assume that the dependence on the dimension d of the parameter space in condition (ii) can be omitted and instead consider the following condition (cf. Kopperschmidt 2005, p. 1290 and Kopperschmidt and Stute 2013, pp. 86, 152):

($\widetilde{\text{ii}}$) There exist constants $\nu > 0$ and $\alpha > 1$ such that

$$\mathbb{E}\,\|X_n(x) - X_n(y)\|^2 \leq \nu\,\|x-y\|^\alpha\ , \quad \text{for all } x,y \in [0,1]^d \text{ and } n \in \mathbb{N}\,.$$

Both authors cite Billingsley 1968 on this, while the article Kopperschmidt and Stute 2013 moreover refers to Bickel and Wichura 1971 as a reference for multi-dimensional generalizations. In fact, in the uni-dimensional case conditions (ii) and ($\widetilde{\text{ii}}$) coincide, and agree with the formulation from said Theorem 12.3 of Billingsley 1968, p. 95. However, we can demonstrate that under condition ($\widetilde{\text{ii}}$), the proof given in Kopperschmidt 2005, pp. 155-157 fails as soon as $d > 1$. The problem arises when Kopperschmidt tries to show that this condition allows the application of Billingsley's Theorem 12.2 (see Billingsley 1968, p. 94). To understand the difficulties, we briefly summarize his procedure up to this step:

1. Without loss of generality, Kopperschmidt assumes that the closure of the parameter space is given by the d-dimensional unit cube (cf. Kopperschmidt 2005, p. 152), that is,
$$\overline{\Theta} = [0,1]^d\ .$$
This is justified in his Theorem A.5.2 (for a more detailed explanation, see also Corollary 2.3.8 from Jakubzik 2017, p. 74). It is also the reason why an additional assumption is made that Θ is simply connected, see Kopperschmidt 2005, p. 28 (this requirement is omitted in Kopperschmidt and Stute 2013 for unknown reasons).

2. For each $k \in \mathbb{N}$, he decomposes the parameter space $\overline{\Theta}$ into $(2k+1)^d$ d-dimensional cubes $K_i^{\delta_k}$ with edge length $(2k+1)^{-1}$. For each $h \in \mathbb{N}$, a lattice is then defined in $K_i^{\delta_k}$ by $(2h+1)^d$ equidistant points. These points are numbered based on a specific scheme and denoted by $v^i(j)$, $j = 1, \ldots, (2h+1)^d$ (the construction of this scheme is outlined in Kopperschmidt 2005, p. 156 and described in Section A.2 of Jakubzik 2017, pp. 148–151). Note that $v^i(j) \in [0,1]^d$ by definition.

3. Kopperschmidt uses condition ($\widetilde{\text{ii}}$) to obtain (cf. Kopperschmidt 2005, p. 156):

$$\mathbb{E}\left\|X_n(v^i(l)) - X_n(v^i(j))\right\|^2 \leq \underbrace{\nu\|v^i(l) - v^i(j)\|^\alpha}_{\in\mathbb{R}}$$

$$\stackrel{?}{=} \underbrace{\left(\sum_{p=j+1}^{l} \nu^{\frac{1}{\alpha}} \left(v^i(p) - v^i(p-1) \right) \right)^{\alpha}}_{\in \mathbb{R}^d} .$$

While the use of a telescoping sum leads to the desired result in the one-dimensional case, the above equality is no longer valid in higher dimensions (as indicated by the question mark). The error continuous in the subsequent estimates, which as a result cannot be interpreted in any meaningful way.

4. The proof in the case $d = 1$ indicates that the tightness criterion can essentially be regarded as an application of Billingsley's Theorem 12.2 (cf. Billingsley 1968, p. 96). With the requirements for this theorem no longer satisfied for $d > 1$, the further proof cannot be transferred and fails.

In an attempt to find a way to apply Billingsley's Theorem 12.2, we worked on a multi-dimensional generalization in Jakubzik 2017, pp. 70–73 that, while fixing Kopperschmidt's mistake, also turned out to be erroneous. We would like to emphasize that even though we have identified the proof as incorrect, we have not refuted the statement itself, nor is that the focus of this dissertation. However, we conjecture that condition ($\widetilde{\text{ii}}$) is generally not sufficient to guarantee tightness. We elaborate on this belief by reviewing some literature on Kolmogorov's tightness criterion:

(i) Bickel and Wichura 1971 & Lachout 1988: The article Bickel and Wichura 1971 serves as "an important reference for multiparameter processes" (Kopperschmidt and Stute 2013, p. 1289). The authors "prove multidimensional analogues of Theorems 12.5 and 15.6 of Billingsley 1968" (Bickel and Wichura 1971, p. 1656). Theorem 12.5 is a strengthening of Theorem 12.1 (cf. Billingsley 1968, p. 98), which in turn is needed to prove the important Theorem 12.2 that Kopperschmidt intended to use. Together, these theorems provide "several fluctuation inequalities for sums of random variables" (Bickel and Wichura 1971, p. 1656). In this way, Theorem 1 of Bickel and Wichura 1971, p. 1658 can be considered as a potential starting point for a generalization of Kolmogorov's tightness criterion, Theorem 12.3 in Billingsley 1968. The requirement $(X, \mu) \in \mathcal{C}(\beta, \gamma)$ of this theorem means that (X, μ) satisfies the condition (β, γ) (see Bickel and Wichura 1971, p. 1658), which in its moment version means that

$$\mathbb{E}\left(m(B, C)\right)^{\gamma} \leq \mu(B \cup C)^{\beta}, \tag{3.63}$$

where $m(B, C) = \min\{|X(B)|, |X(C)|\}$, $X(B)$ is the increment of X around a "block" B, μ is a positive measure and $\beta > 1$, $\gamma > 0$ (for details, see Bickel and Wichura 1971, p. 1658). Of particular interest is that the parameter β on the right-hand side of this inequality does not depend on the dimension d of the blocks B and C, which seems to indicate the existence of a condition similar to ($\widetilde{\text{ii}}$) that is independent of d as well. Moreover, Kopperschmidt and Stute suggest that "simple increments suffice to guarantee tightness" due to the continuity of the considered processes (cf. Kopperschmidt and Stute 2013, p. 1289). Thus, if we were to transfer the condition from Equation (3.63) to our situation, the following formulation would be conceivable (we use the Lebesgue measure λ instead of an arbitrary positive measure μ, which is consistent with the proof of Theorem 1, see Bickel and Wichura

1971, p. 1659):

$$\mathbb{E}\left\|X_n(x)-X_n(y)\right\|^2 \le \nu\lambda\left(B_{x,y}\right)^{\alpha}\ , \text{ for all } x,y\in[0,1]^d \text{ with } x_i<y_i,\ i=1,\dots,d.$$

Here, $B_{x,y}$ is the uniquely determined d-dimensional cuboid (or "block") defined by the vertices x and y. At first glance, this inequality appears to be consistent with condition $\widetilde{\text{(ii)}}$ and to be the multi-dimensional counterpart of Billingsley's moment condition (12.51) envisioned by Kopperschmidt and Stute. However, this is not the case because the dependence on the dimension d is incorporated in the measure. For a simple explanation of this, let us assume that $B_{x,y}$ is a cube with edge length δ (e.g., when $x=(0,\dots,0)^\top$ and $y=(\delta,\dots,\delta)^\top$). Then, $\|x-y\|=\sqrt{d}\delta$, so that the upper bound from condition $\widetilde{\text{(ii)}}$ would be

$$\nu\left(\sqrt{d}\delta\right)^{\alpha}\propto\delta^{\alpha}\,,$$

but on the other hand, $\lambda(B_{x,y})=\delta^d$, and hence

$$\nu\lambda\left(B_{x,y}\right)^{\alpha}=\nu\left(\delta^d\right)^{\alpha}\propto\delta^{d\alpha}\,.$$

Nevertheless, this is consistent with our condition (ii) from Theorem B.2.6, where for $\beta=d(\alpha-1)>0$ (note that $\alpha>1$) we obtain

$$\nu\left\|x-y\right\|^{d+\beta}\propto\delta^{d+\beta}=\delta^{d\alpha}\,.$$

Because of the equivalence of norms as well as the additivity of measures, this proportionality does not depend on the specific choices. Overall, the compatibility of Theorem B.2.6 with Bickel and Wichura 1971 reinforces our conviction that the dependence on the dimension cannot be neglected.
The second article, Lachout 1988, further generalizes the results of Bickel and Wichura 1971 to derive tightness criteria for multiparameter processes (as we have already indicated). Here, again, the dependence on the dimension of the parameter space is hidden in the measures used.

(ii) Totoki 1962, Kunita 1986 & Kunita 1990: The tightness criterion considered within this thesis, Theorem B.2.6, can be seen as a special case of the Theorem 1.4.7 from Kunita 1990, p. 38. A complete proof of this generalization of Kolmogorov's tightness criterion can also be found there (see Kunita 1990, pp. 31–35, 38–39). A related result arising from a combination with Kolmogorov-Chentsov's criterion - which can be looked up in Kunita 1990, pp. 41–42 as Exercise 1.4.19 - is found in Kunita 1986, p. 311, see Theorem 1.1 for the special case of a stochastic flow (note the remarkable similarity between Equation (1.4) there and Equation (B.25) from the appendix of this dissertation). The article Kunita 1986 in turn refers to Totoki 1962, which to our knowledge contains the earliest account of a tightness criterion similar to the one given in Theorem B.2.6 (see Theorem 1, Totoki 1962, p. 183). As with the previous criteria, this theorem shows a dependence on the dimension of the parameter set.

All the tightness criteria studied in our brief literature overview share the same dimensional dependence. Conversely, we could not find any evidence for a tightness criterion independent of the dimension of the parameter space. Altogether, this consolidates our belief that such a criterion might not exist or has yet to be proven.

3.4. Application to the Basquin Load Sharing Model With Multiplicative Damage Accumulation

In Section 3.3, we presented a corrected proof for the asymptotic normality of the minimum distance estimator, but for which the initial assumptions of Kopperschmidt 2005 had to be tightened. The main motivation for finding this proof is to establish the asymptotic distribution of that estimator in the context of the models considered within this dissertation, which so far could only be implied by simulation studies. Therefore, the objective of the current section is to verify that the Basquin load sharing model with multiplicative damage accumulation meets the assumptions of Theorem 3.12 - otherwise, the newly found proof would be useless for our applications. In particular, we need to verify that the preconditions stated in Assumptions 3.5 and 3.8 are fulfilled.
In Assumptions 3.5, conditions (C1) and (C2) were retained from Kopperschmidt and Stute 2013, while the uniform boundedness condition (C3) is a strengthening of the locally uniform integrability demanded in ($\widetilde{\text{C}}$3). However, the Basquin load sharing model with multiplicative damage accumulation satisfies (C3) if $\Theta \subset \mathbb{R}_+^2 \times [-1+\varepsilon, \infty)$, $0 < \varepsilon \leq 1$, because in our setting the experimental runs were observed only up to a random time $\tau_j \leq \tau$ or up to a random number of failed components $C_j \leq I$, whichever occurred first (cf. Equation (2.40) from Corollary 2.17 and choose $p = q = r = 0$). Recall that Assumption 2.3 on the random covariates implied that

$$\tau_1, \tau_2, \ldots \overset{\text{i.i.d.}}{\sim} \mathbb{P}^{\tau_0} \qquad \text{and} \qquad C_1, C_2, \ldots \overset{\text{i.i.d.}}{\sim} \mathbb{P}^{C_0}$$

for appropriate probability measures on $\mathcal{B}(\mathcal{I})$ and $2^{\{0,1,\ldots,I\}}$, respectively. In addition, we allowed the systems to be exposed to different initial stress levels s_j, where

$$s_1, s_2 \ldots \overset{\text{i.i.d.}}{\sim} \mathbb{P}^{s_0}$$

for some probability measure $\mathbb{P}^{s_0}$ on $\mathcal{B}([0, \infty))$. In this section, we demand that

$$\forall j \in \mathbb{N}: \quad \tau_j, C_j, s_j \quad \text{are stochastically independent.}$$

As seen here, the Basquin load sharing model with multiplicative damage accumulation satisfies Assumptions 3.5 if and only if it also satisfies the corresponding assumptions of Kopperschmidt and Stute 2013. Consequently, we first focus on Assumptions 3.8, for which we will require a technical preliminary theorem, and return to Assumptions 3.5 later. By placing minor assumptions on the supports of the above probability measures, we infer that the expected cumulative intensities in the Basquin load sharing model with multiplicative damage accumulation are strictly increasing: We suppose, on the hand, that the experiments are not systematically stopped at the beginning of the observation period and, on the other hand, that the probability for a critical number of component failures of at least one is positive. Finally, we require that strictly positive initial stress levels can be realized.

Theorem 3.28 (Strict Monotonicity of the Expected Cumulative Intensities in the Basquin Load Sharing Model with Multiplicative Damage Accumulation)**.**
If $0 < t \in \operatorname{supp}(\mathbb{P}^{\tau_0})$ *and* $\operatorname{supp}(\mathbb{P}^{s_0}) \neq \{0\} \neq \operatorname{supp}(\mathbb{P}^{C_0})$*, the expected cumulative intensity* $\mathbb{E}^{\times\mathrm{D}}\Lambda_\theta$ *of the Basquin load sharing model with multiplicative damage accumulation is strictly increasing on* $[0,t]$ *for each* $\theta \in \Theta$ *with* $\theta_1 \neq 0$*. In particular, if* $\tau \in \operatorname{supp}(\mathbb{P}^{\tau_0})$*, then it is strictly increasing on the entire interval* $\mathcal{I}$*.*

Proof. We begin with a technical note: By Assumption 2.3, we require that an intrinsic filtration be considered that satisfies $\sigma(\tau_0) \vee \sigma(C_0) \vee \sigma(s_0) \subset \mathcal{G}_0$, which implies that all information about τ_0, C_0 and s_0 is available at the beginning of the experiment. This permits us later to condition on events given in terms of τ_0, C_0 or s_0 in order to relate the cumulative conditional hazard function to the compensator, compare Lemma A.32. We return to this remark towards the end of the proof.
Let us recall the conditional intensity function of the model ${}^{\times}$D, the Basquin load sharing model with multiplicative damage accumulation given in Definition 2.8. Adapted to the notation of this theorem, it reads as follows:

$$^{\times\mathrm{D}}\lambda_\theta(t) := \theta_1 \left(s_0 \frac{I}{I - N_{t-}} \right)^{\theta_2} A(t)^{\theta_3} \cdot \mathbb{1}_{\left\{N_{t-} < C_0\right\} \cap \{t \le \tau_0\}} .$$

If we disregard the indicator function for the moment, this intensity is strictly positive for all $t > 0$ as long as $\theta_1 \neq 0$ and $s_0 \neq 0$, while at $t = 0$ the intensity is equal to 0 due to the vanishing damage accumulation term, $A(0) = 0$ (leaving aside the pathological cases $\theta_3 \le 0$, where the following reasoning applies analogously). Since the cumulative intensity ${}^{\times\mathrm{D}}\Lambda_\theta$ is defined as the integral over the above intensity function, its monotonicity would follow immediately from this positivity. However, the indicator function is crucial, so we opt for the law of total expectation (see Lemma B.4.9) to access events where we can control its value. To formally prove the monotonicity of the expected cumulative intensity on $[0,t]$, let $t_1, t_2 \in [0,t]$ with $t_1 < t_2$. Per definition, any intensity is non-negative, so the (expected) cumulative intensity $\mathbb{E}^{\times\mathrm{D}}\Lambda_\theta$ is non-decreasing by default. It is therefore sufficient to find $t_1 < \tilde{t}_1 < \tilde{t}_2 < t_2$ for which

$$\mathbb{E}^{\times\mathrm{D}}\Lambda_\theta\left(\tilde{t}_1\right) < \mathbb{E}^{\times\mathrm{D}}\Lambda_\theta\left(\tilde{t}_2\right)$$

is satisfied. For this reason, we can assume without loss of generality that $t_1, t_2 \in (0,t)$ holds (i.e., the boundary points need not be examined separately). From here on, we will omit the model indicator ${}^{\times}$D for ease of reading, and proceed to compute:

$$\begin{aligned}
\mathbb{E}\Lambda_\theta\left(t_2\right) - \mathbb{E}\Lambda_\theta\left(t_1\right) &= \mathbb{E}\left[\Lambda_\theta(t_2) - \Lambda_\theta(t_1)\right] = \mathbb{E}\left[\int_{t_1}^{t_2} \lambda_\theta(t)\,\mathrm{d}t\right] \qquad \Big|\text{law of total expectation} \\
&= \mathbb{E}\left[\int_{t_1}^{t_2} \lambda_\theta(t)\,\mathrm{d}t \,\Big|\, \{N_{t_2^-} < C_0\} \cap \{\tau_0 \ge t_2\} \cap \{s_0 > 0\}\right] \\
&\qquad \cdot \mathbb{P}\Big(\underbrace{\{N_{t_2^-} < C_0\} \cap \{\tau_0 \ge t_2\} \cap \{s_0 > 0\}}_{\supset \left\{N_{t_2} < C_0\right\} \cap \{\tau_0 > t_2\} \cap \{s_0 > 0\}} \Big) \\
&\quad + \underbrace{\mathbb{E}\left[\int_{t_1}^{t_2} \lambda_\theta(t)\,\mathrm{d}t \,\Big|\, \{N_{t_2^-} < C_0\}^{\complement} \cup \{\tau_0 \ge t_2\}^{\complement} \cup \{s_0 > 0\}^{\complement}\right]}_{\ge 0}
\end{aligned}$$

$$
\begin{aligned}
&\cdot \mathbb{P}\left(\{N_{t_2^-} < C_0\}^{\complement} \cup \{\tau_0 \geq t_2\}^{\complement} \cup \{s_0 > 0\}^{\complement}\right) \\
\geq\ & \mathbb{E}\left[\int_{t_1}^{t_2} \lambda_\theta(t)\,\mathrm{d}t \,\middle|\, \{N_{t_2^-} < C_0\} \cap \{\tau_0 \geq t_2\} \cap \{s_0 > 0\}\right] \\
&\cdot \mathbb{P}\Big(\{N_{t_2} < C_0\} \cap \{\tau_0 > t_2\} \cap \{s_0 > 0\}\Big)\,.
\end{aligned}
\tag{3.64}
$$

For $\omega \in \{N_{t_2^-} < C_0\} \cap \{\tau_0 \geq t_2\} \cap \{s_0 > 0\}$, the indicator function in λ_θ is identically 1 on $[t_1, t_2]$. Consequently, as $s_0 \neq 0$ is satisfied as well, the intensity is strictly positive here and so is the corresponding integral. The monotonicity of the expectation then yields that the first factor in Equation (3.64) is greater than 0, provided that the event $\{N_{t_2^-} < C_0\} \cap \{\tau_0 \geq t_2\} \cap \{s_0 > 0\}$ is not a $\mathbb{P}$-null set. Verifying this conveniently coincides with the study of the second factor, so we have:

$$
\begin{aligned}
& \mathbb{E}\left[\int_{t_1}^{t_2} \lambda_\theta(t)\,\mathrm{d}t \,\middle|\, \{N_{t_2^-} < C_0\} \cap \{\tau_0 \geq t_2\} \cap \{s_0 > 0\}\right] \\
& \qquad \cdot \mathbb{P}\Big(\{N_{t_2} < C_0\} \cap \{\tau_0 > t_2\} \cap \{s_0 > 0\}\Big) \quad > 0 \\
\iff \quad & \qquad \mathbb{P}\Big(\{N_{t_2} < C_0\} \cap \{\tau_0 > t_2\} \cap \{s_0 > 0\}\Big) \quad > 0\,,
\end{aligned}
$$

which would in turn imply $\mathbb{E}\Lambda_\theta\,(t_2) - \mathbb{E}\Lambda_\theta\,(t_1) > 0$ by virtue of Equation (3.64). Due to the independence of τ_0, C_0 and s_0 we obtain:

$$
\begin{aligned}
&\mathbb{P}\Big(\{N_{t_2} < C_0\} \cap \{\tau_0 > t_2\} \cap \{s_0 > 0\}\Big) \\
&\quad \geq \mathbb{P}\Big(\{N_{t_2} < C_0\} \cap \{\tau_0 > t_2\} \cap \{s_0 > 0\} \cap \{C_0 \geq 1\}\Big) \\
&\quad \geq \mathbb{P}\Big(\{N_{t_2} = 0\} \cap \{\tau_0 > t_2\} \cap \{s_0 > 0\} \cap \{C_0 \geq 1\}\Big) \\
&\quad = \mathbb{P}\left(N_{t_2} = 0 \,\middle|\, \{\tau_0 > t_2\} \cap \{s_0 > 0\} \cap \{C_0 \geq 1\}\right) \\
&\qquad\qquad \cdot \mathbb{P}\Big(\{\tau_0 > t_2\} \cap \{s_0 > 0\} \cap \{C_0 \geq 1\}\Big) \\
&\quad = \mathbb{P}\left(N_{t_2} = 0 \,\middle|\, \{\tau_0 > t_2\} \cap \{s_0 > 0\} \cap \{C_0 \geq 1\}\right) \\
&\qquad\qquad \cdot \mathbb{P}\left(\tau_0 > t_2\right) \cdot \mathbb{P}\left(s_0 > 0\right) \cdot \mathbb{P}\left(C_0 \geq 1\right)\,.
\end{aligned}
$$

Since $t \in \operatorname{supp}\left(\mathbb{P}^{\tau_0}\right)$, $\mathbb{P}\left(\tau_0 \in U\right) > 0$ for any open neighbourhood U of t. In particular, $\mathbb{P}\left(\tau_0 > t_2\right) > 0$, as $(t_2, \tau]$ is an open set (with respect to the subspace topology on $\mathcal{I} \subset \mathbb{R}$) containing t by assumption. Additionally, $\mathbb{P}\left(C_0 \geq 1\right) > 0$, as otherwise $\mathbb{P}\left(C_0 = 0\right) = 1$ would imply $\operatorname{supp}(\mathbb{P}^{C_0}) = \{0\}$. Lastly, $\mathbb{P}\left(s_0 > 0\right) > 0$ with the same argument, so it suffices to show that $\mathbb{P}\left(N_{t_2} = 0 \,\middle|\, \{\tau_0 > t_2\} \cap \{s_0 > 0\} \cap \{C_0 \geq 1\}\right) > 0$ to complete the proof. Recalling the technical remark at the beginning of the proof, we note that $\{\tau_0 > t_2\} \cap \{s_0 > 0\} \cap \{C_0 \geq 1\} \in \mathcal{G}_0$. We further conclude by revisiting Summary 1 and incorporating the ideas of the hazard transformation given in Theorem A.46:

$$
\begin{aligned}
\mathbb{P}\left(N_{t_2} = 0 \,\middle|\, \mathcal{G}_0\right) &= \mathbb{P}\left(T_1 \geq t_2 \,\middle|\, \mathcal{G}_0\right) = S_1\left(t_2 \,\middle|\, \mathcal{G}_0\right) \\
&= \exp\left(-H_1\left(t_2 \,\middle|\, \mathcal{G}_0\right)\right) && \Big|\text{apply Lemma A.32} \\
&= \exp\left(-\int_0^{t_2} \theta_1 s_0^{\theta_2} \left(\frac{1}{\tau}\int_0^t s_0\,\mathrm{d}u\right)^{\theta_3} \cdot \mathbb{1}_{\{t<\tau_0\}} \cdot \mathbb{1}_{\{0<C_0\}}\,\mathrm{d}t\right)
\end{aligned}
$$

$$= \begin{cases} 1\,, & C_0 = 0\,, \\ \exp\left(-\frac{\theta_1 s_0^{\theta_2+\theta_3}}{\tau^{\theta_3}(1+\theta_3)} \min\{t_2, \tau_0\}^{1+\theta_3}\right)\,, & C_0 > 0\,, \end{cases} \tag{3.65}$$

where we applied Lemma A.32 to the one-point process

$$(t, \omega) \longrightarrow \mathbb{1}_{(-\infty,t]}\left(T_1(\omega)\right)$$

to transition from the cumulative conditional hazard function to the associated compensator. As this compensator coincides with ${}^{\times\mathrm{D}}\Lambda_\theta$ on $[0, T_1)$, we directly expressed it as an integrated intensity in order to avoid notational confusion. Notably, Equation (3.65) can also be recognized as the survival function of a Weibull distribution given $\tau_0 = \infty$, $C_0 > 0$ and any $s_0 > 0$. On $\{\tau_0 > t_2\} \cap \{s_0 > 0\} \cap \{C_0 \geq 1\}$, we have

$$\mathbb{P}\left(N_{t_2} = 0 \,\middle|\, \mathcal{G}_0\right) = \exp\left(-\frac{\theta_1 s_0^{\theta_2+\theta_3}}{\tau^{\theta_3}\left(1+\theta_3\right)} t_2^{1+\theta_3}\right) > 0\,.$$

The tower property then yields:

$$\begin{aligned}
&\mathbb{P}\left(N_{t_2} = 0 \,\middle|\, \{\tau_0 > t_2\} \cap \{s_0 > 0\} \cap \{C_0 \geq 1\}\right) \\
&\quad = \mathbb{E}\left(\mathbb{P}\left(N_{t_2} = 0 \,\middle|\, \mathcal{G}_0\right) \,\middle|\, \{\tau_0 > t_2\} \cap \{s_0 > 0\} \cap \{C_0 \geq 1\}\right) \\
&\quad = \mathbb{E}\left(\exp\left(-\frac{\theta_1 s_0^{\theta_2+\theta_3}}{\tau^{\theta_3}\left(1+\theta_3\right)} t_2^{1+\theta_3}\right) \,\middle|\, \{\tau_0 > t_2\} \cap \{s_0 > 0\} \cap \{C_0 \geq 1\}\right) > 0\,,
\end{aligned}$$

which completes the proof. □

The above theorem lays the foundation for two of the following results, which ultimately establish the identifiability condition (C1) as well as the positive definiteness of the standardizing matrix $\Phi_0\left(\theta^*\right)$ by virtue of condition (A4) and Lemma 3.18.

3.4.1. Verifying the Assumptions 3.8 for Asymptotic Normality

We have already indicated earlier that we will leave Assumptions 3.5 for the time being and skip ahead to Assumptions 3.8 instead. On the one hand, this is due to the fact that the main differences compared to the conditions formulated by Kopperschmidt and Stute are present in these assumptions. On the other hand, for the proof of the identifiability condition (C1) we can later revert to the methods used to verify condition (A4), whereas the converse would require more effort. Under the conditions of Corollary 2.17 from Section 2.4, we immediately observe that conditions (A1), (A2) and (A3) are satisfied:

(A1) From the proof of Lemma 2.15 we know that the conditional intensity function ${}^{\times\mathrm{D}}\lambda_\cdot(t)$ as a function of $\theta \in \Theta$ for fixed $t \in \mathcal{I}$ is infinitely often continuously differentiable. The cumulative intensity function inherits the smoothness of the conditional intensity function whenever we are allowed to interchange the derivative with respect to θ and the integral with respect to t. Since this applies for derivatives of arbitrary orders according to Corollary 2.17, it holds:

$$^{\times\mathrm{D}}\Lambda_\cdot(t) \in C^\infty(\Theta)\,,$$

as long as the restrictions placed on Θ are in force. Condition (A1) follows trivially.

(A2) Since ${}^{\times\mathrm{D}}\Lambda$ was modelled as a cumulative intensity, continuity ensues immediately from the integral representation. Note that for this property we once again take advantage of the fact that we are allowed to interchange integration and differentiation up to arbitrary orders by virtue of Corollary 2.17.

(A3) This condition is directly implied by Equation (2.40) from Corollary 2.17.

It only remains to verify condition (A4), which is crucial to derive the positive definiteness of the standardizing matrix $\Phi_0(\theta^*)$ according to Lemma 3.18. We require again the existence of suitable integrable majorants for the partial derivatives of the conditional intensity function, as we have previously done in Corollary 2.17. We will here consider only the case of deterministic initial stress levels. Moreover, we demand that $\operatorname{supp}(\mathbb{P}^{\tau_0})$ is bounded below by some positive constant.

Theorem 3.29 (Positive Definiteness of the Standardizing Matrix in the Basquin Load Sharing Model With Multiplicative Damage Accumulation)**.**
Assume that the initial stress level is deterministic, so that $\mathbb{P}^{s_0} = \delta_s$, the Dirac measure centred on some $s > 0$. Furthermore, suppose that $\min\operatorname{supp}(\mathbb{P}^{\tau_0}) > 0$ and that $\operatorname{supp}(\mathbb{P}^{C_0}) \cap \{2, \ldots, I\} \neq \emptyset$. Then, under the preconditions of Corollary 2.17, the standardizing matrix $\Phi_0(\theta)$ is positive definite for all $\theta \in \Theta$.

Proof. By Corollary 2.17, differentiation with respect to $\theta \in \Theta$ and integration with respect to $t \in \mathcal{I}$ are interchangeable for the cumulative intensity ${}^{\times\mathrm{D}}\Lambda_\theta$. Therefore,

$$\begin{aligned} v^\top \frac{\mathrm{d}}{\mathrm{d}\theta}\, {}^{\times\mathrm{D}}\Lambda_\theta(t)^\top &= \int_0^t v^\top \frac{\mathrm{d}}{\mathrm{d}\theta}\, {}^{\times\mathrm{D}}\lambda_\theta(u)^\top \,\mathrm{d}u \\ &= \int_0^t \left(\frac{v_1}{\theta_1} + v_2 \ln B(u) + v_3 \ln A(u)\right) \cdot {}^{\times\mathrm{D}}\lambda_\theta(u)\,\mathrm{d}u\,, \end{aligned}$$

compare Equations (B.2) and (B.4) in Appendix B.1 for the partial derivatives of ${}^{\times\mathrm{D}}\lambda_\theta(t)$. We distinguish two cases. First, let $v_3 \neq 0$. We exploit that $|\ln A(u)| \to \infty$ as $u \to 0$, while $\left|\frac{v_1}{\theta_1} + v_2 \ln B(u)\right|$ is bounded for all $u \in \mathcal{I}$. More precisely we have:

$$\left|\frac{v_1}{\theta_1} + v_2 \ln B(u)\right| \le \frac{|v_1|}{\theta_1} + |v_2| \max\{-\ln(s), \ln(sI)\} =: C_{v,\theta}\,,$$

where the constant $C_{v,\theta}$ does not depend on the particular realization ω. Moreover, for $u \le \frac{\tau}{sI}$ (so that $A(u) \le 1$ with probability 1),

$$|\ln A(u)| = \left|\ln\Bigg(\underbrace{\frac{1}{\tau}\int_0^u s\frac{I}{I - N_{x^-}}\,\mathrm{d}x}_{\le \frac{usI}{\tau} \le 1}\Bigg)\right| \ge -\ln\left(\frac{usI}{\tau}\right),$$

which again does not depend on ω at all. The reverse triangle inequality yields:

$$\begin{aligned} \left|\frac{v_1}{\theta_1} + v_2 \ln B(u) + v_3 \ln A(u)\right| &\ge |v_3|\,|\ln A(u)| - \left|\frac{v_1}{\theta_1} + v_2 \ln B(u)\right| \\ &\ge -|v_3| \ln\left(\frac{usI}{\tau}\right) - C_{v,\theta} \to \infty \quad (u \to 0)\,, \end{aligned}$$

where for sufficiently small u the sign depends entirely on v_3. Consequently, for any $C > 0$, there exists $t_{v,\theta} \in \mathcal{I}$ such that

$$\begin{aligned}
&\text{either} \qquad \frac{v_1}{\theta_1} + v_2 \ln B(u) + v_3 \ln A(u) \leq -C \qquad\qquad \text{for all } 0 < u \leq t_{v,\theta}, \\
&\text{or} \qquad \frac{v_1}{\theta_1} + v_2 \ln B(u) + v_3 \ln A(u) \geq C \qquad\qquad \text{for all } 0 < u \leq t_{v,\theta}.
\end{aligned}$$

Since the conditional intensity is non-negative for all $u \in \mathcal{I}$, in the first case (the second case can be treated analogously) we then obtain by the monotonicity of the integral:

$$\begin{aligned}
v^\top \frac{\mathrm{d}}{\mathrm{d}\theta} {}^{\times\mathrm{D}}\Lambda_\theta \left(t_{v,\theta}\right)^\top &= \int_0^{t_{v,\theta}} \left(\frac{v_1}{\theta_1} + v_2 \ln B(u) + v_3 \ln A(u) \right) \cdot {}^{\times\mathrm{D}}\lambda_\theta(u) \,\mathrm{d}u \\
&\leq -C \int_0^{t_{v,\theta}} {}^{\times\mathrm{D}}\lambda_\theta(u) \,\mathrm{d}u = -C \, {}^{\times\mathrm{D}}\Lambda_\theta \left(t_{v,\theta}\right) .
\end{aligned}$$

This inequality holds for all ω and hence carries over to the expectation, that is,

$$v^\top \mathbb{E} \frac{\mathrm{d}}{\mathrm{d}\theta} {}^{\times\mathrm{D}}\Lambda_\theta \left(t_{v,\theta}\right)^\top \leq -C \, \mathbb{E}^{\times\mathrm{D}}\Lambda_\theta \left(t_{v,\theta}\right) < 0 \,,$$

where for the last inequality we applied Theorem 3.28. By the continuity of $\mathbb{E}\frac{\mathrm{d}}{\mathrm{d}\theta} {}^{\times\mathrm{D}}\Lambda_\theta$ as a function of $t \in \mathcal{I}$, there then exists $\varepsilon > 0$ such that

$$v^\top \mathbb{E} \frac{\mathrm{d}}{\mathrm{d}\theta} {}^{\times\mathrm{D}}\Lambda_\theta \left(t\right)^\top < 0 \quad \text{for all } t \in (t_{v,\theta} - \varepsilon, t_{v,\theta} + \varepsilon) =: B_v.$$

This Borel set B_v then satisfies condition (A4), as Theorem 3.28 yields that B_v has positive $\mathbb{E}^{\times\mathrm{D}}\Lambda_\theta$-measure. Therefore, the positive definiteness of $\Phi_0(\theta)$ ensues by virtue of Lemma 3.18.

There remains the case that $v_3 = 0$. As before, we get:

$$v^\top \frac{\mathrm{d}}{\mathrm{d}\theta} {}^{\times\mathrm{D}}\Lambda_\theta(t)^\top = \int_0^t \left(\frac{v_1}{\theta_1} + v_2 \ln B(u) \right) \cdot {}^{\times\mathrm{D}}\lambda_\theta(u) \,\mathrm{d}u \,.$$

We need to prove that there exists $t \in \mathcal{I}$ such that

$$0 \neq v^\top \mathbb{E} \frac{\mathrm{d}}{\mathrm{d}\theta} {}^{\times\mathrm{D}}\Lambda_\theta(t)^\top = \mathbb{E} \left(\int_0^t \left(\frac{v_1}{\theta_1} + v_2 \ln B(u) \right) \cdot {}^{\times\mathrm{D}}\lambda_\theta(u) \,\mathrm{d}u \right) .$$

For $v_2 = 0$ (and thus $v_1 \neq 0$), this is a direct consequence of Theorem 3.28. For $v_2 \neq 0$, division by v_2 yields:

$$0 \neq \mathbb{E} \left(\int_0^t \left(\frac{v_1}{v_2 \theta_1} + \ln B(u) \right) \cdot {}^{\times\mathrm{D}}\lambda_\theta(u) \,\mathrm{d}u \right) . \tag{3.66}$$

When the factor $\frac{v_1}{v_2\theta_1} + \ln B(u)$ is almost surely positive, we can immediately infer the positivity of the above expectation from Theorem 3.28. Similarly, the negativity of the expectation ensues when $\frac{v_1}{v_2\theta_1} + \ln B(u)$ is almost surely negative. For certain s, however, this factor can take both negative and positive values, in which case we show that the behavior near $t = 0$ is predominantly determined by the event $\{T_1 > t\}$. In total, we distinguish three cases, noting that the randomness of s could interfere with subsequent estimates. From here on, we thus need the condition that $\mathbb{P}^{s_0} = \delta_s$ holds, whereas the previous part would be provable analogously for any measure with bounded support.

(i) $s > \exp\left(-\frac{v_1}{v_2\theta_1}\right)$:

In this case, we consider the first factor of the integral from Equation (3.66) and show that it is strictly positive for all $u \in \mathcal{I}$. With probability one, we have:

$$\begin{aligned}
\frac{v_1}{v_2\theta_1} + \ln B(u) &= \frac{v_1}{v_2\theta_1} + \ln\left(s\frac{I}{I - N_{u^-}}\right) \\
&\geq \frac{v_1}{v_2\theta_1} + \ln(s) \\
&> \frac{v_1}{v_2\theta_1} + \ln\left(\exp\left(-\frac{v_1}{v_2\theta_1}\right) I\right) = 0\,, \qquad (3.67)
\end{aligned}$$

where we substituted the lower bound for s and exploited the monotonicity of the logarithm. Hence, by virtue of Theorem 3.28, we obtain for each $t > 0$:

$$\mathbb{E}\left(\int_0^t \left(\frac{v_1}{v_2\theta_1} + \ln B(u)\right) \cdot {}^{\times\mathrm{D}}\lambda_\theta(u)\,\mathrm{d}u\right) \geq \underbrace{\left(\frac{v_1}{v_2\theta_1} + \ln(s)\right)}_{>0 \text{ by Eq. (3.67)}} \underbrace{\mathbb{E}\left({}^{\times\mathrm{D}}\Lambda_\theta(t)\right)}_{>0} > 0\,.$$

(ii) $s = \exp\left(-\frac{v_1}{v_2\theta_1}\right)$:

We note that for $u \leq T_1$ (i.e., when no component has yet failed),

$$\frac{v_1}{v_2\theta_1} + \ln B(u) = \frac{v_1}{v_2\theta_1} + \ln s = 0\,,$$

while for $u > T_1$, the monotonicity of $B(u)$ implies that

$$\frac{v_1}{v_2\theta_1} + \ln B(u) > 0\,.$$

Accordingly, the integrand in Equation (3.66) is non-negative. Moreover, for any given $t \in \mathcal{I}$, it is strictly positive on $\{t > T_1\} \cap \{t < \tau_0\} \cap \{1 < C_0\}$. Therefore, we can conclude that the expectation is positive if the set $\{t > T_1\} \cap \{t < \tau_0\} \cap \{1 < C_0\}$ has positive $\mathbb{P}$-measure. We have:

$$\begin{aligned}
&\mathbb{P}\left(\{t > T_1\} \cap \{t < \tau_0\} \cap \{1 < C_0\}\right) \\
&\quad = \mathbb{P}\big(t > T_1 \,\big|\, \underbrace{\{t < \tau_0\} \cap \{1 < C_0\}}_{\in \mathcal{G}_0}\big) \cdot \mathbb{P}\left(\{t < \tau_0\} \cap \{1 < C_0\}\right)\,.
\end{aligned}$$

The conditional probability is positive for each $t > 0$, as can be seen by considering Equation (3.65) once again. Additionally, the independence assumption yields

$$\mathbb{P}\left(\{t < \tau_0\} \cap \{1 < C_0\}\right) = \mathbb{P}\left(t < \tau_0\right) \cdot \mathbb{P}\left(1 < C_0\right)\,, \qquad (3.68)$$

where the latter probability is not equal to 0 because $\operatorname{supp}\left(\mathbb{P}^{C_0}\right) \neq \{0, 1\}$ and the former probability also becomes positive for sufficiently small t as otherwise $\mathbb{P}^{\tau_0} = \delta_0$ would follow. Hence,

$$\mathbb{P}\left(\{t > T_1\} \cap \{t < \tau_0\} \cap \{1 < C_0\}\right) > 0\,.$$

(iii) $s < \exp\left(-\frac{v_1}{v_2\theta_1}\right)$:

This case is the most difficult, because here the sign of the factor $\frac{v_1}{v_2\theta_1} + \ln B(u)$ may change as the number of failures (and hence the load sharing term $B(u)$) increases. Instead of the expectation, we first consider the conditional expectation under $\mathcal{G}_0$ in Equation (3.66), similar to the proof of the previous theorem. Here, $\sigma(s_0)$ is the trivial σ-algebra, so only $\sigma(\tau_0) \vee \sigma(C_0) \subset \mathcal{G}_0$ needs to be fulfilled. We compute:

$$\begin{aligned}
&\mathbb{E}\left(\int_0^t \left(\frac{v_1}{v_2\theta_1} + \ln B(u)\right) \cdot {}^{\times\mathrm{D}}\lambda_\theta(u)\,\mathrm{d}u \,\middle|\, \mathcal{G}_0\right) \\
&\quad = \mathbb{E}\left(\int_0^t \left(\frac{v_1}{v_2\theta_1} + \ln B(u)\right) \cdot {}^{\times\mathrm{D}}\lambda_\theta(u)\,\mathrm{d}u \cdot \mathbb{1}_{\{T_1>t\}} \,\middle|\, \mathcal{G}_0\right) \\
&\qquad + \mathbb{E}\left(\int_0^t \underbrace{\left(\frac{v_1}{v_2\theta_1} + \ln B(u)\right)}_{\leq \frac{v_1}{v_2\theta_1} + \ln(sI)} \cdot {}^{\times\mathrm{D}}\lambda_\theta(u)\,\mathrm{d}u \cdot \mathbb{1}_{\{T_1\leq t\}} \,\middle|\, \mathcal{G}_0\right) \\
&\quad \leq \left(\frac{v_1}{v_2\theta_1} + \ln(s)\right) \cdot \mathbb{E}\left(\int_0^t {}^{\times\mathrm{D}}\lambda_\theta(u)\,\mathrm{d}u \cdot \mathbb{1}_{\{T_1>t\}} \,\middle|\, \mathcal{G}_0\right) \\
&\qquad + \left(\frac{v_1}{v_2\theta_1} + \ln(sI)\right) \cdot \mathbb{E}\left(\int_0^t {}^{\times\mathrm{D}}\lambda_\theta(u)\,\mathrm{d}u \cdot \mathbb{1}_{\{T_1\leq t\}} \,\middle|\, \mathcal{G}_0\right) \\
&\quad \leq \left(\frac{v_1}{v_2\theta_1} + \ln(s)\right) \cdot \frac{\theta_1 s^{\theta_2+\theta_3}}{(\theta_3+1)\,\tau^{\theta_3}} \min\{t, \tau_0\}^{\theta_3+1} \cdot \mathbb{1}_{\{0<C_0\}} \cdot \mathbb{P}\left(T_1 > t \,|\, \mathcal{G}_0\right) \\
&\qquad + \left(\frac{v_1}{v_2\theta_1} + \ln(sI)\right) \cdot \mathbb{E}\left(\int_0^t {}^{\times\mathrm{D}}\lambda_\theta(u)\,\mathrm{d}u \cdot \mathbb{1}_{\{T_1\leq t\}} \,\middle|\, \mathcal{G}_0\right) \\
&\quad \leq \left(\frac{v_1}{v_2\theta_1} + \ln(s)\right) \cdot \frac{\theta_1 s^{\theta_2+\theta_3}}{(\theta_3+1)\,\tau^{\theta_3}} \min\{t, \tau_0\}^{\theta_3+1} \cdot \mathbb{1}_{\{0<C_0\}} \cdot \mathbb{P}\left(T_1 > t \,|\, \mathcal{G}_0\right) \\
&\qquad + \left(\frac{v_1}{v_2\theta_1} + \ln(sI)\right) \cdot \frac{\theta_1 (sI)^{\theta_2+\theta_3}}{(\theta_3+1)\,\tau^{\theta_3}} \min\{t, \tau_0\}^{\theta_3+1} \cdot \mathbb{P}\left(T_1 \leq t \,|\, \mathcal{G}_0\right), \qquad (3.69)
\end{aligned}$$

where in the last step we estimated the indicator function $\mathbb{1}_{\{N_{u-}<C_0\}}$ upward by 1. On $\{C_0 = 0\}$, we have ${}^{\times\mathrm{D}}\lambda_\theta \equiv 0$, which implies that this event does not contribute to the expectation. For this reason, we operate only on $\{C_0 > 0\}$ and obtain for $0 < t \leq \min\operatorname{supp}(\mathbb{P}^{\tau_0})$ by applying Equation (3.65) to Equation (3.69):

$$\begin{aligned}
&\mathbb{E}\left(\int_0^t \left(\frac{v_1}{v_2\theta_1} + \ln B(u)\right) \cdot {}^{\times\mathrm{D}}\lambda_\theta(u)\,\mathrm{d}u \,\middle|\, \mathcal{G}_0\right) \\
&\quad \leq \left(\frac{v_1}{v_2\theta_1} + \ln(s)\right) \cdot \frac{\theta_1 s^{\theta_2+\theta_3}}{(\theta_3+1)\,\tau^{\theta_3}} t^{\theta_3+1} \cdot \exp\left(-\frac{\theta_1 s^{\theta_2+\theta_3}}{(\theta_3+1)\,\tau^{\theta_3}} t^{\theta_3+1}\right) \\
&\qquad + \left(\frac{v_1}{v_2\theta_1} + \ln(sI)\right) \cdot \frac{\theta_1 (sI)^{\theta_2+\theta_3}}{(\theta_3+1)\,\tau^{\theta_3}} t^{\theta_3+1} \cdot \left(1 - \exp\left(-\frac{\theta_1 s^{\theta_2+\theta_3}}{(\theta_3+1)\,\tau^{\theta_3}} t^{\theta_3+1}\right)\right).
\end{aligned}$$

This upper bound is deterministic. Moreover, we observe:

$$\begin{aligned}
&\left(\frac{v_1}{v_2\theta_1} + \ln(s)\right) \cdot \frac{\theta_1 s^{\theta_2+\theta_3}}{(\theta_3+1)\,\tau^{\theta_3}} t^{\theta_3+1} \cdot \exp\left(-\frac{\theta_1 s^{\theta_2+\theta_3}}{(\theta_3+1)\,\tau^{\theta_3}} t^{\theta_3+1}\right) \\
&\quad + \left(\frac{v_1}{v_2\theta_1} + \ln(sI)\right) \cdot \frac{\theta_1 (sI)^{\theta_2+\theta_3}}{(\theta_3+1)\,\tau^{\theta_3}} t^{\theta_3+1} \cdot \left(1 - \exp\left(-\frac{\theta_1 s^{\theta_2+\theta_3}}{(\theta_3+1)\,\tau^{\theta_3}} t^{\theta_3+1}\right)\right) < 0
\end{aligned}$$

$$\iff \quad \left(\frac{v_1}{v_2\theta_1} + \ln(sI)\right) \cdot I^{\theta_2+\theta_3} \cdot \left(1 - \exp\left(-\frac{\theta_1 s^{\theta_2+\theta_3}}{(\theta_3+1)\,\tau^{\theta_3}} t^{\theta_3+1}\right)\right)$$
$$< -\left(\frac{v_1}{v_2\theta_1} + \ln(s)\right) \cdot \exp\left(-\frac{\theta_1 s^{\theta_2+\theta_3}}{(\theta_3+1)\,\tau^{\theta_3}} t^{\theta_3+1}\right) . \tag{3.70}$$

Now, for $t \to 0$, the left-hand side of Equation (3.70) converges to 0, while the right-hand side converges to $-\left(\frac{v_1}{v_2\theta_1} + \ln(s)\right)$. Since $s < \exp\left(-\frac{v_1}{v_2\theta_1}\right)$,

$$-\left(\frac{v_1}{v_2\theta_1} + \ln(s)\right) > -\left(\frac{v_1}{v_2\theta_1} + \ln\left(\exp\left(-\frac{v_1}{v_2\theta_1}\right)\right)\right) = 0 ,$$

so there exists $t \in \mathcal{I}$ for which holds:

$$\mathbb{E}\left(\int_0^t \left(\frac{v_1}{v_2\theta_1} + \ln B(u)\right) \cdot {}^{\times\mathrm{D}}\lambda_\theta(u)\,\mathrm{d}u \,\Big|\, \mathcal{G}_0\right) \begin{cases} = 0 , & \text{on } \{C_0 = 0\}, \\ < 0 , & \text{on } \{C_0 > 0\}. \end{cases}$$

Finally, the law of total expectation yields:

$$\begin{aligned}
&\mathbb{E}\left(\int_0^t \left(\frac{v_1}{v_2\theta_1} + \ln B(u)\right) \cdot {}^{\times\mathrm{D}}\lambda_\theta(u)\,\mathrm{d}u\right) \\
&\quad = \mathbb{E}\left[\mathbb{E}\left(\int_0^t \left(\frac{v_1}{v_2\theta_1} + \ln B(u)\right) \cdot {}^{\times\mathrm{D}}\lambda_\theta(u)\,\mathrm{d}u \,\Big|\, \mathcal{G}_0\right)\right] \\
&\quad = \underbrace{\mathbb{E}\left[\mathbb{E}\left(\int_0^t \left(\frac{v_1}{v_2\theta_1} + \ln B(u)\right) \cdot {}^{\times\mathrm{D}}\lambda_\theta(u)\,\mathrm{d}u \,\Big|\, \mathcal{G}_0\right) \,\Big|\, C_0 = 0\right]}_{=0} \cdot \mathbb{P}\left(C_0 = 0\right) \\
&\qquad + \underbrace{\mathbb{E}\left[\mathbb{E}\left(\int_0^t \left(\frac{v_1}{v_2\theta_1} + \ln B(u)\right) \cdot {}^{\times\mathrm{D}}\lambda_\theta(u)\,\mathrm{d}u \,\Big|\, \mathcal{G}_0\right) \,\Big|\, C_0 > 0\right]}_{<0} \cdot \mathbb{P}\left(C_0 > 0\right) < 0 ,
\end{aligned}$$

as $\mathbb{P}\left(C_0 > 0\right) > 0$ by assumption.

Combining the cases (i) through (iii), the statement of Equation (3.67) is verified, completing the proof of the positive definiteness of the standardizing matrix $\Phi_0(\theta)$. □

Remark 3.30 (Positive Definiteness for Random Initial Stress Levels)**.**
As outlined in the proof of Theorem 3.29, the randomness of the initial stress level s significantly complicates the verification of Equation (3.66). This can already be seen in the example of a discrete uniform distribution with support $\operatorname{supp}\left(\mathbb{P}^{s_0}\right) = \{s^{(1)}, s^{(2)}\}$, where

$$s^{(1)} < \frac{1}{I}\exp\left(-\frac{v_1}{v_2\theta_1}\right) < \exp\left(-\frac{v_1}{v_2\theta_1}\right) < s^{(2)} .$$

The above proof (consider the cases (i) and (iv)) then demonstrates that

$$\mathbb{E}\left(\int_0^t \left(\frac{v_1}{v_2\theta_1} + \ln B(u)\right) \cdot {}^{\times\mathrm{D}}\lambda_\theta(u)\,\mathrm{d}u \,\Big|\, s_0 = s^{(1)}\right) < 0$$
$$< \mathbb{E}\left(\int_0^t \left(\frac{v_1}{v_2\theta_1} + \ln B(u)\right) \cdot {}^{\times\mathrm{D}}\lambda_\theta(u)\,\mathrm{d}u \,\Big|\, s_0 = s^{(2)}\right) , \tag{3.71}$$

from where no conclusions can be drawn about the sign of

$$
\begin{aligned}
&\mathbb{E}\left(\int_0^t \left(\frac{v_1}{v_2\theta_1} + \ln B(u)\right) \cdot {}^{\times\mathrm{D}}\lambda_\theta(u)\,\mathrm{d}u\right) \\
&\quad = \mathbb{E}\left(\int_0^t \left(\frac{v_1}{v_2\theta_1} + \ln B(u)\right) \cdot {}^{\times\mathrm{D}}\lambda_\theta(u)\,\mathrm{d}u \,\middle|\, s_0 = s^{(1)}\right) \cdot \underbrace{\mathbb{P}\left(s_0 = s^{(1)}\right)}_{=\frac{1}{2}} \\
&\qquad + \mathbb{E}\left(\int_0^t \left(\frac{v_1}{v_2\theta_1} + \ln B(u)\right) \cdot {}^{\times\mathrm{D}}\lambda_\theta(u)\,\mathrm{d}u \,\middle|\, s_0 = s^{(2)}\right) \cdot \underbrace{\mathbb{P}\left(s_0 = s^{(2)}\right)}_{=\frac{1}{2}} .
\end{aligned}
$$

If, for instance, both conditional expectations in Equation (3.71) have the same absolute value regardless of $t \in \mathcal{I}$, this expectation is always equal to 0. Thus, unlike in the deterministic case, it is no longer sufficient to state that the expectation is strictly positive or negative for any given s; instead, here we need to quantify these conditional expectations in order to relate them to the probabilities specified by $\mathbb{P}^{s_0}$. In this, we have not even taken into account that $s^{(1)}$ and $s^{(2)}$ could also occur with different frequencies. Since this would go beyond the scope of this thesis, we therefore restrict ourselves to the case of deterministic initial stress levels. We do, however, conjecture that the positive definiteness persists even for random initial stress levels, although the proof presented here fails to substantiate this belief. A proof of this assertion would presumably exploit the different stress levels that can occur during an experiment (i.e., take advantage of the increasing step function $B(u)$), whereas we have focused primarily on the stress that is prevalent at the beginning of an experiment (i.e., $B(0) = s$).

3.4.2. Verifying the Assumptions 3.5 for Strong Consistency

To conclude Section 3.4, we finally turn to the Assumptions 3.5 only touched upon so far. We have already mentioned that condition (C3) follows immediately from the ever-present Corollary 2.17 (set $p = q = r = 0$ in Equation (2.40)). Of the remaining two conditions, we begin with the (simpler) condition (C2).

Lemma 3.31 (Continuous Extensions of the Cumulative Intensities in the Basquin Load Sharing Model With Multiplicative Damage Accumulation)**.**
In the situation of Corollary 2.17, let $0 < \varepsilon \leq 1$ and consider an open and bounded parameter space $\Theta \subset \mathbb{R}_+^2 \times [-1+\varepsilon, \infty)$. Then with probability 1, the cumulative intensity of the Basquin load sharing model with multiplicative damage accumulation,

$$
\begin{aligned}
{}^{\times D}\Lambda_{\cdot}(\cdot) \,:\, \mathcal{I} \times \Theta &\longrightarrow \mathbb{R} \\
(t, \theta) &\longmapsto {}^{\times D}\Lambda_\theta(t) \,,
\end{aligned}
$$

is continuous on $\mathcal{I} \times \Theta$ and has a continuous extension to $\mathcal{I} \times \overline{\Theta}$.

Proof. It suffices to show that the mapping ${}^{\times\mathrm{D}}\Lambda_{\cdot}(\cdot)$ is uniformly continuous on $(0, \tau) \times \Theta$ (the interior of $\mathcal{I} \times \Theta$), because any uniformly continuous function admits a continuous extension to the closure of its domain of definition. Furthermore, the equivalence of norms on the finite-dimensional vector space $\mathbb{R}^{1+d} \supset \mathcal{I} \times \overline{\Theta}$ allows us to consider below the norm

$$
\|(t, \theta)\| := \max\left\{\|t\|_{\mathcal{I}}\,, \|\theta\|_{\overline{\Theta}}\right\} \,,
$$

where $\|\cdot\|_{\mathcal{I}}$ and $\|\cdot\|_{\overline{\Theta}}$ are arbitrary norms on $\mathcal{I}$ and $\overline{\Theta}$, respectively. For simplicity, we omit the indices and settle for the Euclidean norm in both cases. The uniform continuity can be derived from the following two observations:

(i) According to Equation (2.40) from Corollary 2.17, the partial derivatives of ${}^{\times\mathrm{D}}\Lambda_{\theta}(t)$ with respect to θ are bounded by a constant independent of $(t, \theta) \in \mathcal{I} \times \Theta$. Consequently, such a constant $C_1 > 0$ can also be found for the gradient of ${}^{\times\mathrm{D}}\Lambda_{\theta}(t)$, so we obtain:

$$\sup_{(t,\theta)\in\mathcal{I}\times\Theta} \left\| D_\theta\, {}^{\times\mathrm{D}}\Lambda_\theta(t) \right\| \leq C_1 \,. \tag{3.72}$$

(ii) Choosing $p = q = r = 0$ in Equation (2.39) of Corollary 2.17 yields

$$\left| {}^{\times\mathrm{D}}\lambda_\theta(t) \right| \leq C \left(s_0 \frac{t}{\tau} \right)^{\varepsilon - 1} .$$

Hence, for $t_1, t_2 \in \mathcal{I}$ with $t_1 \leq t_2$ we have:

$$\begin{aligned} \left| \int_{t_1}^{t_2} {}^{\times\mathrm{D}}\lambda_\theta(t)\,\mathrm{d}t \right| &\leq \int_{t_1}^{t_2} C \left(s_0 \frac{t}{\tau} \right)^{\varepsilon-1} \mathrm{d}t \\ &= C \left(\frac{s_0}{\tau} \right)^{\varepsilon-1} \left[\frac{t^\varepsilon}{\varepsilon} \right]_{t=t_1}^{t_2} \\ &= C \left(\frac{s_0}{\tau} \right)^{\varepsilon-1} \frac{t_2^\varepsilon - t_1^\varepsilon}{\varepsilon} \,. \end{aligned} \tag{3.73}$$

Moreover, for $0 < \varepsilon \leq 1$ holds:

$$\begin{aligned} 1 - \frac{t_1^\varepsilon}{t_2^\varepsilon} = 1 - \Bigg(\underbrace{\frac{t_1}{t_2}}_{\leq 1} \Bigg)^\varepsilon &\leq \underbrace{1 - \frac{t_1}{t_2}}_{\leq 1} \leq \left(1 - \frac{t_1}{t_2} \right)^\varepsilon \\ \iff \qquad t_2^\varepsilon - t_1^\varepsilon \leq t_2^\varepsilon \left(1 - \frac{t_1}{t_2} \right)^\varepsilon &= (t_2 - t_1)^\varepsilon \,. \end{aligned} \tag{3.74}$$

Combining Equations (3.73) and (3.74), we find that

$$\left| \int_{t_1}^{t_2} {}^{\times\mathrm{D}}\lambda_\theta(t)\,\mathrm{d}t \right| \leq C_2\,(t_2 - t_1)^\varepsilon \tag{3.75}$$

for an appropriate constant $C_2 > 0$ that is independent of t_1, t_2 and θ.

Now let $\eta > 0$. In order to prove uniform continuity of ${}^{\times\mathrm{D}}\Lambda_{\cdot}(\cdot)$, we need to find $\delta > 0$ such that for all $(t_1, \theta_1), (t_2, \theta_2) \in (0, \tau) \times \Theta$ with $\|(t_2, \theta_2) - (t_1, \theta_1)\| < \delta$ holds that $\left\| {}^{\times\mathrm{D}}\Lambda_{\theta_1}(t_1) - {}^{\times\mathrm{D}}\Lambda_{\theta_2}(t_2) \right\| < \eta$. By the triangle inequality, the mean value theorem and Equations (3.72) and (3.75) we have:

$$\begin{aligned} \left\| {}^{\times\mathrm{D}}\Lambda_{\theta_1}(t_1) - {}^{\times\mathrm{D}}\Lambda_{\theta_2}(t_2) \right\| &= \left\| {}^{\times\mathrm{D}}\Lambda_{\theta_1}(t_1) - {}^{\times\mathrm{D}}\Lambda_{\theta_2}(t_1) + {}^{\times\mathrm{D}}\Lambda_{\theta_2}(t_1) - {}^{\times\mathrm{D}}\Lambda_{\theta_2}(t_2) \right\| \\ &\leq \sup_{\theta\in\Theta} \left\| D_\theta\, {}^{\times\mathrm{D}}\Lambda_\theta(t_1) \right\| \|\theta_2 - \theta_1\| + \left\| \int_{t_1}^{t_2} {}^{\times\mathrm{D}}\lambda_{\theta_2}(u)\,\mathrm{d}u \right\| \\ &\leq C_1 \|\theta_2 - \theta_1\| + C_2 \|t_2 - t_1\|^\varepsilon \,. \end{aligned} \tag{3.76}$$

Choose any $\delta \leq \max\left\{\left(\frac{\eta}{C_1+C_2}\right)^{\frac{1}{\varepsilon}}, 1\right\}$. By choice of norm, $\|(t_2, \theta_2) - (t_1, \theta_1)\| < \delta$ trivially implies that both $\|t_2 - t_1\| < \delta$ and $\|\theta_2 - \theta_1\| < \delta$. Continuing from Equation (3.76), we then obtain:

$$C_1 \|\theta_2 - \theta_1\| + C_2 \|t_2 - t_1\|^{\varepsilon} < C_1\delta + C_2\delta^{\varepsilon} \leq (C_1 + C_2)\,\delta^{\varepsilon} \leq (C_1 + C_2)\,\frac{\eta}{C_1 + C_2} = \eta\,,$$

and hence $\left\|{}^{\times\mathrm{D}}\Lambda_{\theta_1}(t_1) - {}^{\times\mathrm{D}}\Lambda_{\theta_2}(t_2)\right\| < \eta$, which concludes the proof. □

This leaves only Condition (C1), which can be shown using the methods introduced in Theorem 3.29. Consequently, we will rely on the same assumptions, with the only actual limitation again being the necessity of a deterministic initial stress level.

Theorem 3.32 (Identifiability in the Basquin Load Sharing Model With Multiplicative Damage Accumulation)**.**
Let the initial stress level be deterministic, so that $\mathbb{P}^{s_0} = \delta_s$ *for some* $s > 0$*. Suppose that* $\min\operatorname{supp}(\mathbb{P}^{\tau_0}) > 0$ *and that* $\operatorname{supp}(\mathbb{P}^{C_0}) \cap \{2, \dots, I\} \neq \emptyset$*. Then, under the preconditions of Corollary 2.17, the true parameter* θ^* *is identifiable in the Basquin load sharing model with multiplicative damage accumulation.*

Proof. Let $\theta \neq \theta^*$. By Equation (3.9), it suffices to find $t \in \mathcal{I}$ such that

$$\mathbb{E}\left({}^{\times\mathrm{D}}\Lambda_{\theta^*}(t) - {}^{\times\mathrm{D}}\Lambda_{\theta}(t)\right) \neq 0\,. \tag{3.77}$$

Because the expected cumulative intensity is continuous by Remark 3.6, there then exists an interval $\mathcal{I}_\theta \subset \mathcal{I}$ so that Equation (3.77) is fulfilled for all $t \in \mathcal{I}_\theta$. Moreover, Theorem 3.28 implies that $\mu(\mathcal{I}_\theta) > 0$, where μ is the Borel measure on $\mathcal{I}$ induced by ${}^{\times\mathrm{D}}\Lambda_{\theta^*}$, see Equation (3.2).
Similar to the proof of Theorem 3.29, we distinguish whether $\theta_3^* = \theta_3$ or $\theta_3^* \neq \theta_3$. We start with the case $\theta_3^* \neq \theta_3$ and assume without loss of generality that $\theta_3^* > \theta_3$ (otherwise we swap θ^* and θ). By Fubini's theorem,

$$\mathbb{E}\left({}^{\times\mathrm{D}}\Lambda_{\theta^*}(t) - {}^{\times\mathrm{D}}\Lambda_{\theta}(t)\right) = \int_0^t \mathbb{E}\left({}^{\times\mathrm{D}}\lambda_{\theta^*}(u) - {}^{\times\mathrm{D}}\lambda_{\theta}(u)\right)\mathrm{d}u\,. \tag{3.78}$$

We proceed by showing that the integrand is strictly negative in a (deterministic) neighbourhood of 0, which carries over to the expectation and the outer integral. Considering that $\theta_3 < 0$ is allowed, on $\{N_{u^-} < C_0\} \cap \{u \leq \tau_0\}$ we estimate:

$$L(\theta)\,u^{\theta_3} \leq {}^{\times\mathrm{D}}\lambda_{\theta}(u) = \theta_1\left(s\frac{I}{I - N_{u^-}}\right)^{\theta_2}\left(\frac{1}{\tau}\int_0^u s\frac{I}{I - N_{x^-}}\,\mathrm{d}x\right)^{\theta_3} \leq U(\theta)\,u^{\theta_3}\,,$$

$$\text{where } L(\theta) := \begin{cases} \theta_1 s^{\theta_2}\left(\frac{s}{\tau}\right)^{\theta_3}, & \text{if } \theta_3 \geq 0\,,\\ \theta_1 s^{\theta_2}\left(\frac{sI}{\tau}\right)^{\theta_3}, & \text{if } \theta_3 < 0\,,\end{cases}$$

$$\text{and} \quad U(\theta) := \begin{cases} \theta_1 (sI)^{\theta_2}\left(\frac{sI}{\tau}\right)^{\theta_3}, & \text{if } \theta_3 \geq 0\,,\\ \theta_1 (sI)^{\theta_2}\left(\frac{s}{\tau}\right)^{\theta_3}, & \text{if } \theta_3 < 0\,.\end{cases}$$

Accordingly, on $\{N_{u^-} < C_0\} \cap \{u \leq \tau_0\}$ we obtain that

$${}^{\times\mathrm{D}}\lambda_{\theta^*}(u) - {}^{\times\mathrm{D}}\lambda_{\theta}(u) \leq U(\theta^*)\,u^{\theta_3^*} - L(\theta)\,u^{\theta_3}\,.$$

However, $U(\theta^*)$ and $L(\theta)$ are deterministic positive constants. As $\theta_3^* - \theta_3 > 0$, we have for $u > 0$:

$$U(\theta^*)\, u^{\theta_3^*} - L(\theta)\, u^{\theta_3} < 0 \quad \iff \quad u^{\theta_3^*-\theta_3} < \frac{L(\theta)}{U(\theta^*)} \quad \iff \quad u < \underbrace{\left(\frac{L(\theta)}{U(\theta^*)}\right)^{\frac{1}{\theta_3^*-\theta_3}}}_{>0},$$

and therefore

$$^{\times\mathrm{D}}\lambda_{\theta^*}(u) - {}^{\times\mathrm{D}}\lambda_{\theta}(u) < 0 \quad \text{for all } 0 < u < \left(\frac{L(\theta)}{U(\theta^*)}\right)^{\frac{1}{\theta_3^*-\theta_3}}. \tag{3.79}$$

Since the conditional intensity vanishes on $\left(\{N_{u^-} < C_0\} \cap \{u \leq \tau_0\}\right)^{\complement}$ regardless of its parameter, Equation (3.79) also transfers to the expectation,

$$\mathbb{E}\left({}^{\times\mathrm{D}}\lambda_{\theta^*}(u) - {}^{\times\mathrm{D}}\lambda_{\theta}(u)\right) = \mathbb{E}\left[\left({}^{\times\mathrm{D}}\lambda_{\theta^*}(u) - {}^{\times\mathrm{D}}\lambda_{\theta}(u)\right) \cdot \mathbb{1}_{\{N_{u^-}<C_0\}\cap\{u\leq\tau_0\}}\right] < 0\,, \tag{3.80}$$

at least if $\mathbb{P}\left(\{N_{u^-} < C_0\} \cap \{u \leq \tau_0\}\right) > 0$. This is trivially the case for any $u \leq \min \operatorname{supp}(\mathbb{P}^{\tau_0})$ (as the proof of Theorem 3.28 shows, it suffices for this property if $u < t$ for some $t \in \operatorname{supp}(\mathbb{P}^{\tau_0})$). Now let $t \in \mathcal{I}$ with

$$0 < t < \min\left\{\min \operatorname{supp}(\mathbb{P}^{\tau_0}), \left(\frac{L(\theta)}{U(\theta^*)}\right)^{\frac{1}{\theta_3^*-\theta_3}}\right\}.$$

Then, substituting Equation (3.80) into Equation (3.78) yields

$$\mathbb{E}\left({}^{\times\mathrm{D}}\Lambda_{\theta^*}(t) - {}^{\times\mathrm{D}}\Lambda_{\theta}(t)\right) < 0\,,$$

so that Equation (3.77) is satisfied in the case $\theta_3^* \neq \theta_3$.
We turn to the case $\theta_3^* = \theta_3$. If we assume that we operate on $\{N_{u^-} < C_0\} \cap \{u \leq \tau_0\}$ as before, then

$$^{\times\mathrm{D}}\lambda_{\theta^*}(u) - {}^{\times\mathrm{D}}\lambda_{\theta}(u) = \left(\theta_1^*\left(s\frac{I}{I-N_{u^-}}\right)^{\theta_2^*} - \theta_1\left(s\frac{I}{I-N_{u^-}}\right)^{\theta_2}\right) \cdot A(u)^{\theta_3}\,. \tag{3.81}$$

We will use this identity to compute the conditional expectation given $\mathcal{G}_0$ in Equation (3.78), for which the procedure used to derive Equation (3.69) can be adopted. Let us first suppose that $\theta_1^* s^{\theta_2^*} - \theta_1 s^{\theta_2} < 0$. Then,

$$\begin{aligned}
&\mathbb{E}\left({}^{\times\mathrm{D}}\Lambda_{\theta^*}(t) - {}^{\times\mathrm{D}}\Lambda_{\theta}(t) \,\middle|\, \mathcal{G}_0\right) \\
&= \mathbb{E}\left(\int_0^t \left(\theta_1^*\left(s\frac{I}{I-N_{u^-}}\right)^{\theta_2^*} - \theta_1\left(s\frac{I}{I-N_{u^-}}\right)^{\theta_2}\right) \cdot A(u)^{\theta_3} \cdot \mathbb{1}_{\{T_1>t\}}\,\mathrm{d}u \,\middle|\, \mathcal{G}_0\right) \\
&\quad + \mathbb{E}\left(\int_0^t \left(\theta_1^*\left(s\frac{I}{I-N_{u^-}}\right)^{\theta_2^*} - \theta_1\left(s\frac{I}{I-N_{u^-}}\right)^{\theta_2}\right) \cdot A(u)^{\theta_3} \cdot \mathbb{1}_{\{T_1\leq t\}}\,\mathrm{d}u \,\middle|\, \mathcal{G}_0\right) \\
&\leq \left(\theta_1^* s^{\theta_2^*} - \theta_1 s^{\theta_2}\right) \frac{s^{\theta_3}}{(\theta_3+1)\tau^{\theta_3}} \min\{t, \tau_0\}^{\theta_3+1} \cdot \mathbb{1}_{\{0<C_0\}} \cdot \mathbb{P}\left(T_1 > t \,\middle|\, \mathcal{G}_0\right)
\end{aligned}$$

$$+\left(\theta_1^*\,(sI)^{\theta_2^*}-\theta_1 s^{\theta_2}\right)\frac{(sI)^{\theta_3}}{(\theta_3+1)\tau^{\theta_3}}\min\left\{t,\tau_0\right\}^{\theta_3+1}\cdot\mathbb{P}\left(T_1\le t\,\middle|\,\mathcal{G}_0\right)\,. \tag{3.82}$$

On $\{C_0>0\}$ and for $0<t\le\min\operatorname{supp}\left(\mathbb{P}^{\tau_0}\right)$, substituting Equation (3.65) into Equation (3.82) yields:

$$\begin{aligned}&\mathbb{E}\left({}^{\times\mathrm{D}}\Lambda_{\theta^*}(t)-{}^{\times\mathrm{D}}\Lambda_\theta(t)\,\middle|\,\mathcal{G}_0\right)\\&\quad\le\left(\theta_1^* s^{\theta_2^*}-\theta_1 s^{\theta_2}\right)\frac{s^{\theta_3}}{(\theta_3+1)\tau^{\theta_3}}t^{\theta_3+1}\cdot\exp\left(-\frac{\theta_1^* s^{\theta_2^*+\theta_3^*}}{(\theta_3^*+1)\,\tau^{\theta_3^*}}t^{\theta_3^*+1}\right)\\&\qquad+\left(\theta_1^*\,(sI)^{\theta_2^*}-\theta_1 s^{\theta_2}\right)\frac{(sI)^{\theta_3}}{(\theta_3+1)\tau^{\theta_3}}t^{\theta_3+1}\cdot\left(1-\exp\left(-\frac{\theta_1^* s^{\theta_2^*+\theta_3^*}}{(\theta_3^*+1)\,\tau^{\theta_3^*}}t^{\theta_3^*+1}\right)\right)\,.\end{aligned}$$

With the same arguments as in the proof of Theorem 3.29 (see case (ii)), we observe:

$$\begin{aligned}&\left(\theta_1^* s^{\theta_2^*}-\theta_1 s^{\theta_2}\right)\frac{s^{\theta_3}}{(\theta_3+1)\tau^{\theta_3}}t^{\theta_3+1}\cdot\exp\left(-\frac{\theta_1^* s^{\theta_2^*+\theta_3^*}}{(\theta_3^*+1)\,\tau^{\theta_3^*}}t^{\theta_3^*+1}\right)\\&\quad+\left(\theta_1^*\,(sI)^{\theta_2^*}-\theta_1 s^{\theta_2}\right)\frac{(sI)^{\theta_3}}{(\theta_3+1)\tau^{\theta_3}}t^{\theta_3+1}\cdot\left(1-\exp\left(-\frac{\theta_1^* s^{\theta_2^*+\theta_3^*}}{(\theta_3^*+1)\,\tau^{\theta_3^*}}t^{\theta_3^*+1}\right)\right)<0\\\iff\;&\underbrace{\left(\theta_1^*\,(sI)^{\theta_2^*}-\theta_1 s^{\theta_2}\right)\cdot I^{\theta_3}\cdot\left(1-\exp\left(-\frac{\theta_1^* s^{\theta_2^*+\theta_3^*}}{(\theta_3^*+1)\,\tau^{\theta_3^*}}t^{\theta_3^*+1}\right)\right)}_{\longrightarrow 0\quad(t\longrightarrow 0)}\\&\quad<\underbrace{-\left(\theta_1^* s^{\theta_2^*}-\theta_1 s^{\theta_2}\right)}_{>0}\underbrace{\exp\left(-\frac{\theta_1^* s^{\theta_2^*+\theta_3^*}}{(\theta_3^*+1)\,\tau^{\theta_3^*}}t^{\theta_3^*+1}\right)}_{\longrightarrow 1\quad(t\longrightarrow 0)}\,,\end{aligned}$$

which is always satisfied for sufficiently small t. For such t, we consequently have

$$\mathbb{E}\left({}^{\times\mathrm{D}}\Lambda_{\theta^*}(t)-{}^{\times\mathrm{D}}\Lambda_\theta(t)\,\middle|\,\mathcal{G}_0\right)\begin{cases}=0\,, & \text{on }\;\{C_0=0\}\,,\\<0\,, & \text{on }\;\{C_0>0\}\,.\end{cases}$$

From here, the law of total expectation can be used to infer the negativity of the expectation from Equation (3.77) (see again case (ii) from the proof of Theorem 3.29 for details).

If instead we suppose that $\theta_1^* s^{\theta_2^*}-\theta_1 s^{\theta_2}>0$, we can repeat the above computations for

$$-\mathbb{E}\left({}^{\times\mathrm{D}}\Lambda_{\theta^*}(t)-{}^{\times\mathrm{D}}\Lambda_\theta(t)\,\middle|\,\mathcal{G}_0\right)=\mathbb{E}\left({}^{\times\mathrm{D}}\Lambda_\theta(t)-{}^{\times\mathrm{D}}\Lambda_{\theta^*}(t)\,\middle|\,\mathcal{G}_0\right)\,,$$

which then implies

$$\mathbb{E}\left({}^{\times\mathrm{D}}\Lambda_{\theta^*}(t)-{}^{\times\mathrm{D}}\Lambda_\theta(t)\,\middle|\,\mathcal{G}_0\right)\begin{cases}=0\,, & \text{on }\;\{C_0=0\}\,,\\>0\,, & \text{on }\;\{C_0>0\}\,,\end{cases}$$

and hence the positivity of the expectation from Equation (3.77).

There remains the situation where $\theta_1^* s^{\theta_2^*}=\theta_1 s^{\theta_2}$ and thus $\theta_1^* s^{\theta_2^*}-\theta_1 s^{\theta_2}=0$. Note that this automatically excludes the case $\theta_2=\theta_2^*$, since otherwise $\theta_1=\theta_1^*$ would follow, which contradicts the premise $\theta \neq \theta^*$. The ideas below are similar to case (iii) in the proof of Theorem 3.29: We define the auxiliary function

$$g_{\theta,\theta^*}(x) := \theta_1^* x^{\theta_2^*} - \theta_1 x^{\theta_2} \,.$$

The only root of g_{θ,θ^*} can be determined analytically as $x_0 = \left(\frac{\theta_1^*}{\theta_1}\right)^{\frac{1}{\theta_2 - \theta_2^*}}$. In our situation, this means that $B(0) = s = x_0$ must hold, where as usual $B(t) = s\frac{I}{I - N_{t-}}$ denotes the load sharing term. Since g_{θ,θ^*} is continuous, we have either $g_{\theta,\theta^*}(x) > 0$ or $g_{\theta,\theta^*}(x) < 0$ for all $x > x_0$. By construction, $B(t)$ is non-decreasing and $B(t) > B(0)$ for $t > T_1$ (as here $N_{t-} \geq 1$), which implies that either $g_{\theta,\theta^*}(B(t)) > 0$ or $g_{\theta,\theta^*}(B(t)) < 0$ applies for all $t > T_1$. Now, Equation (3.81) can be rewritten as

$${}^{\times\mathrm{D}}\lambda_{\theta^*}(u) - {}^{\times\mathrm{D}}\lambda_{\theta}(u) = g_{\theta,\theta^*}(B(u)) \cdot A(u)^{\theta_3} \,,$$

and hence

$${}^{\times\mathrm{D}}\Lambda_{\theta^*}(t) - {}^{\times\mathrm{D}}\Lambda_{\theta}(t) = \int_0^t {}^{\times\mathrm{D}}\lambda_{\theta^*}(u) - {}^{\times\mathrm{D}}\lambda_{\theta}(u)\,\mathrm{d}u$$

is either strictly positive or strictly negative on the set $\{t > T_1\} \cap \{t < \tau_0\} \cap \{1 < C_0\}$, while on the complement it either vanishes or takes the same sign because the sign of $g_{\theta,\theta^*}(B(u))$ never changes from minus to plus or vice versa. To conclude the proof, we then only have to prove that $\{t > T_1\} \cap \{t < \tau_0\} \cap \{1 < C_0\}$ has positive $\mathbb{P}$-measure. But this was shown as part of case (ii) in the proof of Theorem 3.29. $\square$

4. Statistical Inference Based on the K-Sign Depth

After the detailed treatment of the minimum distance estimator in the context of "self-exciting" point processes, we now turn to a fundamentally different method of statistical inference in the K-sign depth test. Although its foundations were laid early in the regression depth of Rousseeuw and Hubert 1999 and have since been further developed in Müller 2005, Kustosz, Leucht and Müller 2016 and Kustosz, Wendler and Müller 2016, the K-sign depth is a quite novel notion of depth. Recently, results on

(i) the asymptotic distribution of the K-sign depth and its efficient implementation (Malcherczyk, Leckey and Müller 2021),

(ii) the power and robustness of the K-sign depth test (Leckey et al. 2023),

(iii) applications of the K-sign depth test in multiple regression (Horn and Müller 2023)

have been published. These studies culminated in consistency conditions for the K-sign depth test (Leckey, Jakubzik and Müller 2023). In this chapter, we demonstrate how the K-sign depth test can be used in an intensity-based framework. We prove that the consistency conditions can be satisfied in the case $K = 3$ and derive further applications for our models, in particular the Basquin load sharing model with multiplicative damage accumulation.

4.1. The K-Sign Depth

In this first section, we introduce the K-sign depth in Subsection 4.1.1. We then learn about its asymptotic distribution and use it to derive the asymptotic K-sign depth test in Subsection 4.1.2.

4.1.1. Definition of the K-Sign Depth

The conceptual origin of the K-sign depth lies in the regression depth. Rousseeuw and Hubert coined the notion of regression depth when they applied the halfspace (location) depth of Tukey 1975 to a regression setting (see Rousseeuw and Hubert 1999). In essence, the regression depth is intended to gauge how well a - potentially multidimensional - parameter of a regression model fits the observed data.
The simplicial (location) depth of Liu 1990 extended the halfspace depth of Tukey 1975. Translating the relationship between halfspace depth and simplicial depth into the regression context yielded the simplicial regression depth (this was already considered in Rousseeuw and Hubert 1999, see also Kustosz, Wendler and Müller 2016, p. 126).
In an approach to provide a unified notion of depth, Mizera 2002 proposed the tangent depth which includes both the halfspace depth and the regression depth as a special case (cf. Müller 2005, p. 154). Replacing the regression depth with the tangent depth in the extension of Liu 1990 then further generalized the simplicial regression depth to the simplicial tangent depth.
In a classical regression model with d-dimensional model parameter θ, the depth is calculated from the residuals at θ. Under certain conditions given in Kustosz, Wendler and Müller 2016, the simplicial tangent depth amounts to the relative proportion of $(d+1)$-tuples of residuals with alternating signs. This insight motivates the notion of a K-sign depth, where the depth of a parameter θ is defined directly "via alternating

signs of residuals in K-tuples" (Leckey et al. 2023, p. 859). Therefore, the K-sign depth and the simplicial tangent depth coincide for $K = d + 1$, but other choices of K may be considered. This paves the way for a formal definition of the K-sign depth. We provide an overview of how the various concepts of depth are connected in Figure 4.

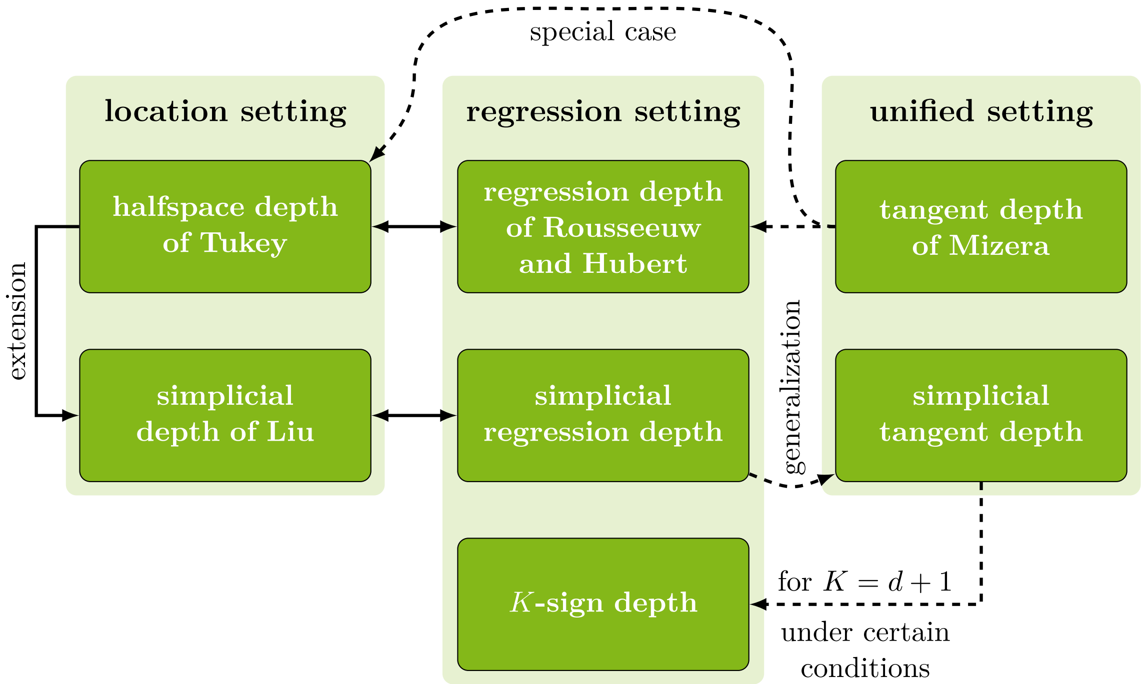


Figure 4: Schematic of the relationships between the different notions of depth.

In order to work in an intensity-based framework, we need to define the K-sign depth in a more general context than usual. For this, let $\Theta \subset \mathbb{R}^d$ with $d \in \mathbb{N}$, $N \in \mathbb{N}$ and consider a parametric family of real-valued random variables,

$$\left\{ R_n^\theta \, : \, n \in \{1, \ldots, N\}, \theta \in \Theta \right\} . \tag{4.1}$$

As usual, $\theta^* \in \Theta$ denotes an unknown "true" parameter. For instance, in a regression setting, we have

$$Y_n = f\left(X_n, \theta^*\right) + E_n \, , \quad n \in \{1, \ldots, N\} \, , \tag{4.2}$$

with a dependent variable Y_n taking values in $\mathbb{R}$, (potentially random) explanatory variables $X_1, \ldots, X_N \in \mathbb{R}^p$, $p \in \mathbb{N}$, a regression function $f : \mathbb{R}^p \times \Theta \to \mathbb{R}$ and unobservable random error terms $E_1, \ldots, E_N$.

In this context, we can define the random variable R_n^θ as the residual of the model at $\theta \in \Theta$, that is:

$$R_n^\theta := Y_n - f\left(X_n, \theta\right) \, , \quad n \in \{1, \ldots, N\} \, , \quad \theta \in \Theta \, .$$

In intensity-based modelling, however, we do not obtain such residuals, so we opt instead for the general approach of Equation (4.1). As such a family of random variables may comprise transformations of point processes, this later enables us to invoke the standardized hazard transforms of Section 2.5 in place of the usual residuals.

After these minor preliminaries, we can now define the K-sign depth.

Definition 4.1 (K-Sign Depth; cf. Leckey et al. 2023, p. 861)**.**
Let $K \in \mathbb{N} \setminus \{1\}$. The K-sign depth of $R^\theta := (R_1^\theta, \ldots, R_N^\theta)$ is defined as[15]

$$d_K(R^\theta) := \frac{1}{\binom{N}{K}} \sum_{1 \leq n_1 < \ldots < n_K \leq N} \left(\prod_{k=1}^{K} \mathbb{1}\left\{(-1)^k R_{n_k}^\theta > 0\right\} + \prod_{k=1}^{K} \mathbb{1}\left\{(-1)^k R_{n_k}^\theta < 0\right\} \right).$$

By definition, the K-sign depth is the relative proportion of ordered K-tuples with alternating signs. Notably, the event $\{\omega : R_n^\theta(\omega) = 0 \text{ for some } n \in \{1, \ldots, N\}\}$ interacts unfavourably with the above definition, since tuples $(n_1, \ldots, n_K)$ with $R_{n_k}^\theta = 0$ for some $k \in \{1, \ldots, K\}$ never contribute to the total depth. Accordingly, we will often demand that this event be a $\mathbb{P}_{\theta^*}$-null set for each $\theta \in \Theta$ (e.g., when the R_n^θ follow a continuous distribution under $\mathbb{P}_{\theta^*}$).
Since the parametric family from Equation (4.1) does not assign any meaning to the parameter θ, a true parameter value θ^* has no inherent relevance. Similarly, the calculated depth at θ also has no meaningful interpretation. It is only true in the modelling context (e.g., in the regression setting) that a larger depth tends to indicate a "good fit" of the parameter θ.
A crucial aspect in depth-related statistical inference is the chosen order (cf. Leckey et al. 2023, p. 861).

Remark 4.2 (On Ordering in the Context of K-Sign Depth)**.**
It is pointed out in Malcherczyk, Leckey and Müller 2021 that the power of the K-sign depth test heavily depends on the chosen order. For example, in a regression setting, ordering with respect to a single univariate explanatory variable X_n (instead of the canonical ordering by the number n of the observation) is often advisable (Malcherczyk, Leckey and Müller 2021, p. 346). In case of a multivariate explanatory variable (i.e., $p > 1$), Horn and Müller 2023 propose a multitude of data-driven orderings. We note that Equation (4.2) is also suitable for autoregressive models, where an AR(1) model can be implemented by choosing $X_n = Y_{n-1}$ (see Kustosz, Leucht and Müller 2016). In this situation, the ordering is based on previous observations, which in the population case leads to a random ordering. Notably, under basic assumptions for the random variables R_n^θ, the asymptotic distribution of the K-sign depth is independent from the chosen ordering (cf. Malcherczyk 2022, p. 8). These assumptions are given in the next subsection.

4.1.2. The K-Sign Depth Test

The K-sign depth test is an asymptotic test proposed by Leckey et al. 2023. Its construction relies on the quantiles of the asymptotic distribution of the K-sign depth. The asymptotic distribution of the K-sign depth in the cases $K = 2$ and $K = 3$ have already been derived by Müller 2005 and Kustosz, Leucht and Müller 2016, respectively. In Malcherczyk, Leckey and Müller 2021, a proof for general K is given. Since the K-sign depth test for $K = 2$ is equivalent to the classical sign test (see Leckey et al. 2023, p. 874), we concentrate on the cases where $K \geq 3$. We restate the main results of Malcherczyk, Leckey and Müller 2021 here, adapting them to our situation where necessary. For this, we start with the assumptions hinted at in Remark 4.2. To maintain consistent notation,

[15]Normally, we use the notation $\mathbb{1}_{\{\ldots\}}(\omega)$ for indicator functions, often omitting the argument ω. Since this makes the events appear subscript and thus smaller, we occasionally prefer the modified notation $\mathbb{1}\{\ldots\}$ throughout this chapter to ensure better readability.

we will continue to denote the true probability measure with $\mathbb{P}_{\theta^*}$ in this subsection. This simplifies the application to an intensity-based framework later on.

Assumption 4.3 (Requirements for the Asymptotic Distribution of the K-Sign Depth)**.**
For $\theta \in \Theta$, we say that $R^\theta = (R_1^\theta, \dots, R_N^\theta)$ satisfies the assumptions (K1) and (K2) if

$$R_1^\theta, \dots, R_N^\theta \qquad \text{are independent (w.r.t. } \mathbb{P}_{\theta^*}), \tag{K1}$$

$$\mathbb{P}_{\theta^*}(R_n^\theta > 0) = \mathbb{P}_{\theta^*}(R_n^\theta < 0) = \frac{1}{2} \qquad \text{for } n = 1, \dots, N. \tag{K2}$$

Whether assumptions (K1) and (K2) are satisfied therefore also depends on the true parameter θ^*. The letter K is used to associate the assumptions with the $\underline{K}$-sign depth. The letter D (for $\underline{\text{d}}$epth) is reserved for future assumptions.

In the regression setting, the assumptions (K1) and (K2) can be translated into requirements for the error terms $E_1, \dots, E_N$. At the true parameter θ^*,

$$R_n^{\theta^*} = E_n\,, \quad n = 1, \dots, N\,.$$

If the error terms are assumed to follow a continuous distribution, then Assumption 4.3 demands that they are independent and (median-)centred. These comparatively minor assumptions are sufficient to derive the asymptotic distribution of the K-sign depth.

Theorem 4.4 (Asymptotic Distribution of the K-Sign Depth; Theorem 2.2. of Malcherczyk, Leckey and Müller 2021, p. 346)**.**
Let $K \geq 3$. If $R_1^\theta, \dots, R_N^\theta$ satisfy the assumptions (K1) *and* (K2)*, then, as $N \to \infty$,*

$$N\left(d_K(R_1^\theta, \dots, R_N^\theta) - \left(\frac{1}{2}\right)^{K-1}\right) \xrightarrow{\text{d}} \Psi_K(W)\,,$$

where $W = (W_t)_{t\in[0,1]}$ denotes a standard Brownian motion and

$$\begin{aligned}
&\Psi_3(W) := \frac{3}{4}\left(1 - \int_0^1 (W_1 - 2W_t)^2 \,\mathrm{d}t\right)\,,\\
&\Psi_K(W)\\
&:= -\frac{K!}{4(K-4)!}\int_{-0.5}^1 \int_{t\vee 0}^{t+0.5} \left(\frac{1}{2} + t - s\right)^{K-4} \left((W_{s\wedge 1} - W_{t\vee 0})^2 - ((s\wedge 1) - (t\vee 0))\right)\mathrm{d}s\,\mathrm{d}t\\
&\quad - \frac{K!}{2(K-4)!}\int_{0.5}^1 \int_0^{t-0.5} \left(\frac{1}{2} + s - t\right)^{K-4} W_s(W_1 - W_t)\,\mathrm{d}s\,\mathrm{d}t\,, \qquad K \geq 4\,.
\end{aligned}$$

Proof. The complete proof is given in Malcherczyk, Leckey and Müller 2021, supplementary explanations can be found in Chapter 3 of Malcherczyk 2022.
We sketch the proof idea for $K = 3$ to motivate the introduction of a "normalized" K-sign depth. Moreover, both the K-sign depth test and its associated consistency conditions from Subsection 4.2.1 build on the ideas embodied in this proof. The central component is the stochastic process $\mathcal{W}^N(\theta) = (\mathcal{W}_t^N(\theta))_{t\in[0,1]}$ defined via

$$\mathcal{W}_t^N(\theta) := \frac{1}{\sqrt{N}}\sum_{n=1}^{\lfloor tN\rfloor} \operatorname{sgn} R_n^\theta\,, \quad t \in [0,1]\,, \tag{4.3}$$

with the convention that $\mathcal{W}_t^N(\theta) := 0$ for $t < \frac{1}{N}$. Here, sgn is the sign function given by

$$\operatorname{sgn} x := \begin{cases} -1\,, & x < 0\,, \\ 0\,, & x = 0\,, \\ 1\,, & x > 0\,. \end{cases}$$

We mostly omit the argument θ and write $\mathcal{W}_t^N$ or $\mathcal{W}^N$ instead of $\mathcal{W}_t^N(\theta)$ or $\mathcal{W}^N(\theta)$, respectively. Under the assumptions (K1) and (K2), $\mathcal{W}^N$ is a random walk rescaled by $\frac{1}{\sqrt{N}}$. Since the paths of $\mathcal{W}^N$ are càdlàg, $\mathcal{W}^N$ takes values in the Skorokhod space $\mathcal{D}([0,1])$ of càdlàg functions on $[0,1]$ (see Definition B.3.3). Donsker's invariance principle (Theorem 16.1 of Billingsley 1968) then yields[16]

$$\mathcal{W}^N \xrightarrow{\mathrm{d}} W \quad (N \to \infty)\,.$$

As $\Psi_3 : \mathcal{D}([0,1]) \to \mathbb{R}$ is continuous in f for all $f \in C([0,1]) \subset \mathcal{D}([0,1])$ and W is almost surely continuous, the continuous mapping theorem grants

$$\Psi_3(\mathcal{W}^N) \xrightarrow{\mathrm{d}} \Psi_3(W) \quad (N \to \infty)\,.$$

The most tedious part of the proof is to show that $\Psi_3(\mathcal{W}^N)$ can be identified with a normalized 3-sign depth. More precisely, it can be shown that

$$\Psi_3(\mathcal{W}^N) = \frac{(N-1)(N-2)}{N}\left(d_3(R_1^\theta,\ldots,R_N^\theta) - \frac{1}{4}\right) \quad \mathbb{P}_{\theta^*}\text{-almost surely.} \tag{4.4}$$

From here, the limit law of the 3-sign depth follows by Slutzky's theorem. □

Equation (4.4) prompts us to think of $\Psi_3(\mathcal{W}^N)$ as a normalized 3-sign depth. The following theorem generalizes this relation to arbitrary $K \geq 3$. In Malcherczyk, Leckey and Müller 2021, it is incorporated into the proof of Theorem 4.4.

Theorem 4.5 (Normalized K-Sign Depth; Theorem 2.3. of Malcherczyk, Leckey and Müller 2021, p. 346)**.**
Let Ψ_K, $K \geq 3$, be the functional from Theorem 4.4 and let $\mathcal{W}^N$ be as in Equation (4.3)*.*

(i) Suppose that

$$\mathbb{P}_{\theta^*}(R_n^\theta = 0) = 0 \quad \textit{for } n = 1,\ldots,N.$$

Then, $\mathbb{P}_{\theta^*}$-almost surely,

$$N\left(d_3(R_1^\theta,\ldots,R_N^\theta) - \frac{1}{4}\right) = \frac{N^2}{(N-1)(N-2)}\Psi_3(\mathcal{W}^N)\,.$$

(ii) If $R_1^\theta,\ldots,R_N^\theta$ satisfy assumptions (K1) *and* (K2)*, then*

$$N\left(d_K(R_1^\theta,\ldots,R_N^\theta) - \left(\frac{1}{2}\right)^{K-1}\right) = \frac{N^K(N-K)!}{N!}\Psi_K(\mathcal{W}^N) + o_{\mathbb{P}}(1)\,,$$

[16] Note that $\mathbb{E}_{\theta^*}\left(\operatorname{sgn} R_n^\theta\right) = 0$ and $\operatorname{Var}_{\theta^*}\left(\operatorname{sgn} R_n^\theta\right) = 1$ according to (K2).

where $o_{\mathbb{P}}(1)$ converges to 0 in probability[17] *as $N \to \infty$.*

We denote hereafter the random variable

$$\Psi_K(\mathcal{W}^N(\theta)) = \frac{N!}{N^{K-1}(N-K)!}\left(d_K(R_1^\theta, \dots, R_N^\theta) - \left(\frac{1}{2}\right)^{K-1}\right)$$

as the normalized K-sign depth at θ.

For $K \geq 4$, Theorem 4.5 demonstrates that the integral representation of the K-sign depth holds only asymptotically. This further substantiates our focus on the case $K = 3$. We can take the normalized K-sign depth as the test statistic of an asymptotic hypothesis test based on the quantiles of $\Psi_K(W)$. For this, we consider hypotheses of the form

$$\mathcal{H}_0 : \theta^* \in \Theta_0 \quad \text{vs.} \quad \mathcal{H}_1 : \theta^* \in \Theta \setminus \Theta_0 \,, \tag{4.5}$$

where $\Theta_0 \subset \Theta$ is an arbitrary subset. Note that we still have not given any meaning to the "true" parameter. We will now rectify this by introducing a third assumption (K3).

Assumption 4.6 (Requirement for the K-Sign Depth Test)**.**
The parametric family from Equation (4.1) satisfies the assumption (K3) if for each true parameter value $\theta^* \in \Theta$,

$$R^{\theta^*} \text{ satisfies the assumptions (K1) and (K2).} \tag{K3}$$

We emphasize that (K3) does not require that (K1) and (K2) must apply for all $\theta \in \Theta$. We only demand that the assumptions (K1) and (K2) are satisfied at the true parameter θ^*, regardless of its value. Assumption (K3) ensures that the normalized K-sign depth converges to $\Psi_K(W)$ at the true parameter. This finally assigns a meaning to θ^* also in the context of Equation (4.1). Under assumption (K3), we can define an asymptotic level α test for the hypotheses from Equation (4.5): the K-sign depth test.

Definition 4.7 (K-Sign Depth Test)**.**
Let $\alpha \in (0, 1)$ be the desired significance level and suppose that the parametric family

$$\left\{R_n^\theta : n \in \{1, \dots, N\}, \theta \in \Theta\right\} \tag{4.1}$$

satisfies assumption (K3)*. For $K \geq 3$, the K-sign depth test for the hypotheses*

$$\mathcal{H}_0 : \theta^* \in \Theta_0 \quad \textit{vs.} \quad \mathcal{H}_1 : \theta^* \in \Theta \setminus \Theta_0$$

is the asymptotic level α test given by

$$\varphi_N = \mathbb{1}\left\{\sup_{\theta \in \Theta_0} \Psi_K(\mathcal{W}^N(\theta)) < q_\alpha(\Psi_K)\right\},$$

where $q_\alpha(\Psi_K)$ denotes the α-quantile of $\Psi_K(W)$ for a standard Brownian motion W.

Less formally, the K-sign depth test is given by the rule:

$$\text{reject } \mathcal{H}_0 \text{ if } \sup_{\theta \in \Theta_0} \Psi_K(\mathcal{W}^N(\theta)) < q_\alpha(\Psi_K)\,,$$

[17] The convergence holds with respect to $\mathbb{P}_{\theta^*}$. For aesthetic reasons we omit parameters in the Bachmann-Landau notation whenever they are clear from the context.

which means that the null hypothesis is rejected if the normalized K-sign depth for all parameters $\theta \in \Theta_0$ lies below the α-quantile of the asymptotic distribution at the true parameter. The conception of this test is borrowed from Müller 2005, p. 160, although the use of the supremum tends to result in conservative tests. We can easily verify that the asymptotic significance level of the test is at least α. For $\theta^* \in \Theta_0$, it follows from the monotonicity of the probability measure that

$$\begin{aligned}\lim_{N\to\infty} \mathbb{P}_{\theta^*}\left(\varphi_N = 1\right) &= \lim_{N\to\infty} \mathbb{P}_{\theta^*}\Big(\sup_{\theta\in\Theta_0} \Psi_K\big(\mathcal{W}^N(\theta)\big) < q_\alpha(\Psi_K)\Big) \\ &\leq \lim_{N\to\infty} \mathbb{P}_{\theta^*}\Big(\Psi_K\big(\mathcal{W}^N(\theta^*)\big) \leq q_\alpha(\Psi_K)\Big)\,.\end{aligned}$$

The condition (K3) ensures that $\Psi_K\big(\mathcal{W}^N(\theta^*) \to \Psi_K(W)$ in distribution, and hence

$$\lim_{N\to\infty} \mathbb{P}_{\theta^*}\Big(\Psi_K\big(\mathcal{W}^N(\theta^*)\big) \leq q_\alpha(\Psi_K)\Big) = \mathbb{P}_{\theta^*}\left(\Psi_K(W) \leq q_\alpha(\Psi_K)\right) = \alpha\,.$$

Combining these equations yields the desired significance level, since then

$$\lim_{N\to\infty} \mathbb{P}_{\theta^*}\left(\varphi_N = 1\right) \leq \alpha\,.$$

Having introduced the K-sign test for arbitrary natural $K \geq 3$, in the following we will restrict ourselves entirely to the case $K = 3$.

4.2. Consistency of the 3-Sign Depth Test

In order that a level α test φ_N is consistent for the hypotheses (4.5), it must hold that

$$\lim_{N\to\infty} \mathbb{P}_{\theta^*}\left(\varphi_N = 1\right) = 1 \quad \text{for all } \theta^* \notin \Theta_0\,, \tag{4.6}$$

which means that the power of the test converges to 1 for each fixed parameter under the alternative hypothesis. If we want to show that the 3-sign depth test from Definition 4.7 is consistent for any significance level $\alpha \in (0,1)$, we therefore have to prove that

$$\lim_{N\to\infty} \mathbb{P}_{\theta^*}\Big(\sup_{\theta\in\Theta_0} \Psi_3\big(\mathcal{W}^N(\theta)\big) < q_\alpha(\Psi_3)\Big) = 1 \quad \text{for all } \alpha \in (0,1) \text{ and } \theta^* \notin \Theta_0. \tag{4.7}$$

In Subsection 4.2.1, we will first examine conditions under which Equation (4.7) is satisfied, so that the 3-sign depth test is consistent. We then transfer these conditions to an intensity-based framework in Subsection 4.2.2.

4.2.1. General Consistency Conditions for the 3-Sign Depth Test

With the introduction of the rescaled random walk $\mathcal{W}^N(\theta)$ and the normalized K-sign depth $\Psi_K\big(\mathcal{W}^N(\theta)\big)$, a compact definition of the K-sign depth test became possible. While neither would be needed for an abstract definition of this test, they prove useful for formulating consistency conditions: As shown in Leckey, Jakubzik and Müller 2023, the consistency of the 3-sign depth test can be derived from the asymptotic behavior of

$$\left\|\mathcal{W}^N(\theta)\right\|_\infty := \sup_{t\in[0,1]} \left|\mathcal{W}^N_t(\theta)\right|,$$

that is, the maximum of $\mathcal{W}^N(\theta)$ on the interval $[0,1]$. This is particularly advantageous because the functional Ψ_3 constitutes the major complexity of the normalized 3-sign depth and even more so of its asymptotic distribution. We restate here the main theorem of Leckey, Jakubzik and Müller 2023.

Theorem 4.8 (Consistency of the 3-Sign Depth Test via Asymptotic Behavior of $\mathcal{W}^N(\theta)$; Theorem 3.1 of Leckey, Jakubzik and Müller 2023)**.**
Let Ψ_3 be the functional from Theorem 4.4 and let $\mathcal{W}^N(\theta)$ be as in Equation (4.3).
If, for $\theta^* \notin \Theta_0$,

$$\mathbb{P}_{\theta^*}\left(\liminf_{N\to\infty} \frac{\inf_{\theta\in\Theta_0} \|\mathcal{W}^N(\theta)\|_\infty}{\sqrt{N}} > 0\right) = 1\,, \tag{4.8}$$

then

$$\mathbb{P}_{\theta^*}\left(\lim_{N\to\infty} \sup_{\theta\in\Theta_0} \Psi_3(\mathcal{W}^N(\theta)) = -\infty\right) = 1\,. \tag{4.9}$$

Proof. Section 6.1 of Leckey, Jakubzik and Müller 2023 is dedicated to the proof of Theorem 4.8. For the basic proof idea, we recall that an integral appears in the definition of $\Psi_3(\mathcal{W}^N(\theta))$. This integral contains a transformation of $\mathcal{W}^N(\theta)$ as the integrand, namely

$$\left(\mathcal{W}_1^N(\theta) - 2\mathcal{W}_t^N(\theta)\right)^2.$$

If the condition from Equation (4.8) applies, this transformation tends to infinity on an interval of positive length. Consequently, the integral itself converges to infinity. Since the integral in $\Psi_3(\mathcal{W}^N(\theta))$ carries a negative sign, Equation (4.9) ensues. The main work lies in embedding this argument into a probabilistic context. □

If Equation (4.9) holds for all $\theta^* \notin \Theta_0$, then Equation (4.7) is satisfied and the 3-sign depth test is consistent. To see this, note that it follows immediately from Equation (4.9) that

$$\mathbb{P}_{\theta^*}\left(\lim_{N\to\infty} \sup_{\theta\in\Theta_0} \Psi_3(\mathcal{W}^N(\theta)) < q_\alpha(\Psi_3)\right) = 1 \quad \text{for all } \alpha \in (0,1)\,.$$

From here, the σ-continuity of the probability measure $\mathbb{P}_{\theta^*}$ then yields for all $\alpha \in (0,1)$:

$$\begin{aligned}
1 &= \mathbb{P}_{\theta^*}\left(\liminf_{N\to\infty}\left\{\sup_{\theta\in\Theta_0} \Psi_3(\mathcal{W}^N(\theta)) < q_\alpha(\Psi_3)\right\}\right) \\
&= \mathbb{P}_{\theta^*}\left(\lim_{N\to\infty} \bigcap_{n=N}^{\infty}\left\{\sup_{\theta\in\Theta_0} \Psi_3(\mathcal{W}^n(\theta)) < q_\alpha(\Psi_3)\right\}\right) \\
&= \lim_{N\to\infty} \mathbb{P}_{\theta^*}\left(\bigcap_{n=N}^{\infty}\left\{\sup_{\theta\in\Theta_0} \Psi_3(\mathcal{W}^n(\theta)) < q_\alpha(\Psi_3)\right\}\right) \\
&\le \lim_{N\to\infty} \mathbb{P}_{\theta^*}\left(\sup_{\theta\in\Theta_0} \Psi_3(\mathcal{W}^N(\theta)) < q_\alpha(\Psi_3)\right)\,.
\end{aligned}$$

Thus, we recognize that Equation (4.8) does indeed provide a sufficient condition for the consistency of the 3-sign depth test. Leckey, Jakubzik and Müller 2023 record this conclusion in their Corollary 3.2.
In order for Equation (4.8) to be fulfilled, it often suffices to find an interval $[t_1,t_2] \subset [0,1]$, $t_1 < t_2$ such that the increment $|\mathcal{W}_{t_2}^N(\theta) - \mathcal{W}_{t_1}^N(\theta)|$ is atypically large. As this increment is

a rescaled sum of the signs of consecutive random variables $R^\theta_m, R^\theta_{m+1}, \ldots, R^\theta_n$ for suitable integers $m < n$, we need to identify index regions at which these random variables tend to have either unusually many negative or positive signs (cf. Subsection 3.1 of Leckey, Jakubzik and Müller 2023). To formalize this, we define

$$\mathcal{N}^+_{m,n}(\theta) := \sum_{j=m}^{n} \mathbb{1}\{R^\theta_j > 0\} \quad \text{and} \quad \mathcal{N}^-_{m,n}(\theta) := \sum_{j=m}^{n} \mathbb{1}\{R^\theta_j < 0\} .$$

The following lemma relates these random variables to $\|\mathcal{W}^N(\theta)\|_\infty$.

Lemma 4.9 (Sufficient Condition for Large $\|\mathcal{W}^N(\theta)\|_\infty$; Lemma 3.3 of Leckey, Jakubzik and Müller 2023)**.**
Let $\varepsilon > 0$ and $N \in \mathbb{N}$. Then, for all $\theta \in \Theta$ holds:

$$\sup_{n-m+1 \geq \varepsilon N} \frac{\max\left\{\mathcal{N}^+_{m,n}(\theta), \mathcal{N}^-_{m,n}(\theta)\right\}}{n-m+1} - \frac{1}{2} > \varepsilon \quad \Longrightarrow \quad \|\mathcal{W}^N(\theta)\|_\infty \geq \varepsilon^2\sqrt{N} , \tag{4.10}$$

where the supremum is taken over all $m, n \in \{1, \ldots, N\}$ with $n - m + 1 \geq \varepsilon N$.

Proof. The proof is due to Leckey, Jakubzik and Müller 2023 and is given there in Subsection 6.2. Since it is rather short and motivates the following steps, we reproduce it here for convenience. Without loss of generality, we may start by assuming that

$$\frac{\mathcal{N}^+_{m,n}(\theta)}{n-m+1} - \frac{1}{2} > \varepsilon \tag{4.11}$$

for some integers $m < n$ with $n - m + 1 \geq \varepsilon N$, since replacing R^θ_j by $-R^\theta_j$ has no effect on $\|\mathcal{W}^N(\theta)\|_\infty$. Moreover, we observe that

$$\mathcal{N}^+_{m,n}(\theta) + \mathcal{N}^-_{m,n}(\theta) \leq n - m + 1 , \tag{4.12}$$

where equality holds if the R^θ_j are non-zero for all $j \in \{m, \ldots, n\}$. By combining Equations (4.11) and (4.12), we obtain:

$$\begin{aligned}
\mathcal{W}^N_{\frac{n}{N}}(\theta) - \mathcal{W}^N_{\frac{m-1}{N}}(\theta) &= \frac{1}{\sqrt{N}} \sum_{j=m}^{n} \operatorname{sgn} R^\theta_j = \frac{1}{\sqrt{N}} \left(\mathcal{N}^+_{m,n}(\theta) - \mathcal{N}^-_{m,n}(\theta)\right) && |\text{apply Eq. (4.12)} \\
&\geq \frac{1}{\sqrt{N}} \left(\mathcal{N}^+_{m,n}(\theta) - \left((n-m+1) - \mathcal{N}^+_{m,n}(\theta)\right)\right) \\
&= \frac{1}{\sqrt{N}} \left(2\mathcal{N}^+_{m,n}(\theta) - (n-m+1)\right) && |\text{apply Eq. (4.11)} \\
&\geq \frac{1}{\sqrt{N}} 2\varepsilon(n-m+1) \geq 2\varepsilon^2\sqrt{N} ,
\end{aligned}$$

as $n - m + 1 \geq \varepsilon N$ by assumption. The triangle inequality then yields

$$\varepsilon^2\sqrt{N} \leq \frac{1}{2}\left|\mathcal{W}^N_{\frac{n}{N}}(\theta) - \mathcal{W}^N_{\frac{m-1}{N}}(\theta)\right| \leq \|\mathcal{W}^N(\theta)\|_\infty$$

and therefore the assertion. □

If for $\varepsilon > 0$ and $N \in \mathbb{N}$ we define the event

$$\mathbb{A}_{N,\varepsilon}(\theta) := \left\{ \sup_{n-m+1 \geq \varepsilon N} \frac{\max\left\{\mathcal{N}^+_{m,n}(\theta), \mathcal{N}^-_{m,n}(\theta)\right\}}{n-m+1} - \frac{1}{2} > \varepsilon \right\}, \tag{4.13}$$

then the implication from Equation (4.10) translates to

$$\mathbb{A}_{N,\varepsilon}(\theta) \subset \left\{ \left\|\mathcal{W}^N(\theta)\right\|_\infty \geq \varepsilon^2 \sqrt{N} \right\}.$$

Hence,

$$\bigcap_{\theta \in \Theta_0} \mathbb{A}_{N,\varepsilon}(\theta) \subset \bigcap_{\theta \in \Theta_0} \left\{ \left\|\mathcal{W}^N(\theta)\right\|_\infty \geq \varepsilon^2 \sqrt{N} \right\} = \left\{ \inf_{\theta \in \Theta_0} \left\|\mathcal{W}^N(\theta)\right\|_\infty \geq \varepsilon^2 \sqrt{N} \right\}.$$

This property carries over to the set-theoretic limit inferior, and we further obtain:

$$\begin{aligned}
\liminf_{N \to \infty} \bigcap_{\theta \in \Theta_0} \mathbb{A}_{N,\varepsilon}(\theta) &\subset \liminf_{N \to \infty} \left\{ \inf_{\theta \in \Theta_0} \left\|\mathcal{W}^N(\theta)\right\|_\infty \geq \varepsilon^2 \sqrt{N} \right\} \\
&\subset \left\{ \liminf_{N \to \infty} \frac{\inf_{\theta \in \Theta_0} \left\|\mathcal{W}^N(\theta)\right\|_\infty}{\sqrt{N}} \geq \varepsilon^2 \right\} \\
&\subset \left\{ \liminf_{N \to \infty} \frac{\inf_{\theta \in \Theta_0} \left\|\mathcal{W}^N(\theta)\right\|_\infty}{\sqrt{N}} > 0 \right\}.
\end{aligned} \tag{4.14}$$

For the step from Equation (4.14), note that we have

$$\begin{aligned}
&\omega \in \liminf_{N \to \infty} \left\{ \inf_{\theta \in \Theta_0} \left\|\mathcal{W}^N(\theta)\right\|_\infty \geq \varepsilon^2 \sqrt{N} \right\} \\
\iff \quad & \exists N_0 = N_0(\omega) \in \mathbb{N}\ \forall N \geq N_0 \ : \ \underbrace{\frac{\inf_{\theta \in \Theta_0} \left\|\mathcal{W}^N(\theta)\right\|_\infty}{\sqrt{N}}}_{\text{depends on } \omega} \geq \varepsilon^2 \\
\implies \quad & \liminf_{N \to \infty} \frac{\inf_{\theta \in \Theta_0} \left\|\mathcal{W}^N(\theta)\right\|_\infty}{\sqrt{N}} \geq \varepsilon^2 .
\end{aligned}$$

This train of thought leads us to (and proves) another consistency criterion for the 3-sign depth test based on index regions with either atypically many negative or positive signs among their associated random variables. It is our criterion of choice in an intensity-based framework and we will encounter it several times in the remainder of this chapter.

Corollary 4.10 (Consistency of the 3-Sign Depth Test via Atypical Index Regions; Corollary 3.4 of Leckey, Jakubzik and Müller 2023)**.**
Let $\mathbb{A}_{N,\varepsilon}$ be defined as in Equation (4.13)*. If for $\theta^* \notin \Theta_0$ there exists $\varepsilon > 0$ such that*

$$\mathbb{P}_{\theta^*}\left(\liminf_{N \to \infty} \bigcap_{\theta \in \Theta_0} \mathbb{A}_{N,\varepsilon}(\theta) \right) = 1, \tag{4.15}$$

then Equation (4.8) *of Theorem 4.8 holds. In particular, if Equation* (4.15) *holds for every $\theta^* \notin \Theta_0$, then the 3-sign depth test is consistent.*

4.2.2. Consistency Conditions for the 3-Sign Depth Test in Intensity-Based Models

If we wish to apply statistical methods based on the K-sign depth in an intensity-based framework, we quickly encounter a central issue: how can we obtain proper random variables R_n^θ from the observable point processes, and according to which order should they be arranged? Our aim is to find a suitable counterpart to the residuals of a classical regression approach (cf. Subsection 4.2.1) to which the 3-sign depth test can be applied. The flexible (and in a sense distribution free[18]) modelling based on cumulative intensities prevents the definition of residuals directly based on the distribution of point processes, although this is occasionally feasible depending on the specific model. Our solution is to consider transformations of the corresponding point processes that meet certain requirements to have them act as substitutes for the residuals.
In order to formulate such transformations and their requirements, we first specify the framework in analogy to Definition 3.2 of the previous chapter. Unlike the framework of Section 2.1, we allow that the counting processes follow one of up to $L \in \mathbb{N}$ distinct distributions. This later permits us to incorporate L different experimental conditions without the need for random covariates. The main reason to avoid random covariates here is that they interact unfavourably with our chosen transformation - namely, the hazard transformation.

Definition 4.11 (Intensity-Based Framework for a Consistent 3-Sign Depth Test)**.**
Let $\mathring{N}^{(1)}, \ldots, \mathring{N}^{(L)}$, $L \in \mathbb{N}$, be independent, adapted counting processes on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in \mathcal{I}}, \mathbb{P})$, where $\mathcal{I} = [0, \infty)$. For each $j \in \mathbb{N}$, let $N^{(j)}$ be an i.i.d. copy of one of these processes and let $l_j \in \{1, \ldots, L\}$ indicate which distribution this counting process follows (i.e., the jth counting process is an i.i.d. copy of $\mathring{N}^{(l_j)}$). The $(\mathbb{P}, \{\mathcal{F}_t\}_{t \geq 0})$-compensator of $N^{(j)}$ is denoted by $\Lambda^{(j)}$. Let $\theta \in \Theta$ be the parameter of interest, where $\Theta \subset \mathbb{R}^d$, $d \in \mathbb{N}$, is the parameter space. For each $l \in \{1, \ldots, L\}$, let a parametric model be given by a class $\mathcal{M}_l$ of cumulative intensities,

$$\mathcal{M}_l = \{\Lambda_{l,\theta} \, : \, \theta \in \Theta\} \, .$$

It is noteworthy that Definition 4.11 still incorporates the previous i.i.d. case by choosing $L = 1$, since then all processes are i.i.d. copies of the single counting process $\mathring{N}^{(1)}$. In comparison to the framework for the minimum distance estimation, we were able to apply three mitigations here: First, we did not demand absolute continuity of the compensators; second, we did not impose any restrictions on the parameter range up to this point; and third, we weakened the i.i.d. requirement and allowed for up to L different distributions. The last point, though, entails its own set of assumptions, specifically that the counting processes are terminating and that all L distributions occur with an asymptotically non-vanishing proportion. This ensures that when the number of replications tends to infinity, infinitely many i.i.d. observations are made for all of these distributions or, as we will regularly call them, “classes”. We record these first assumptions, along with other general requirements on the parametric model, in Assumption 4.12 below.

[18]In the applications we consider, the probability structure of the counting processes is already fully determined by the model according to Proposition A.35. Nevertheless, in general we will not deal with the distribution of the associated point process.

Assumption 4.12 (General Requirements for the Parametric Model)**.**
As we did for the minimum distance estimator, we accumulate a number of assumptions throughout this section. We refer to them only by the letter D (for depth), since here the total number is much more moderate.

(D1) For each $l \in \{1, \ldots, L\}$, the model $\mathcal{M}_l$ contains the compensator Λ_l of $\mathring{N}^{(l)}$, so there exists a true parameter $\theta^* \in \Theta$ such that

$$\Lambda_l = \Lambda_{l,\theta^*} .$$

The true parameter θ^* is the same for all models and thus does not depend on l.

(D2) There exist constants $c_l \in \mathbb{N}$, $l \in \{1, \ldots, L\}$, such that $\mathring{N}_t^{(l)} \to c_l$ as $t \to \infty$ with probability one.

(D3) Each of the L distributions occurs with an asymptotic proportion of $p_l > 0$, that is,

$$\lim_{J \to \infty} \frac{1}{J} \sum_{j=1}^{J} \mathbb{1}_{\{l\}}(l_j) =: p_l > 0 , \quad \text{for each } l \in \{1, \ldots, L\}.$$

The total number of observations is denoted by $\eta := \sum_{j=1}^{J} c_{l_j}$.

The condition (D1) is a limiting factor regarding the freedom of modelling, since cumulative intensities may vary between the different classes, but must still share a true parameter. Therefore, the most common application will be that the processes are allowed to differ in terms of deterministic covariates (e.g., different initial stress levels) where they were previously subject to random external effects. However, the respective cumulative intensities may also contain different non-parametric components.
Note that, as a consequence of (D2), the counting processes $\mathring{N}^{(1)}, \ldots, \mathring{N}^{(L)}$ are almost surely terminating (the boundedness condition here is obviously even stronger).
The third condition (D3) prevents asymptotically negligible classes, that is, classes that occur only finitely often or infinitely often with a vanishing proportion. Possible attenuations of this condition are the subject of the following remark.

Remark 4.13 (On Asymptotically Negligible Classes)**.**
If condition (D3) is violated because for some $l \in \{1, \ldots, L\}$ we have

$$\frac{1}{J} \sum_{j=1}^{J} \mathbb{1}_{\{l\}}(l_j) \to 0 \quad (J \to \infty) ,$$

we say that the class l is asymptotically negligible. The results of this section require in principle only that $L \geq 1$ asymptotically non-negligible classes exist. Inconveniences then arise in the precise mathematical formulation, since the asymptotically negligible classes must always be treated separately. This unnecessarily dilutes the following proofs, which is why we will restrict ourselves here to the case of asymptotically non-negligible classes.

We have now reached the point where we can introduce proper transformations of the observable point processes. These transformations may also take into account the past of the processes. In a sense, this mirrors the autoregressive models mentioned earlier in Remark 4.2. The further conditions are necessary to later derive the consistency of the 3-sign depth test by virtue of Corollary 4.10.

Assumption 4.14 (Requirements Regarding the Transformation)**.**
For each $l \in \{1, \ldots, L\}$ and $i \in \{1, \ldots, c_l\}$, let $f_{l,i}$ be a real-valued function on $\Theta \times [0, \infty)^i$ and define the transformation

$$R_{j,i}^{\theta} := f_{l_j,i}\left(\theta, T_i^{(j)}, \ldots, T_1^{(j)}\right) .$$

The transforms $R_{j,i}^{\theta}$ are subject to the following conditions:

(D4) At $\theta = \theta^*$, the transforms have the distributional properties given as follows:

(i) The family $\left(R_{j,i}^{\theta^*}\right)_{\substack{j \in \mathbb{N}, \\ i \in \{1,\ldots,c_{l_j}\}}}$ is independent (with respect to $\mathbb{P}_{\theta^*}$).

(ii) For all $j \in \mathbb{N}$ and $i \in \{1, \ldots, c_{l_j}\}$ holds:

$$\mathbb{P}_{\theta^*}\left(R_{j,i}^{\theta^*} > 0\right) = \frac{1}{2} = \mathbb{P}_{\theta^*}\left(R_{j,i}^{\theta^*} < 0\right) .$$

(D5) For each $l \in \{1, \ldots, L\}$ with $c_l > 1$ and $i \in \{2, \ldots, c_l\}$, let

$$\mathbb{P}_{\theta^*}\left(R_{j,i}^{\theta} > 0 \,\middle|\, T_{1:(i-1)}^{(j)}\right) = g_{l_j,i}\left(\theta^*, \theta, A_{j,i}\right) ,$$

where $g_{l,i} \,:\, \Theta^2 \times \mathbb{R} \to \mathbb{R}$ is a continuous function and

$$A_{j,i} := a_{l_j,i} \circ \left(T_1^{(j)}, \ldots, T_{i-1}^{(j)}\right) \quad \text{for some function } a_{l,i} \,:\, [0,\infty)^{i-1} \to \mathbb{R} .$$

(D6) For each $l \in \{1, \ldots, L\}$ with $c_l > 1$ and $i \in \{2, \ldots, c_l\}$, there exists an open interval $U_{l,i}$ such that on $\{A_{j,i} \in U_{l_j,i}\}$ the random variable $R_{j,i}^{\theta}$ is monotone in each component of θ. Formally, we demand that some $\rho = (\rho_1, \ldots, \rho_d) \in \{0,1\}^d$ exists such that for all $\theta = (\theta_1, \ldots, \theta_d) \in \Theta$, $k \in \{1, \ldots, d\}$ and $\omega \in \{A_{j,i} \in U_{l_j,i}\}$ we have

$$\theta_k \leq \tilde{\theta}_k \quad \Longrightarrow \quad (-1)^{\rho_k} R_{j,i}^{(\theta_1,\ldots,\theta_k,\ldots,\theta_d)}(\omega) \leq (-1)^{\rho_k} R_{j,i}^{(\theta_1,\ldots,\tilde{\theta}_k,\ldots,\theta_d)}(\omega) .$$

In particular, ρ must not depend on θ or ω.

The property (D4) emulates the conditions usually imposed on residuals in a classical regression model. Except for the indexing of the random variables, the conditions (K3) and (D4) agree. In particular, condition (D4) ensures that Assumption 4.6 is satisfied. It is needed for the 3-sign depth test to meet the requirements of a statistical hypothesis test, because otherwise the type I error could not be controlled. On the other hand, the conditions (D5) and (D6) later provide the consistency of the test. In (D5), the function $a_{l,i}$ comprises all the information that the past of the process contributes to the probability of a positive transform, while the function $g_{l,i}$ links this information with the influence of the parameters. Often $A_{j,i}$ will be a weighted sum of the time points $T_1^{(j)}, \ldots, T_{i-1}^{(j)}$. Technically (and analogously to Remark 4.13), the relationship given in (D5) need not exist for all i and l. It is sufficient if it holds for selected indices, so that the prerequisites stated later are fulfilled. The same applies to the conditional monotonicity condition (D6). The corresponding Theorem 4.17 can very easily be adapted to reflect this situation, but this would again needlessly inflate its formulation.

We postpone the formulation of this theorem for a while to address one remaining problem: Given that the 3-sign depth is highly dependent on the ordering of the observations, how should the double-indexed transforms $R^{\theta}_{j,i}$ be ordered? While in a finite sample the transforms apparently posses a canonical order with respect to i for each fixed $j \in \{1, \dots, J\}$, we need to specify the ordering with respect to j as $J \to \infty$. Clearly, it is not useful to simply append new observations to the end and completely neglect the concept of depth. Instead, assumption (D5) allows us to apply a random ordering based on the past of the counting processes, the total order $\leq_{\text{acc}}$. Moreover, this ordering preserves the canonical order with respect to i for each $j \in \mathbb{N}$.

Definition 4.15 (Total Order $\leq_{\text{acc}}$).
Let $A_{j,i}$ be defined as in Assumption 4.14 and set $A_{j,1} \equiv 0$ for all $j \in \mathbb{N}$. The total order $\leq_{acc}$ on the set

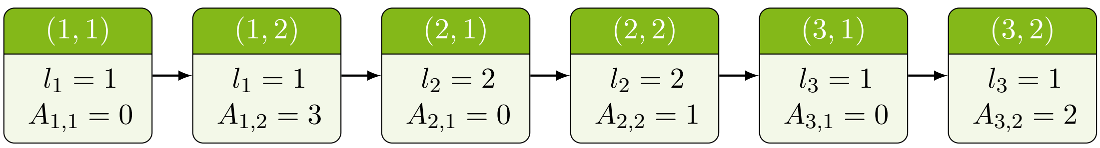
$$\mathcal{R}_{observed} := \left\{ (j,i) \in \mathbb{N} \times \mathbb{N} \,:\, 1 \leq i \leq c_{l_j} \right\}$$

is then defined by

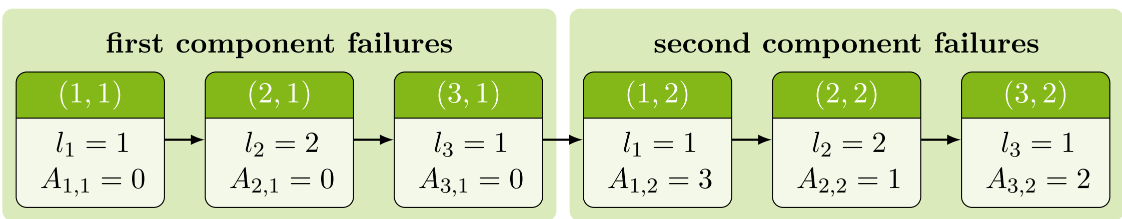
$$\begin{aligned} (j,i) \leq_{acc} (n,m) \quad :\Longleftrightarrow \quad i < m \quad & \vee \quad (i = m \ \wedge \ l_j < l_n) \\ & \vee \quad (i = m \ \wedge \ l_j = l_n \ \wedge \ A_{j,i} \leq A_{n,m}) \,. \end{aligned}$$

This total order hence sorts the observed transforms first by the number i of realized time points, then by the underlying distribution l_j, and finally by the essential past $A_{j,i}$ of the process. We illustrate this three-step ordering in Example 4.16.

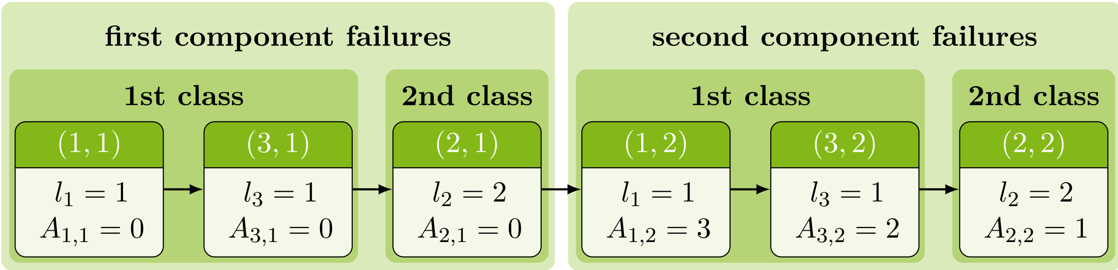
Example 4.16 (Visualization of the Total Order $\leq_{\text{acc}}$).
To demonstrate the ordering from Definition 4.15, we consider $L = 2$ two classes. Suppose that we run $J = 3$ experiments with $l_1 = l_3 = 1$ and $l_2 = 2$, which means that the first and third experiments were conducted under the same experimental conditions, while the second experiment had a different setup. In both cases, we assume that $c_1 = c_2 = 2$ component failures can be observed. If we perform the experiments sequentially, we then obtain the observations in the order $(1,1)$ (first experiment, first failure), $(1,2)$ (first experiment, second failure), $(2,1)$ (second experiment, first failure) and so on:



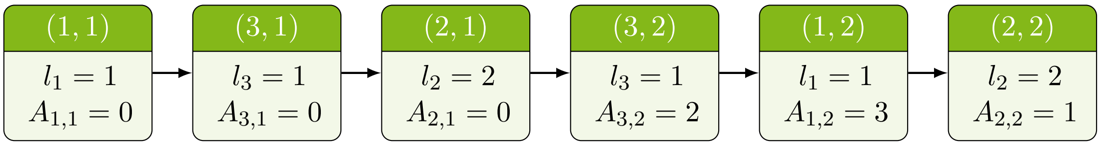
Here we have included the realizations $A_{j,i}(\omega)$ of each of the random variables $A_{j,i}$ (we omit the argument ω for brevity), which will be used in the last step of the ordering.
In the first step, the observations are ordered by the number of component failures, which simply corresponds to the second component of the (j,i):



In the second step, we then further order the observations by their corresponding class l_j:



In the third and final step, we arrange the observations with the same number of component failures and from the same class by their essential past $A_{j,i}$: Because of $A_{1,1}(\omega) = 0 = A_{3,1}(\omega)$, the order of $(1,1)$ and $(3,1)$ is interchangeable, but $A_{1,2}(\omega) = 3 > 2 = A_{3,2}(\omega)$ implies $(1,2) \not\leq_{\text{acc}} (3,2)$ so that $(3,2)$ and $(1,2)$ must be switched. Overall, this gives us the following ordering with respect to $\leq_{\text{acc}}$:



We have now assembled all the components to state the central theorem of this section. It represents an application of Corollary 4.10 to an intensity-based framework, and itself provides a consistency criterion for the 3-sign depth test. The essential condition of Theorem 4.17 - given in Equation (4.16) - will turn out to be particularly easy to verify for the load-sharing models under consideration.

Theorem 4.17 (Consistency of the 3-Sign Depth Test for Compact Θ_0)**.**
In the framework of Definition 4.11, suppose that the conditions (D1) *through* (D6) *of Assumptions 4.12 and 4.14 are fulfilled. Let* $\Theta = \Theta_0 \cup \Theta_1$ *with* $\Theta_0 \cap \Theta_1 = \emptyset$*, where* Θ_0 *is compact. If for each* $\theta^* \in \Theta_1$ *and* $\theta \in \Theta_0$ *there exist* $l \in \{1, \dots, L\}$*,* $i \in \{2, \dots, c_l\}$ *and* $x \in \operatorname{supp}(A_{j,i}) \cap U_{l,i}$ *for* j *with* $l_j = l$ *satisfying*

$$g_{l_j,i}(\theta^*, \theta, x) \neq \frac{1}{2}, \tag{4.16}$$

then the 3*-sign depth test with transforms* $R^{\theta}_{j,i}$ *ordered according to* $\leq_{acc}$ *is consistent for the hypotheses*

$$\mathcal{H}_0 : \theta^* \in \Theta_0 \quad \textit{vs.} \quad \mathcal{H}_1 : \theta^* \in \Theta_1 .$$

The proof of Theorem 4.17 requires a basic lemma, which we state here in advance.

Lemma 4.18 (Limit Superior of Finite Unions of Asymptotic Null Sets)**.**
Let $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and* $A^{(i)} := (A_n^{(i)})_{n\in\mathbb{N}} \subset \mathcal{A}$ *for* $i \in \mathbb{I} \subset \mathbb{N}$ *with* $|\mathbb{I}| < \infty$*. Suppose that* $\mathbb{P}(\limsup_{n\to\infty} A_n^{(i)}) = 0$ *holds for each* $i \in \mathbb{I}$*. Then,*

$$\mathbb{P}\left(\limsup_{n\to\infty} \bigcup_{i\in\mathbb{I}} A_n^{(i)}\right) = 0 . \tag{4.17}$$

Proof. At first, note that for each $i \in \mathbb{I}$ we have by assumption and the continuity of the probability measure $\mathbb{P}$ that

$$\lim_{n\to\infty} \mathbb{P}\left(\bigcup_{m\geq n} A_m^{(i)}\right) = \mathbb{P}\left(\lim_{n\to\infty} \bigcup_{m\geq n} A_m^{(i)}\right) = \mathbb{P}\left(\limsup_{n\to\infty} A_n^{(i)}\right) = 0\,. \tag{4.18}$$

For an arbitrary $\varepsilon > 0$, there exists $n_0 \in \mathbb{N}$ so that according to Equation (4.18) the following holds simultaneously for all $i \in \mathbb{I}$:

$$\mathbb{P}\left(\bigcup_{m\geq n_0} A_m^{(i)}\right) < \frac{\varepsilon}{|\mathbb{I}|}\,, \quad \text{for all } i \in \mathbb{I}\,. \tag{4.19}$$

From Equation (4.19), we easily obtain by the subadditivity of the probability measure $\mathbb{P}$:

$$\mathbb{P}\left(\limsup_{n\to\infty} \bigcup_{i\in\mathbb{I}} A_n^{(i)}\right) \leq \mathbb{P}\left(\bigcup_{m\geq n_0} \bigcup_{i\in\mathbb{I}} A_m^{(i)}\right) \leq \sum_{i\in\mathbb{I}} \mathbb{P}\left(\bigcup_{m\geq n_0} A_m^{(i)}\right) < \varepsilon\,, \tag{4.20}$$

therefore completing the proof. □

We now return to the proof of Theorem 4.17.

Proof of Theorem 4.17. By the equivalence of norms on finite dimensional spaces, we can use the maximum norm $\|\cdot\|_\infty$ throughout this proof. This becomes useful later when we exploit the conditional monotonicity of the transforms in each component of θ.
Fix any $\theta^* \in \Theta_1$. Then, for each $\theta \in \Theta_0$, there exist $l = l(\theta)$, $i = i(\theta)$ and $x = x(\theta) \in \operatorname{supp}(A_{j,i}) \cap U_{l,i}$ (where $l_j = l$) such that

$$g_{l_j,i}(\theta^*, \theta, x) \neq \frac{1}{2}\,.$$

The random variables $A_{j,i}$ for j with $l_j = l$ and fixed i are i.i.d. by construction. We choose $j_0 := \min\{j : l_j = l\}$ to avoid ambiguity later on.
Due to the continuity of $g_{l_j,i}$, for any given $\varepsilon = \varepsilon(\theta) < \left|g_{l_j,i}(\theta^*, \theta, x) - \frac{1}{2}\right|$ we now find $\delta = \delta(\theta) > 0$ and an open interval $\mathrm{Int} = \mathrm{Int}(\theta) \ni x$ with $\mathrm{Int} \subset U_{l,i}$ (since $x \in U_{l,i}$ and $U_{l,i}$ is open) and $\mathbb{P}_{\theta^*}(A_{j,i} \in \mathrm{Int}) > 0$ (since $x \in \operatorname{supp}(A_{j,i})$) so that either

$$g_{l_j,i}\left(\theta^*, \tilde{\theta}, \tilde{x}\right) > \frac{1}{2} + \varepsilon \quad \text{for all } \tilde{\theta} \in \mathrm{B}_{2\delta}(\theta)\,, \tilde{x} \in \mathrm{Int}\,, \tag{4.21}$$

or

$$g_{l_j,i}\left(\theta^*, \tilde{\theta}, \tilde{x}\right) < \frac{1}{2} - \varepsilon \quad \text{for all } \tilde{\theta} \in \mathrm{B}_{2\delta}(\theta)\,, \tilde{x} \in \mathrm{Int}\,. \tag{4.22}$$

For convenience, we can assume below that Int is a compact interval with the same properties. Note the factor 2 used here, which allows us some leeway later when we take a closer look at the boundaries of the smaller sets $\mathrm{B}_\delta(\theta)$. Moreover, these open neighbourhoods $\mathrm{B}_\delta(\theta)$ form an open cover of the compact set Θ_0, for which a finite subcover can be found (we write δ_ν instead of $\delta(\theta_\nu)$):

$$\Theta_0 \subset \bigcup_{\nu=1}^{\nu_{\max}} \mathrm{B}_{\delta_\nu}(\theta_\nu)\,.$$

Let $\varepsilon_0 := \min_\nu \varepsilon(\theta_\nu)$. For each $\nu = 1, \dots, \nu_{\max}$, we proceed to show that, for some $\varepsilon > 0$,

$$\mathbb{P}_{\theta^*}\left(\liminf_{\eta\to\infty} \bigcap_{\theta \in \mathrm{B}_{\delta_\nu}(\theta_\nu)} \mathbb{A}_{\eta,\varepsilon}(\theta)\right) = 1\,,$$

where $\mathbb{A}_{\eta,\varepsilon}(\theta)$ is given as in Equation (4.13), that is,

$$\mathbb{A}_{\eta,\varepsilon}(\theta) := \left\{\sup_{n-m+1\geq\varepsilon\eta} \frac{\max\left\{\mathcal{N}^+_{m,n}(\theta), \mathcal{N}^-_{m,n}(\theta)\right\}}{n-m+1} - \frac{1}{2} > \varepsilon\right\}.$$

We will treat only the case from Equation (4.21) and examine the events

$$\left\{\sup_{n-m+1\geq\varepsilon\eta} \frac{\mathcal{N}^+_{m,n}(\theta)}{n-m+1} - \frac{1}{2} > \varepsilon\right\} \subset \mathbb{A}_{\eta,\varepsilon}(\theta)\,.$$

The case from Equation (4.22) can be handled analogously by replacing $\mathcal{N}^+_{m,n}(\theta)$ with $\mathcal{N}^-_{m,n}(\theta)$. Choose any $\nu \in \{1, \dots, \nu_{\max}\}$ and define

$$\mathcal{N}^+_{\mathrm{Int}}(\theta) := \sum_{j\in J_l} \mathbb{1}\left\{R^\theta_{j,i} > 0,\ A_{j,i} \in \mathrm{Int}\right\}\,, \quad \theta \in \mathrm{B}_{2\delta_\nu}(\theta_\nu)\,, \tag{4.23}$$

where $J_l := \{j \in \{1, \dots, J\} : l_j = l\}$ and $i = i(\theta_\nu)$, $l = l(\theta_\nu)$, $\mathrm{Int} = \mathrm{Int}(\theta_\nu)$ are fixed (we suppress the dependence on the indices i and l in the notation $\mathcal{N}^+_{\mathrm{Int}}(\theta)$). By construction, the summation takes place over independent (the index of summation is j) and identically distributed ($l_j = l$ for each $j \in J_l$) random variables. Since $|J_l| \to \infty$ as $J \to \infty$, the strong law of large numbers provides :

$$\begin{aligned}\frac{1}{|J_l|}\mathcal{N}^+_{\mathrm{Int}}(\theta) &= \frac{1}{|J_l|}\sum_{j\in J_l} \mathbb{1}\left\{R^\theta_{j,i} > 0,\ A_{j,i} \in \mathrm{Int}\right\} \\ &\overset{J\to\infty}{\longrightarrow} \mathbb{P}_{\theta^*}\left(R^\theta_{j_0,i} > 0,\ A_{j_0,i} \in \mathrm{Int}\right) \quad \mathbb{P}_{\theta^*}\text{-almost surely.}\end{aligned}$$

Likewise, we obtain

$$\frac{1}{|J_l|}\sum_{j\in J_l} \mathbb{1}\left\{A_{j,i} \in \mathrm{Int}\right\} \overset{J\to\infty}{\longrightarrow} \mathbb{P}_{\theta^*}\left(A_{j_0,i} \in \mathrm{Int}\right) \quad \mathbb{P}_{\theta^*}\text{-almost surely.}$$

Therefore, combining the above findings yields:

$$\begin{aligned}\frac{\mathcal{N}^+_{\mathrm{Int}}(\theta)}{\sum_{j\in J_l} \mathbb{1}\left\{A_{j,i} \in \mathrm{Int}\right\}} - \frac{1}{2} &\overset{J\to\infty}{\longrightarrow} \frac{\mathbb{P}_{\theta^*}\left(R^\theta_{j_0,i} > 0,\ A_{j_0,i} \in \mathrm{Int}\right)}{\mathbb{P}_{\theta^*}\left(A_{j_0,i} \in \mathrm{Int}\right)} - \frac{1}{2} \\ &\overset{(*)}{=} \frac{\int_{\left\{A_{j_0,i}\in\mathrm{Int}\right\}} g_{l_{j_0},i}\left(\theta^*, \theta, A_{j_0,i}\right) \mathrm{d}\mathbb{P}_{\theta^*}}{\mathbb{P}_{\theta^*}\left(A_{j_0,i} \in \mathrm{Int}\right)} - \frac{1}{2} \\ &> \frac{\left(\frac{1}{2} + \varepsilon_0\right)\mathbb{P}_{\theta^*}\left(A_{j_0,i} \in \mathrm{Int}\right)}{\mathbb{P}_{\theta^*}\left(A_{j_0,i} \in \mathrm{Int}\right)} - \frac{1}{2} = \varepsilon_0 \quad \mathbb{P}_{\theta^*}\text{-almost surely.}\end{aligned}$$

The identity $(*)$ follows from condition (D5): For this, note that for arbitrary random variables $Y = h(Z)$, X, Z and sets B, C the property $\mathbb{P}(X \in B \,|\, Z) = g(h(Z))$ implies with the tower property that

$$\begin{aligned}\mathbb{P}\left(X \in B, Y \in C\right) &= \mathbb{E}\left(\mathbb{1}_B(X) \cdot \mathbb{1}_C(h(Z))\right) = \mathbb{E}\left[\mathbb{E}\left(\mathbb{1}_B(X) \cdot \mathbb{1}_C(h(Z)) \,|\, Z\right)\right] \\ &= \mathbb{E}\left[\mathbb{1}_C(h(Z)) \cdot \mathbb{E}\left(\mathbb{1}_B(X) \,|\, Z\right)\right] = \mathbb{E}\left[\mathbb{1}_C(h(Z)) \cdot \mathbb{P}\left(X \in B \,|\, Z\right)\right] \\ &= \mathbb{E}\left[\mathbb{1}_C(h(Z)) \cdot g(h(Z))\right] = \int_{\{Y \in C\}} g(h(Z)) \,\mathrm{d}\mathbb{P}\,.\end{aligned}$$

Here, $X = R^{\theta}_{j_0,i}$, $Y = A_{j_0,i}$, $Z = \left(T_1^{(j_0)}, \dots, T_{i-1}^{(j_0)}\right)^\top$, $g = g_{l_{j_0},i}$, $h = a_{l_{j_0},i}$ and $B = (0,\infty)$, $C = \mathrm{Int}$. To proceed with the proof, note that $|J_l| = \sum_{j=1}^J \mathbb{1}_{\{l\}}(l_j)$ by definition and thus $\frac{|J_l|}{J} \overset{J\to\infty}{\longrightarrow} p_l$ according to condition (D3). Due to condition (D2) we can write

$$\eta = \sum_{j=1}^{J} c_{l_j} = \sum_{l=1}^{L} \sum_{j \in J_l} c_{l_j} = \sum_{l=1}^{L} |J_l| \cdot c_l\,,$$

and hence obtain by again condition (D3):

$$\frac{1}{\eta} = \frac{1}{|J_l|}\frac{|J_l|}{\eta} = \frac{1}{|J_l|}\frac{\frac{|J_l|}{J}}{\sum_{l=1}^{L}\frac{|J_l|}{J}\cdot c_l}\,,$$
$$\text{where} \quad \frac{\frac{|J_l|}{J}}{\sum_{l=1}^{L}\frac{|J_l|}{J}\cdot c_l} \overset{J\to\infty}{\longrightarrow} \frac{p_l}{\sum_{l=1}^{L} p_l \cdot c_l}\,.$$

Obviously, $J \to \infty$ if and only if $\eta \to \infty$ since $c_l < \infty$ for all $l \in \{1, \dots, L\}$. We can conclude:

$$\frac{1}{\eta}\sum_{j\in J_l} \mathbb{1}\left\{A_{j,i} \in \mathrm{Int}\right\} \overset{\eta\to\infty}{\longrightarrow} \underbrace{\frac{p_l}{\sum_{l=1}^{L} p_l \cdot c_l}\mathbb{P}_{\theta^*}\left(A_{j_0,i(\theta_\nu)} \in \mathrm{Int}(\theta_\nu)\right)}_{=:\tilde{\varepsilon}(\theta_\nu)}$$
$$\geq \varepsilon_1 > 0 \quad \mathbb{P}_{\theta^*}\text{-almost surely,}$$

where $\varepsilon_1 := \min_\nu \tilde{\varepsilon}(\theta_\nu)$. Until now we have considered an arbitrary $\theta \in \mathrm{B}_{2\delta_\nu}(\theta_\nu)$. We now want to move to a "reference parameter" θ^0_ν that lies on the edge of $\mathrm{B}_{\delta_\nu}(\theta_\nu)$ (recall that we use here the maximum norm). The local monotonicity condition (D6) and choice of norm yield the existence of some $\theta_\nu^0 \in \overline{\mathrm{B}_{\delta_\nu}(\theta_\nu)}$ such that

$$R^{\theta}_{j,i} \geq R^{\theta^0_\nu}_{j,i} \quad \text{on } \{A_{j,i} \in U_{l_j,i}\} \text{ for all } \theta \in \mathrm{B}_{\delta_\nu}(\theta_\nu).$$

For this, we can simply choose θ^0_ν as a vertex of the cube $\overline{\mathrm{B}_{\delta_\nu}(\theta_\nu)}$. Because of $\mathrm{Int} \subset U_{l,i}$, we then infer that

$$\mathcal{N}^{+}_{\mathrm{Int}}(\theta) \geq \mathcal{N}^{+}_{\mathrm{Int}}(\theta^0_\nu)\,.$$

Now let $\varepsilon < \min\{\varepsilon_0, \varepsilon_1\}$. There then exists $\Omega_0 = \Omega_0(\theta_\nu)$ with $\mathbb{P}_{\theta^*}(\Omega_0) = 1$ so that for any $\omega \in \Omega_0$ there exists η_ω such that for all $\eta \geq \eta_\omega$ simultaneously holds:

$$\sum_{j\in J_l} \mathbb{1}\left\{A_{j,i(\theta_\nu)}(\omega) \in \mathrm{Int}\right\} > \varepsilon\eta\,, \tag{4.24}$$

$$\frac{\mathcal{N}^{+}_{\mathrm{Int}}(\theta)(\omega)}{\sum_{j\in J_l} \mathbb{1}\left\{A_{j,i}(\omega)\in \mathrm{Int}\right\}} - \frac{1}{2} \geq \frac{\mathcal{N}^{+}_{\mathrm{Int}}(\theta^0_\nu)(\omega)}{\sum_{j\in J_l} \mathbb{1}\left\{A_{j,i}(\omega)\in \mathrm{Int}\right\}} - \frac{1}{2} > \varepsilon\,, \quad \text{for all } \theta \in \mathrm{B}_{\delta_\nu}(\theta_\nu).$$

As we order the transforms with respect to $\leq_{\mathrm{acc}}$, the $R^{\theta}_{j,i}$ with $A_{j,i} \in \mathrm{Int}$ appear in succession. Accordingly, from Equation (4.24) we can infer that at least $\lceil \varepsilon\eta \rceil$ consecutive observations satisfy $A_{j,i}(\omega) \in \mathrm{Int}$, and all these $R^{\theta}_{j,i}(\omega)$ are counted in $\mathcal{N}^{+}_{\mathrm{Int}}(\theta)(\omega)$ (since the indicator function returns 1 for all of them). This implies that, for all $\theta \in \mathrm{B}_{\delta_\nu}(\theta_\nu)$ and $\eta \geq \eta_\omega$,

$$\sup_{n-m+1\geq\varepsilon\eta} \frac{\mathcal{N}^{+}_{m,n}(\theta)(\omega)}{n-m+1} - \frac{1}{2} \geq \frac{\mathcal{N}^{+}_{\mathrm{Int}}(\theta)(\omega)}{\sum_{j\in J_l} \mathbb{1}\left\{A_{j,i}(\omega)\in \mathrm{Int}\right\}} - \frac{1}{2} > \varepsilon\,,$$

which means that $\omega \in \mathbb{A}_{\eta,\varepsilon}(\theta)$. Hence,

$$\omega \in \bigcap_{\theta\in \mathrm{B}_{\delta_\nu}(\theta_\nu)} \mathbb{A}_{\eta,\varepsilon}(\theta) \quad \forall \eta \geq \eta_\omega \quad \Longrightarrow \quad \omega \in \liminf_{\eta\to\infty} \bigcap_{\theta\in \mathrm{B}_{\delta_\nu}(\theta_\nu)} \mathbb{A}_{\eta,\varepsilon}(\theta)\,,$$

and thus we obtain:

$$1 = \mathbb{P}_{\theta*}(\Omega_0) \leq \mathbb{P}_{\theta^*}\left(\liminf_{\eta\to\infty} \bigcap_{\theta\in \mathrm{B}_{\delta_\nu}(\theta_\nu)} \mathbb{A}_{\eta,\varepsilon}(\theta)\right).$$

Lemma 4.18 (by taking complements in Equation (4.17)) then provides that

$$1 = \mathbb{P}_{\theta^*}\left(\liminf_{\eta\to\infty} \bigcap_{\nu=1}^{\nu_{\max}} \bigcap_{\theta\in \mathrm{B}_{\delta_\nu}(\theta_\nu)} \mathbb{A}_{\eta,\varepsilon}(\theta)\right) = \mathbb{P}_{\theta^*}\left(\liminf_{\eta\to\infty} \bigcap_{\theta\in\Theta_0} \mathbb{A}_{\eta,\varepsilon}(\theta)\right),$$

which implies the consistency of the 3-sign depth test by Corollary 4.10. □

Obviously, for Theorem 4.17 to hold, there must be at least one $c_l > 1$. Otherwise, no $i > 1$ can be found, so no observable past exists for any of the transforms. This is represented by the trivial random variables $A_{j,1} \equiv 0$. Moreover, it is also reflected in conditions (D5) and (D6), where only those l with $c_l > 1$ are considered. While our primary goal is to cover cases where processes are distinguishable only when their past is taken into account, this would exclude all one-point processes as an undesirable side effect. However, Equation (4.16) basically only ensures that for some l and $i > 1$ the conditional probability of any transform $R^{\theta}_{j,i} > 0$ with $j \in J_l$ given a feasible past $T^{(j)}_{1:(i-1)} = t^{(j)}_{1:(i-1)}$ differs from $\frac{1}{2}$. This condition can be easily extended to the case $i = 1$ by acknowledging the associated *un*conditional probability as well, that is, $\mathbb{P}_{\theta^*}(R^{\theta}_{j,1} > 0)$. If we allow

$$\mathbb{P}_{\theta^*}(R^{\theta}_{j,1} > 0) \neq \frac{1}{2}$$

to be satisfied as an alternative to Equation (4.16), then with slight adjustments to the conditions (D5) and (D6) the consistency of the 3-sign depth test can still be achieved. Generally speaking, this case is actually the easier one, since we do not have to deal with the intricacies of a random ordering, conditional monotonicity or dependence on the past.

Assumption 4.19 (Further Requirements Regarding the Transformation)**.**
The conditions (D5) and (D6) are supplemented as follows to include the case $i = 1$:

($\widetilde{\text{D}}$5) For each $j \in \mathbb{N}$, the probability $\mathbb{P}_{\theta^*}\big(R^\theta_{j,1} > 0\big)$ depends continuously on θ^* and θ.

($\widetilde{\text{D}}$6) For each $j \in \mathbb{N}$, the random variable $R^\theta_{j,1}$ is monotone in each component of θ.

Corollary 4.20 (Extension of Theorem 4.17)**.**
In the framework of Definition 4.11, suppose that the conditions (D1) *through* (D6) *of Assumptions 4.12 and 4.14 are fulfilled and that the conditions* ($\widetilde{\text{D}}$5) *and* ($\widetilde{\text{D}}$6) *of Assumptions 4.19 apply. Let $\Theta = \Theta_0 \cup \Theta_1$ with $\Theta_0 \cap \Theta_1 = \emptyset$, where Θ_0 is compact. Assume that for each $\theta^* \in \Theta_1$ and $\theta \in \Theta_0$, there exists $l = l(\theta, \theta^*) \in \{1, \ldots, L\}$ such that one of the following conditions holds:*

(i) There are $i \in \{2, \ldots, c_l\}$ and $x \in \operatorname{supp}(A_{j,i}) \cap U_{l,i}$ for j with $l_j = l$ satisfying

$$g_{l_j,i}(\theta^*, \theta, x) \neq \frac{1}{2}\,.$$

(ii) For all j with $l_j = l$, it holds:

$$\mathbb{P}_{\theta^*}\big(R^\theta_{j,1} > 0\big) \neq \frac{1}{2}\,.$$

Then, the 3-sign depth test with transforms $R^\theta_{j,i}$ ordered according to $\leq_{acc}$ is consistent for the hypotheses

$$\mathcal{H}_0 \,:\, \theta^* \in \Theta_0 \quad vs. \quad \mathcal{H}_1 \,:\, \theta^* \in \Theta_1\,.$$

Corollary 4.20 extends Theorem 4.17 to the extent that Equation (4.16) - here case (i) - does not necessarily have to be fulfilled, but case (ii) may occur alternatively.

Proof. We start as in the proof of Theorem 4.17 and will use the maximum norm here too. We fix any $\theta^* \in \Theta_1$. Then, by premise, for each $\theta \in \Theta_0$ there exists $l \in \{1, \ldots, L\}$ such that either case (i) or case (ii) holds. In case (i), we then construct an open neighbourhood $\mathrm{B}_\delta(\theta)$ as in the previous proof. In case (ii), we mirror this procedure: condition ($\widetilde{\text{D}}$5) ensures that for any given $\varepsilon = \varepsilon(\theta) < \big|\mathbb{P}_{\theta^*}\big(R^\theta_{j,1} > 0\big) - \frac{1}{2}\big|$ we find $\delta = \delta(\theta) > 0$ so that either

$$\mathbb{P}_{\theta^*}\left(R^{\tilde\theta}_{j,1} > 0\right) > \frac{1}{2} + \varepsilon \quad \text{for all } \tilde\theta \in \mathrm{B}_{2\delta}(\theta)\,, \tag{4.25}$$

or

$$\mathbb{P}_{\theta^*}\left(R^{\tilde\theta}_{j,1} > 0\right) < \frac{1}{2} - \varepsilon \quad \text{for all } \tilde\theta \in \mathrm{B}_{2\delta}(\theta)\,. \tag{4.26}$$

Consequently, we also obtain an open neighbourhood $\mathrm{B}_\delta(\theta)$ in this case. Altogether, these neighbourhoods (whether they come from case (i) or case (ii)) form an open cover of Θ_0, for which again a finite subcover can be found,

$$\Theta_0 \subset \underbrace{\bigcup_{\nu=1}^{\nu^{(i)}_{\max}} \mathrm{B}_{\delta_\nu}(\theta_\nu)}_{\text{case (i)}} \cup \underbrace{\bigcup_{\nu=\nu^{(i)}_{\max}+1}^{\nu^{(i)}_{\max}+\nu^{(ii)}_{\max}} \mathrm{B}_{\delta_\nu}(\theta_\nu)}_{\text{case (ii)}}\,,$$

where the first union is over $\nu_{\max}^{(i)}$ parameters belonging to case (i) and the second union is over $\nu_{\max}^{(ii)}$ parameters belonging to case (ii). Clearly, this partition is neither uniquely determined nor can $\nu_{\max}^{(i)} = 0$ or $\nu_{\max}^{(ii)} = 0$ be ruled out. If $\nu_{\max}^{(i)} > 0$, the proof of Theorem 4.17 immediately provides that, for some suitable $\varepsilon > 0$,

$$\mathbb{P}_{\theta^*}\left(\liminf_{\eta\to\infty} \bigcap_{\nu=1}^{\nu_{\max}^{(i)}} \bigcap_{\theta\in \mathrm{B}_{\delta_\nu}(\theta_\nu)} \mathbb{A}_{\eta,\varepsilon}(\theta)\right) = 1\,.$$

To complete the proof, we only need to show that an analogous identity also holds for the intersection over the parameters belonging to case (ii), insofar as $\nu_{\max}^{(ii)} > 0$.
For this, let $\varepsilon_0 := \min_\nu \varepsilon(\theta_\nu)$. We will again treat only the case from Equation (4.25) and proceed as before. Instead of $\mathcal{N}_{\mathrm{Int}}^+$, we define

$$\mathcal{N}^+(\theta) := \sum_{j\in J_l} \mathbb{1}\left\{R_{j,1}^\theta > 0\right\}\,, \quad \text{for } \theta \in \mathrm{B}_{2\delta_\nu}(\theta_\nu) \text{ and } \nu \in \{\nu_{\max}^{(i)}+1,\ldots,\nu_{\max}^{(ii)}\}\,,$$

which is already considerably easier than its counterpart from Equation (4.23). This simplifies the rest of the proof: For all $\theta \in \mathrm{B}_{2\delta_\nu}(\theta_\nu)$, the law of large numbers yields

$$\frac{1}{|J_l|}\mathcal{N}^+(\theta) - \frac{1}{2} \xrightarrow{J\to\infty} \mathbb{P}_{\theta^*}\left(R_{j,1}^\theta > 0\right) - \frac{1}{2} > \varepsilon \quad \mathbb{P}_{\theta^*}\text{-almost surely.}$$

Furthermore, we obtain deterministically that

$$\frac{|J_l|}{\eta} \xrightarrow{J\to\infty} \frac{p_l}{\sum_{l=1}^{L} p_l \cdot c_l} =: \varepsilon_1\,.$$

The monotonicity condition ($\widetilde{\mathrm{D}}6$) allows us to determine a reference parameter $\theta_\nu^0 \in \overline{\mathrm{B}_{\delta_\nu}(\theta_\nu)}$ such that

$$\mathcal{N}^+(\theta) \geq \mathcal{N}^+(\theta_\nu^0) \quad \text{for all } \theta \in \mathrm{B}_{\delta_\nu}(\theta_\nu).$$

For any $\varepsilon < \min\{\varepsilon_0, \varepsilon_1\}$, there then exists $\Omega_0 = \Omega_0(\theta_\nu)$ with $\mathbb{P}_{\theta^*}(\Omega_0) = 1$ and the property that for any $\omega \in \Omega_0$ there exists η_ω such that for all $\eta \geq \eta_\omega$ we have

$$|J_l| > \varepsilon\eta \quad \text{and} \quad \frac{1}{|J_l|}\mathcal{N}^+(\theta) - \frac{1}{2} > \varepsilon\,, \quad \text{for all } \theta \in \mathrm{B}_{\delta_\nu}(\theta_\nu).$$

With respect to $\leq_{\mathrm{acc}}$, the transforms with $i = 1$ are ordered by their class l. Hence, the $R_{j,1}^\theta$ with $j \in J_l$ appear in succession and therefore

$$\sup_{n-m+1\geq\varepsilon\eta} \frac{\mathcal{N}_{m,n}^+(\theta)(\omega)}{n-m+1} - \frac{1}{2} \geq \frac{\mathcal{N}^+(\theta)(\omega)}{|J_l|} - \frac{1}{2} > \varepsilon\,, \text{ for all } \theta \in \mathrm{B}_{\delta_\nu}(\theta_\nu),\, \omega \in \Omega_0 \text{ and } \eta \geq \eta_\omega.$$

From here, the rest of the proof then is completely analogous to that of Theorem 4.17. □

Even though the conditions on the transformations $f_{l,i}$ have been formulated more generally, the framework of Assumptions 4.14 is tailor-made for the application of the hazard transformation. We can use Corollary 4.20 to derive the consistency of the 3-sign depth test based on hazard transforms for many load sharing models. As an example, we demonstrate this for the equal load sharing model of Kvam and Peña 2005, see Equation

(2.14). We only need to abandon type I censoring in order to conform to the framework for hazard transforms in load sharing models of Definition 2.23.

Example 4.21 (Consistency of the 3-Sign Depth Test in the Equal Load Sharing Model of Kvam and Peña 2005).
We consider the load sharing model given by the parametric conditional hazard functions

$$h_{l,i}^{\theta}\Big(t \,|\, t_{1:(i-1)}^{(j)}\Big) = \begin{cases} \alpha(t)\theta_{i-1}\left(I-(i-1)\right), & \text{if } i \leq c_l, \\ 0, & \text{otherwise,} \end{cases}$$

where $c_l \in \{1, \ldots, I\}$ determines the number of critical component failures, $l \in \{1, \ldots, L\}$ and $i \in \{1, \ldots, I\}$. In comparison to the model from Equation (2.14), we have dispensed with type I censoring, but in exchange the number of critical component failures is allowed to depend on the particular experiment. We may also admit an experiment-specific baseline hazard α_j, but we refrain from doing so only for the sake of readability.
The basic Assumptions 4.12 are easily satisfied: (D1) is the basic modelling assumption that is always accepted to be true. (D2) is satisfied if $\alpha > 0$ almost everywhere (for the Basquin load sharing models, this is proved in Lemma 4.22; the proof is easily transferred). (D3) can be accomplished by an appropriate experimental design: If a finite assortment of experimental conditions $d = (d_1, \ldots, d_\nu) \in \{1, \ldots, L\}^\nu$, $\nu \in \mathbb{N}$, is repeated consecutively as the total number of systems tends to infinity, then

$$p_l = \frac{1}{\nu}\sum_{j=1}^{\nu} \mathbb{1}_{\{l\}}(d_j), \quad l \in \{1, \ldots, L\},$$

which means that the asymptotic proportion of class l is equal to the relative proportion of class l in the design d. Consequently, no class occurring in d is asymptotically negligible.
Let us assume that $\alpha > 0$ almost everywhere. In practical terms, this prohibits time periods during which no component failures can occur (e.g., a pause in the experiment). If we consider the standardized[19] hazard transform process of Definition 2.29, that is,

$$R_{j,i}^{\theta} := H_{l_j,i}^{\theta}\Big(T_i^{(j)} \,|\, T_{1:(i-1)}^{(j)}\Big) - \ln(2),$$

then (D4) is satisfied according to Theorem 2.20 (cf. Equation (2.60)). For the remaining conditions (D5) and (D6), let

$$G(t) := \int_0^t \alpha(u)\,\mathrm{d}u, \quad t \in [0, \infty).$$

Then, for $j \in \mathbb{N}$ and $i \leq c_{l_j}$, the cumulative conditional hazard function on $[T_{i-1}^{(j)}, \infty)$ is

$$\begin{aligned} H_{l_j,i}^{\theta}\Big(t \,|\, T_{1:(i-1)}^{(j)}\Big) &= \int_{T_{i-1}^{(j)}}^{t} \alpha(u)\theta_{i-1}(I-(i-1))\,\mathrm{d}u \\ &= \theta_{i-1}(I-(i-1))\Big[G(t) - G\Big(T_{i-1}^{(j)}\Big)\Big], \end{aligned}$$

[19] In Definition 2.29, the notation $\tilde{R}_{j,i}^{\theta}$ is used to distinguish the standardized transforms from the ordinary transforms. Since we only use the standardized hazard transforms here, we drop the tilde.

with an inverse on $[0,\infty)$ given via

$$\left(H^{\theta}_{l_j,i}\right)^{-1}\left(u\,|\,T^{(j)}_{1:(i-1)}\right) = G^{-1}\left(\frac{u}{\theta_{i-1}(I-(i-1))} + G\left(T^{(j)}_{i-1}\right)\right).$$

Therefore,

$$\begin{aligned}
&H^{\theta^*}_{l_j,i}\left(\left(H^{\theta}_{l_j,i}\right)^{-1}\left(u\,|\,T^{(j)}_{1:(i-1)}\right)\Big|\,T^{(j)}_{1:(i-1)}\right)\\
&\quad= \theta^*_{i-1}(I-(i-1))\left[G\left(G^{-1}\left(\frac{u}{\theta_{i-1}(I-(i-1))} + G\left(T^{(j)}_{i-1}\right)\right)\right) - G\Big(T^{(j)}_{i-1}\Big)\right]\\
&\quad= \theta^*_{i-1}(I-(i-1))\left[\frac{u}{\theta_{i-1}(I-(i-1))}\right]\\
&\quad= \frac{\theta^*_{i-1}}{\theta_{i-1}}u\,,
\end{aligned}$$

so from Equation (2.49) we get:

$$\mathbb{P}_{\theta^*}\left(R^{\theta}_{j,i} > 0\,\Big|\,T^{(j)}_{1:(i-1)}\right) = \exp\left(-\frac{\theta^*_{i-1}}{\theta_{i-1}}\ln(2)\right) = \left(\frac{1}{2}\right)^{\frac{\theta^*_{i-1}}{\theta_{i-1}}} = g_{l_j,i}\left(\theta^*,\theta,A_{j,i}\right)\,,$$

where for $l\in\{1,\dots,L\}$ and $i\in\{1,\dots,c_l\}$ the function $g_{l,i}\,:\,\Theta^2\times\mathbb{R}\to\mathbb{R}$ is defined as

$$g_{l,i}\left(\theta^*,\theta,x\right) = \left(\frac{1}{2}\right)^{\frac{\theta^*_{i-1}}{\theta_{i-1}}}\,. \tag{4.27}$$

Obviously, $g_{l,i}$ is continuous. Since this function does not depend on x, we can choose $a_{l,i}\equiv 0$ for all $l\in\{1,\dots,L\}$ and $i\in\{1,\dots,c_l\}$ to satisfy condition (D5). As a consequence, $A_{j,i}\sim\delta_0$ for all $j\in\mathbb{N}$ and $i\in\{1,\dots,c_{l_j}\}$. Finally,

$$R^{\theta}_{j,i} = \theta_{i-1}(I-(i-1))\left[G\Big(T^{(j)}_{i}\Big) - G\Big(T^{(j)}_{i-1}\Big)\right] - \ln(2)\,,$$

which is strictly increasing in θ_{i-1} and constant in all other parameters with probability one. Accordingly, (D6) is fulfilled for arbitrary intervals $U_{l,i}\ni 0$.
For the consistency of the 3-sign depth test, let $\theta^*\in\Theta_1$ and $\theta\in\Theta_0$. Since $\theta^*\neq\theta$, there is some $i_0\in\{1,\dots,I\}$ with $\theta^*_{i_0-1}\neq\theta_{i_0-1}$. By Corollary 4.20, we only have to show that

$$g_{l,i}\left(\theta^*,\theta,0\right)\neq\frac{1}{2}\quad\text{for some } l\in\{1,\dots,L\}\text{ and } i\in\{1,\dots,c_l\}.$$

By Equation (4.27), this is equivalent to the condition that

$$\theta^*_{i-1}\neq\theta_{i-1}\quad\text{for some } l\in\{1,\dots,L\}\text{ and } i\in\{1,\dots,c_l\},$$

which is satisfied if there exists $l\in\{1,\dots,L\}$ with $c_l\geq i_0$. From this we can infer that the 3-sign depth test is consistent if and only if for each $\theta^*\in\Theta_1$ and $\theta\in\Theta_0$ there exists $i_0\leq\max_{l\in\{1,\dots,L\}}c_l$ such that $\theta^*_{i_0-1}\neq\theta_{i_0-1}$. This can be understood as a minimum requirement, since otherwise the hazard functions at θ^* and θ would agree for all $j\in\mathbb{N}$ and $i\in\{1,\dots,c_{l_j}\}$. Heuristically, this means that the parameters cannot be distinguished by the model itself, which precludes any test from being consistent.

In the above example, we saw that the condition from Corollary 4.20 as well as the technical requirements (D5) and (D6) were easily verified. This is partly because the model does not take into account the past of the process, which is reflected in the trivial essential past $A_{j,i} \equiv 0$. The complexity of Corollary 4.20 (but also its potency) becomes apparent only when we consider models with damage accumulation. We do so in the following section, where we apply our results to the model $^{\times}$D.

4.3. Application to the Basquin Load Sharing Model With Multiplicative Damage Accumulation

In analogy to the previous chapter on minimum distance estimation, we round out the theoretical framework on the K-sign depth by examining its application to the model $^{\times}$D. For this specific model, the consistency results of Section 4.2 can be further extended. This upcoming corollary allows us to test for the presence of damage accumulation (i.e., to consider the null hypothesis $\mathcal{H}_0 : \theta_3^* = 0$, which here corresponds to the non-compact set $\Theta_0 = (0, \infty) \times \mathbb{R} \times \{0\}$). However, the flexibility of these hypotheses comes at the price of additional constraints on the model: We assume that the initial stress level s_j of the jth experiment is chosen deterministically and that the observation horizon is unbounded. For the model, this means that the interval $\mathcal{I} = [0, \infty)$ is no longer compact and we formally require $\tau_0 \sim \delta_\infty$. What seems to be a limitation at first sight, in practice allows a simplified implementation, since experiments are no longer subject to a rigorous i.i.d. assumption and can instead be carried out under different initial conditions that have been determined in advance. Moreover, these experiments may also differ substantially in terms of their non-parametric parts. We can also mitigate some requirements on the parameter range Θ, which is only assumed to satisfy $\pi_1(\Theta) \subset (0, \infty)$ and $\pi_3(\Theta) \subset (-1, \theta)$ (i.e., $\theta_1 > 0$ and $\theta_3 > -1$ for all $\theta \in \Theta$). In particular, we do not impose any restrictions on the parameter θ_2. The remainder of this section is divided into two parts: In the first subsection, we will review the conditions (D1) through (D6) for the Basquin load sharing model with multiplicative damage accumulation and apply Theorem 4.17. In the second subsection, we prove the consistency of the test for damage accumulation. We postpone further discussion of possible extensions to the outlook.

4.3.1. Application of Theorem 4.17 to the Basquin Load Sharing Model With Multiplicative Damage Accumulation

Example 4.21 provides an outline of how the conditions (D1) through (D6) can be checked for the model $^{\times}$D. As we may not consider random covariates in the sense of Assumptions 2.3, we replace them with up to $L \in \mathbb{N}$ deterministic experimental conditions. The models $\mathcal{M}_l$, $l \in \{1, \dots, L\}$, are then allowed to differ in terms of the following three experimental conditions:

(i) $c_l \in \{1, \dots, I\}$, the number of observable failures,

(ii) $s_l > 0$, the initial stress level,

(iii) $\tau_l > 0$, which serves as a normalizing constant here.

In other words, for the jth experiment, we assume that the intensity is of the shape

$$
{}^{\times\mathrm{D}}\lambda_\theta^{(j)}(t) = \theta_1 \left(s_{l_j} \frac{I}{I - N_{t^-}^{(j)}} \right)^{\theta_2} \left(\frac{1}{\tau_{l_j}} \int_0^t s_{l_j} \frac{I}{I - N_{u^-}^{(j)}} \,\mathrm{d}u \right)^{\theta_3} \cdot \mathbb{1}_{\left\{ N_{t^-}^{(j)} < c_{l_j} \right\}}, \tag{4.28}
$$

where $l_j \in \{1, \ldots, L\}$. Compared to the formula from Definition 2.8, we removed the restriction $\{t \leq \tau_{l_j}\}$. The reason for this is that type I censoring undesirably distorts the distribution of the transformed point processes, which is addressed in Remark 2.22 for the special case of the hazard transformation. It is precisely this transformation that we will invoke later. We must therefore adhere to the framework from Definition 2.23, which necessitates the requirement of deterministic covariates.

Verifying Conditions (D1) Through (D3)

We start by verifying the conditions (D1), (D2) and (D3) from Assumption 4.12. The first condition (D1) can be understood as a plausibility assumption. If it is not satisfied, all efforts of statistical inference are futile. It can never be verified in practice, whereas it always holds for simulation studies, since the realized processes are always drawn from the model in question. On the other hand, the fulfillment of conditions (D2) and even more so (D3) is part of the experimental design. Specifically, the variable c_l plays the same role in the experimental condition (i) as in condition (D2). We record this coherence in a first lemma.

Lemma 4.22 (Component Failures in the Basquin Load Sharing Model)**.**
If the counting process $N^{(j)}$ admits the intensity ${}^{\times D}\lambda_\theta^{(j)}$ from Equation (4.28)*, then $N_t^{(j)} \to c_{l_j}$ as $t \to \infty$ with probability one.*

Proof. Choose any sequence $(t_n)_{n\in\mathbb{N}}$ with $t_n \to \infty$ as $n \to \infty$. Since the intensity ${}^{\times\mathrm{D}}\lambda_\theta^{(j)}$ vanishes on $\{N_{t^-}^{(j)} \geq c_{l_j}\}$, it immediately follows that $N_t^{(j)} \leq c_{l_j}$ for all $t \geq 0$ with probability one. Hence, as $N_t^{(j)}$ is $\mathbb{P}_\theta$-almost surely non-decreasing in t, it holds that

$$
\mathbb{P}_\theta \left(\lim_{n\to\infty} N_{t_n}^{(j)} = c_{l_j} \right) = \mathbb{P}_\theta \left(\lim_{n\to\infty} N_{t_n}^{(j)} \geq c_{l_j} \right) = \lim_{n\to\infty} \mathbb{P}_\theta \left(N_{t_n}^{(j)} \geq c_{l_j} \right), \tag{4.29}
$$

by the σ-continuity of the probability measure. Equation (A.4) yields that

$$
\mathbb{P}_\theta \left(\lim_{n\to\infty} N_{t_n}^{(j)} = c_{l_j} \right) = \lim_{n\to\infty} \mathbb{P}_\theta \left(T_{c_{l_j}}^{(j)} \leq t_n \right),
$$

or, equivalently, by taking complements:

$$
\mathbb{P}_\theta \left(\lim_{n\to\infty} N_{t_n}^{(j)} < c_{l_j} \right) = \mathbb{P}_\theta \left(\lim_{n\to\infty} N_{t_n}^{(j)} \neq c_{l_j} \right) = \lim_{n\to\infty} \mathbb{P}_\theta \left(T_{c_{l_j}}^{(j)} > t_n \right). \tag{4.30}
$$

The absolute continuity of the compensator of $N^{(j)}$, trivially given by the fact that it can be expressed as a cumulative intensity, implies that $T_{c_{l_j}}^{(j)}$ follows a continuous distribution. This is illustrated in the proof of Lemma A.37 in Appendix A.4, where the Lebesgue densities are successively reconstructed from the compensator by virtue of Proposition A.35. The tower property of the conditional expectation then provides:

$$
\mathbb{P}_\theta \left(T_{c_{l_j}}^{(j)} > t_n \right) = \mathbb{E}_\theta \left(\mathbb{1}_{\left\{ T_{c_{l_j}}^{(j)} > t_n \right\}} \right) = \mathbb{E}_\theta \left[\mathbb{E}_\theta \left(\mathbb{1}_{\left\{ T_{c_{l_j}}^{(j)} > t_n \right\}} \,\middle|\, T_{1:(c_{l_j}-1)}^{(j)} \right) \right]
$$

$$
\begin{aligned}
&= \mathbb{E}_\theta\Big[\mathbb{P}_\theta\Big(T^{(j)}_{c_{l_j}} > t_n \,\Big|\, T^{(j)}_{1:(c_{l_j}-1)}\Big)\Big] \\
&= \mathbb{E}_\theta\Big[S^\theta_{l_j,c_{l_j}}\Big(t_n \,\Big|\, T^{(j)}_{1:(c_{l_j}-1)}\Big)\Big] \\
&= \mathbb{E}_\theta\Big[\exp\Big(-H^\theta_{l_j,c_{l_j}}\Big(t_n \,\Big|\, T^{(j)}_{1:(c_{l_j}-1)}\Big)\Big)\Big] \\
&= \mathbb{E}_\theta\Big[\exp\Big(-\int_{T^{(j)}_{c_{l_j}-1}}^{t_n} h^\theta_{l_j,c_{l_j}}\Big(u \,\Big|\, T^{(j)}_{1:(c_{l_j}-1)}\Big)\,\mathrm{d}u\Big)\Big]\,. \qquad (4.31)
\end{aligned}
$$

Substituting the conditional hazard function of the model $^\times$D from Lemma 2.24, we get:

$$
\begin{aligned}
&\mathbb{E}_\theta\Big[\exp\Big(-\int_{T^{(j)}_{c_{l_j}-1}}^{t_n} h^\theta_{l_j,c_{l_j}}\Big(u \,\Big|\, T^{(j)}_{1:(c_{l_j}-1)}\Big)\,\mathrm{d}u\Big)\Big] \\
&\quad = \mathbb{E}_\theta\Big[\exp\Big(-\int_{T^{(j)}_{c_{l_j}-1}}^{t_n} \theta_1 B^{\theta_2}_{j,c_{l_j}}\Big[\frac{1}{\tau_{l_j}}\Big(B_{j,c_{l_j}}\Big(u - T^{(j)}_{c_{l_j}-1}\Big) + A_{j,c_{l_j}}\Big)\Big]^{\theta_3}\,\mathrm{d}u\Big)\Big]\,.
\end{aligned}
$$

On $\Big\{0 < T^{(j)}_{c_{l_j}-1} < \infty\Big\}$ (note that $T^{(j)}_{c_{l_j}-1} > 0$ with probability one by Definition A.3),

$$
\theta_1 B^{\theta_2}_{j,c_{l_j}}\Big[\frac{1}{\tau_{l_j}}\Big(B_{j,c_{l_j}}\Big(u - T^{(j)}_{c_{l_j}-1}\Big) + A_{j,c_{l_j}}\Big)\Big]^{\theta_3} \geq \theta_1 B^{\theta_2}_{j,c_{l_j}}\left(\frac{A_{j,c_{l_j}}}{\tau_{l_j}}\right)^{\theta_3} > 0\,, \quad \text{for } u > T^{(j)}_{c_{l_j}-1}.
$$

Hence, for each $\omega \in \Big\{0 < T^{(j)}_{c_{l_j}-1} < \infty\Big\}$, we have:

$$
\underbrace{\exp\Big(-\int_{T^{(j)}_{c_{l_j}-1}(\omega)}^{t_n} \theta_1 B^{\theta_2}_{j,c_{l_j}}\Big[\frac{1}{\tau_{l_j}}\Big(B_{j,c_{l_j}}\Big(u - T^{(j)}_{c_{l_j}-1}(\omega)\Big) + A_{j,c_{l_j}}(\omega)\Big)\Big]^{\theta_3}\,\mathrm{d}u\Big)}_{\leq 1} \longrightarrow 0 \quad (n \to \infty)\,,
$$

and the dominated convergence theorem yields that

$$
\mathbb{E}_\theta\Big[\exp\Big(-\int_{T^{(j)}_{c_{l_j}-1}}^{t_n} h^\theta_{l_j,c_{l_j}}\Big(u \,\Big|\, T^{(j)}_{1:(c_{l_j}-1)}\Big)\,\mathrm{d}u\Big)\cdot \mathbb{1}_{\big\{0<T^{(j)}_{c_{l_j}-1}<\infty\big\}}\Big] \longrightarrow 0 \quad (n \to \infty)\,. \quad (4.32)
$$

However, we have not yet accounted for the case $\Big\{T^{(j)}_{c_{l_j}-1} = \infty\Big\}$, and by splitting Equation (4.31) we obtain:

$$
\begin{aligned}
&\mathbb{E}_\theta\Big[\exp\Big(-\int_{T^{(j)}_{c_{l_j}-1}}^{t_n} h^\theta_{l_j,c_{l_j}}\Big(u \,\Big|\, T^{(j)}_{1:(c_{l_j}-1)}\Big)\,\mathrm{d}u\Big)\Big] \\
&\quad = \underbrace{\mathbb{E}_\theta\Big[\exp\Big(-\int_{T^{(j)}_{c_{l_j}-1}}^{t_n} h^\theta_{l_j,c_{l_j}}\Big(u \,\Big|\, T^{(j)}_{1:(c_{l_j}-1)}\Big)\,\mathrm{d}u\Big)\cdot \mathbb{1}_{\big\{0<T^{(j)}_{c_{l_j}-1}<\infty\big\}}\Big]}_{=\,o(1)\text{ by Equation (4.32)}}
\end{aligned}
$$

$$+\,\mathbb{E}_\theta\Bigg[\underbrace{\exp\Bigg(-\int_{T^{(j)}_{c_{l_j}-1}}^{t_n} h^\theta_{l_j,c_{l_j}}\Big(u\,\Big|\,T^{(j)}_{1:(c_{l_j}-1)}\Big)\,\mathrm{d}u\Bigg)}_{\le 1}\cdot\mathbb{1}_{\{T^{(j)}_{c_{l_j}-1}=\infty\}}\Bigg]$$

$$\le o(1)+\mathbb{E}_\theta\Big[\mathbb{1}_{\{T^{(j)}_{c_{l_j}-1}=\infty\}}\Big]$$

$$\le \mathbb{P}_\theta\Big(T^{(j)}_{c_{l_j}-1}>t_n\Big)+o(1)\,.$$

Altogether, we can infer that

$$\mathbb{P}_\theta\left(T^{(j)}_{c_{l_j}}>t_n\right)\le\mathbb{P}_\theta\Big(T^{(j)}_{c_{l_j}-1}>t_n\Big)+o(1)\,.$$

If we successively iterate this exact procedure, we therefore arrive at

$$\mathbb{P}_\theta\left(T^{(j)}_{c_{l_j}}>t_n\right)\le\mathbb{P}_\theta\Big(T^{(j)}_{c_{l_j}-1}>t_n\Big)+o(1)\le\ldots\le\mathbb{P}_\theta\Big(T^{(j)}_1>t_n\Big)+o(1)\,. \tag{4.33}$$

This estimate has the advantage that $T^{(j)}_1$ as the first component failure does not depend on the past of the process. More precisely, by Lemma 2.24, we have

$$\mathbb{P}_\theta\Big(T^{(j)}_1>t_n\Big)=\mathbb{E}_\theta\left[\exp\left(-H^\theta_{l_j,1}(t_n)\right)\right]=\mathbb{E}_\theta\left[\exp\left(-\frac{\theta_1 s_{l_j}^{\theta_2+\theta_3}}{\tau_{l_j}^{\theta_3}(\theta_3+1)}t_n^{\theta_3+1}\right)\right]\,,$$

since $B_{j,1}=s_j$ and $A_{j,1}=0$. Unlike before, this is the expected value of a constant, because s_{l_j} and τ_{l_j} are assumed to be deterministic. Hence,

$$\mathbb{P}_\theta\Big(T^{(j)}_1>t_n\Big)=\exp\left(-\frac{\theta_1 s_{l_j}^{\theta_2+\theta_3}}{\tau_{l_j}^{\theta_3}(\theta_3+1)}t_n^{\theta_3+1}\right)\longrightarrow 0\quad(n\to\infty)\,.$$

Overall, from Equation (4.33) we can conclude that

$$\mathbb{P}_\theta\left(T^{(j)}_{c_{l_j}}>t_n\right)\le\mathbb{P}_\theta\Big(T^{(j)}_1>t_n\Big)+o(1)\longrightarrow 0\quad(n\to\infty)\,,$$

and thus the assertion follows from Equation (4.30). □

From Lemma 4.22, the condition (D2) is evident. Note that Equation (4.33) is particularly intriguing, since the monotonicity of a point process guarantees that

$$\mathbb{P}_\theta\left(T^{(j)}_{c_{l_j}}>t_n\right)\ge\mathbb{P}_\theta\Big(T^{(j)}_{c_{l_j}-1}>t_n\Big)\ge\ldots\ge\mathbb{P}_\theta\Big(T^{(j)}_1>t_n\Big)\,.$$

In the given situation, we thus have for each $i\in\{1,\ldots,c_{l_j}-1\}$:

$$\mathbb{P}_\theta\left(T^{(j)}_i>t_n\right)\le\mathbb{P}_\theta\left(T^{(j)}_{i+1}>t_n\right)\le\mathbb{P}_\theta\left(T^{(j)}_i>t_n\right)+o(1)\,,$$

so that these probabilities coincide asymptotically. We now turn to condition (D3): Whether condition (D3) is satisfied is a matter of experimental design. Let us assume that in order to increase the number of trials, a finite design d with

$$d = (d_1, \dots, d_\nu) \in \{1, \dots, L\}^\nu , \quad \nu \in \mathbb{N} ,$$

is repeated successively. For the jth experiment, this means that

$$l_j := d_{j \mod \nu} , \quad j \in \mathbb{N} ,$$

with the convention that $d_0 := d_\nu$, see Table 1 for an illustration. Similar to Example 4.21, we then have

$$p_l = \lim_{J\to\infty} \frac{1}{J} \sum_{j=1}^{J} \mathbb{1}_{\{l\}}(l_j) = \frac{1}{\nu} \sum_{j=1}^{\nu} \mathbb{1}_{\{l\}}(d_j) .$$

Therefore, for each $l \in \{1, \dots, L\}$,

$$p_l > 0 \quad \Longleftrightarrow \quad \sum_{j=1}^{\nu} \mathbb{1}_{\{l\}}(d_j) > 0 ,$$

which is the case if the class l appears at least once in d. In other terms, any class present in the design is asymptotically non-negligible. Consequently, if every class is present in the design, then condition (D3) is met. However, without loss of generality, this can always be achieved by discarding the classes that do not occur in d.

j	1	2	$\cdots$	$\nu-1$	ν	$\nu+1$	$\cdots$	$2\nu-1$	2ν	$2\nu+1$	$\cdots$
l_j	d_1	d_2	$\cdots$	$d_{\nu-1}$	$d_0 = d_\nu$	d_1	$\cdots$	$d_{\nu-1}$	$d_0 = d_\nu$	d_1	$\cdots$

Table 1: How to determine the class l_j of the jth experiment when repeating the design d.

Verifying Conditions (D4) Through (D6)

To prove conditions (D4) to (D6) from Assumption 4.14, the transformations $f_{l,i}$ must be suitably chosen. Analogous to Example 4.21, we consider the standardized hazard transform of Definition 2.29, that is,

$$R^\theta_{j,i} := {}^{\times\mathrm{D}}H^\theta_{l_j,i}\Big(T^{(j)}_i \,\big|\, T^{(j)}_{1:(i-1)}\Big) - \ln(2) , \quad \theta \in \Theta,\, j \in \{1, \dots, J\},\, i \in \{1, \dots, c_{l_j}\} .$$

Again, the condition (D4) is then due to Theorem 2.20 (see also Equation (2.60)). Similarly, condition (D5) has already been shown in Remark 2.30. For each $l \in \{1, \dots, L\}$ with $c_l > 1$, $i \in \{2, \dots, c_l\}$ and $\theta, \theta^* \in \Theta$, we have seen there that

$$\mathbb{P}_{\theta^*}\left(R^\theta_{j,i} > 0 \,\middle|\, T^{(j)}_{1:(i-1)}\right) = g_{l_j,i}\left(\theta^*, \theta, A_{j,i}\right) ,$$

where $g_{l,i} : \Theta \times \Theta \times \mathbb{R} \to \mathbb{R}$ is the deterministic function defined by

$$g_{l,i}(\theta, \theta^*, x) := \exp\left(- \frac{\theta^*_1 B_{l,i}^{\theta^*_2 - 1}}{\tau_l^{\theta^*_3}\left(\theta^*_3 + 1\right)} \left[\left(\frac{\tau_l^{\theta_3}\left(\theta_3 + 1\right)}{\theta_1 B_{l,i}^{\theta_2 - 1}} \ln(2) + x^{\theta_3 + 1} \right)^{\frac{\theta^*_3 + 1}{\theta_3 + 1}} - x^{\theta^*_3 + 1} \right] \right) . \tag{4.34}$$

We remark that here we have

$$B_{l,i} = s_l \frac{I}{I-(i-1)}\,,$$

since the load sharing term $B_{j,i}$ depends on the jth experiment only by its class l_j. If $\Theta \subset (0,\infty)\times\mathbb{R}\times(-1,\infty)$, then $g_{l,i}$ is continuous on $\Theta^2\times\mathbb{R}$. Moreover, from Equation (2.23) we know that

$$A_{j,i} = \sum_{k=1}^{i-1} B_{l_j,k}\left(T_k^{(j)} - T_{k-1}^{(j)}\right) =: a_{l_j,i}\left(T_1^{(j)},\dots,T_{i-1}^{(j)}\right).$$

Recall that $T_0^{(j)} \equiv 0$, which is why $T_0^{(j)}$ does not appear in the argument of $a_{l_j,i}$. There remains only condition (D6). Corollary 2.26 provides that

$$R_{j,i}^{\theta} = \frac{\theta_1 B_{l_j,i}^{\theta_2-1}}{\tau_{l_j}^{\theta_3}(\theta_3+1)} \underbrace{\left[A_{j,i+1}^{\theta_3+1} - A_{j,i}^{\theta_3+1}\right]}_{\geq 0} - \ln(2)\,. \tag{4.35}$$

For each $\omega\in\Omega$, $R_{j,i}^{\theta}(\omega)$ is obviously monotonically increasing in θ_1 and θ_2, so we can choose $\rho_1=\rho_2=0$ for condition (D6). The monotonicity in θ_3 is more difficult to derive from Equation (4.35) because it may depend on the respective value of the two damage accumulation terms $A_{j,i}$ and $A_{j,i+1}$ and thus on $\omega\in\Omega$. This is the reason why we allow a restriction to $\{A_{j,i}\in U_{l_j,i}\}\subset\Omega$ for some open interval $U_{l_j,i}$ in condition (D6) to admit only selected ω. The key insight here is that the monotonicity of $R_{j,i}^{\theta}$ is essentially inherited from the conditional hazard function ${}^{\times\mathrm{D}}h_{l_j,i}^{\theta}$. For this, we observe:

$$\begin{aligned}
R_{j,i}^{\theta} &= {}^{\times\mathrm{D}}H_{l_j,i}^{\theta}\Big(T_i^{(j)} \,\big|\, T_{1:(i-1)}^{(j)}\Big) - \ln(2)\\
&= \int_{T_{i-1}^{(j)}}^{T_i^{(j)}} {}^{\times\mathrm{D}}h_{l_j,i}^{\theta}\Big(u \,\big|\, T_{1:(i-1)}^{(j)}\Big)\,\mathrm{d}u - \ln(2)\\
&= \int_{T_{i-1}^{(j)}}^{T_i^{(j)}} \theta_1 B_{l_j,i}^{\theta_2} A_j(u)^{\theta_3}\,\mathrm{d}u - \ln(2)\,.
\end{aligned}$$

This representation of the integrand is only valid because the integration domain ensures that $u\in[T_{i-1}^{(j)},T_i^{(j)}]$. If for any given $\omega\in\Omega$ we have

$$\theta_3 \leq \tilde{\theta}_3 \quad\Longrightarrow\quad \underbrace{A_j(u)^{\theta_3} \leq A_j(u)^{\tilde{\theta}_3}}_{\text{depends on }\omega} \quad \text{for all } u\in\left[T_{i-1}^{(j)}(\omega),T_i^{(j)}(\omega)\right], \tag{4.36}$$

then the monotonicity of the integral yields (we omit the arguments ω in the integral bounds for brevity)

$$\theta_3 \leq \tilde{\theta}_3 \quad\Longrightarrow\quad \int_{T_{i-1}^{(j)}}^{T_i^{(j)}} \theta_1 B_{l_j,i}^{\theta_2} A_j(u)^{\theta_3}\,\mathrm{d}u - \ln(2) \leq \int_{T_{i-1}^{(j)}}^{T_i^{(j)}} \theta_1 B_{l_j,i}^{\theta_2} A_j(u)^{\tilde{\theta}_3}\,\mathrm{d}u - \ln(2)\,.$$

This means that $R^{\theta}_{j,i}$ is increasing in θ_3 for each ω such that Equation (4.36) applies. For $x > 1$, the function

$$\mathbb{R} \longrightarrow \mathbb{R} \,:\, \theta_3 \longmapsto x^{\theta_3} \,,$$

is monotonically increasing, which is easily seen by differentiating with respect to θ_3. Equation (4.36) therefore holds if $A_j(u) > 1$ can be ensured for all $u \in [T^{(j)}_{i-1}(\omega), T^{(j)}_{i}(\omega)]$. But for such u,

$$A_j(u) = B_{l_j}\left(u - T^{(j)}_{i-1}(\omega)\right) + A_{j,i}(\omega) \geq A_{j,i}(\omega) \,.$$

Accordingly, $A_j(u) > 1$ follows if $A_{j,i}(\omega) > 1$. We can therefore let $U_{l_j,i} \subset (1,\infty)$, since for any $\omega \in \{A_{j,i} \in U_{l_j,i}\}$ we then obtain for all $u \in [T^{(j)}_{i-1}(\omega), T^{(j)}_{i}(\omega)]$:

$$A_{j,i}(\omega) \in U_{l_j,i} \subset (1,\infty) \quad \Longrightarrow \quad A_{j,i}(\omega) > 1 \quad \Longrightarrow \quad A_j(u) > 1 \,.$$

Altogether, condition (D6) is thus satisfied for $\rho_3 = 0$ (hence, $\rho = (0,0,0)$) and arbitrary $U_{l,i} \subset (1,\infty)$ as long as there exists $l \in \{1,\dots,L\}$ with $c_l > 1$. Recall that if $c_l \leq 1$ for all $l \in \{1,\dots,L\}$, model $^{\times}$D may not be the appropriate choice anyway since neither the effects of load sharing nor damage accumulation can be captured from a single component failure. Note that in this situation, consistency can still be accomplished using Corollary 4.20 instead of Theorem 4.17, but we leave this pathological case aside for the rest of the thesis.

Verifying the Preconditions of Theorem 4.17

Having established that conditions (D1) through (D6) are easily met, we still need to prove that Equation (4.16) of Theorem 4.17 can be satisfied. Consequently, we have to show that for all $\theta \in \Theta_0$ and $\theta^* \in \Theta_1$ there exist $l \in \{1,\dots,L\}$, $i \in \{2,\dots,c_l\}$ and $x \in \operatorname{supp}(A_{j,i}) \cap U_{l,i}$ for j with $l_j = l$ such that

$$g_{l_j,i}(\theta^*,\theta,x) \neq \frac{1}{2} \,,$$

where $g_{l,i}$ is the function from Equation (4.34). We have already ruled out the case where $c_l \leq 1$ for all $l \in \{1,\dots,L\}$, because otherwise conditions (D5) and (D6) would not be satisfied. Moreover, for each l with $c_l > 1$, we found that $U_{l,i} = (1,\infty)$ is sufficient to fulfill condition (D6). We can now proceed to verify the precondition of Theorem 4.17 in three steps:

(i) Show that the support of $A_{j,i}$ is unbounded if $c_{l_j} > 1$ and $i \in \{2,\dots,c_{l_j}\}$.

(ii) If $\theta_3 \neq \theta_3^*$, then the function $x \mapsto g_{l,i}(\theta^*,\theta,x)$ is either strictly decreasing or strictly increasing. Additionally, either $g_{l,i}(\theta^*,\theta,x) \to 0$ or $g_{l,i}(\theta^*,\theta,x) \to 1$ as $x \to \infty$.

(iii) If $\theta_3 = \theta_3^*$, then $g_{l,i}(\theta^*,\theta,x)$ is constant with respect to x. Moreover, if $c_l > 1$, then[20]

$$g_{l,i}(\theta^*,\theta,x) = \frac{1}{2} \quad \text{for all } i \in \{1,\dots,c_l\} \quad \Longleftrightarrow \quad \theta^* = \theta \,.$$

Each of these steps corresponds to one of the following three Lemmas.

[20] This technically requires $c_l > 2$ for some $l \in \{1,\dots,L\}$, since we defined $g_{l,i}$ for $i \geq 2$ only. However, we can easily extend the definition to $i = 1$. Formally, this means that Corollary 4.20 must be used instead of Theorem 4.17.

Lemma 4.23 (Unboundedness of the Accumulated Damage $A_{j,i}$).
Let $l \in \{1, \dots, L\}$ with $c_l > 1$. Then, for each $j \in \mathbb{N}$ with $l_j = l$ and $i \in \{2, \dots, c_{l_j}\}$, the support of $A_{j,i}$ is unbounded.

Proof. In the proof of Theorem 3.28, we noted that the conditional distribution of T_1 given $\tau_0 = \infty$, $C_0 > 0$ and $s_0 > 0$ follows a Weibull distribution (cf. Equation (3.65)). Since we are dealing with deterministic covariates here, the proof can be transferred to the unconditional distribution of $T_1^{(j)}$. By recalling that $T_0 \equiv 0$, $A_{j,1} \equiv 0$ and $B_{j,1} = s_{l_j}$, we obtain:

$$\begin{aligned}\mathbb{P}_{\theta^*}\left(T_1^{(j)} > t\right) &= \exp\left(-{}^{\times\mathrm{D}}H_{l_j,1}^{\theta^*}(t)\right)\\ &= \exp\left(-\frac{\theta_1^* B_{j,1}^{\theta_2^*-1}}{\tau_{l_j}^{\theta_3^*}(\theta_3^*+1)}\left[\left(B_{j,1}\left(t - T_0^{(j)}\right) + A_{j,1}\right)^{\theta_3^*+1} - A_{j,1}^{\theta_3^*+1}\right]\right)\\ &= \exp\left(-\frac{\theta_1^* s_{l_j}^{\theta_2^*-1}}{\tau_{l_j}^{\theta_3^*}(\theta_3^*+1)}\left(s_{l_j} t\right)^{\theta_3^*+1}\right) = \exp\left(-\left(\frac{t}{\sigma}\right)^a\right),\end{aligned}$$

which is the survival function of a Weibull distribution, where the scale parameter $\sigma = \sigma(\theta^*)$ and the shape parameter $a = a(\theta^*)$ are given by

$$\sigma(\theta^*) = \left(\frac{\tau_{l_j}^{\theta_3^*}(\theta_3^*+1)}{\theta_1^* s_{l_j}^{\theta_2^*+\theta_3^*}}\right)^{\frac{1}{\theta_3^*+1}} \qquad \text{and} \quad a(\theta^*) = \theta_3^* + 1\,.$$

Notice that this is consistent with Corollary A.47, because for the model without damage accumulation (i.e., $\theta_3^* = 0$), the shape parameter of the Weibull distribution is 1 (and a Weibull distribution with shape parameter 1 is an exponential distribution).
For $j \in \mathbb{N}$ with $c_{l_j} > 1$ and $i \in \{2, \dots, c_{l_j}\}$, we have

$$A_{j,i} = \sum_{k=1}^{i-1} B_{l_j,k}\left(T_k^{(j)} - T_{k-1}^{(j)}\right) \geq B_{l_j,1}\left(T_1^{(j)} - T_0^{(j)}\right) = s_{l_j} T_1^{(j)}\,,$$

and therefore

$$\mathbb{P}_{\theta^*}\left(A_{j,i} > t\right) \geq \mathbb{P}_{\theta^*}\left(T_1^{(j)} > \frac{t}{s_{l_j}}\right) = \exp\left(-\left(\frac{t}{s_{l_j}\sigma}\right)^a\right) > 0 \quad \text{for all } t \geq 0. \tag{4.37}$$

If the support of $A_{j,i}$ were bounded by some $t_0 > 0$, it would hold that

$$\mathbb{P}_{\theta^*}\left(A_{j,i} \leq t_0\right) = 1 \quad \Longleftrightarrow \quad \mathbb{P}_{\theta^*}\left(A_{j,i} > t_0\right) = 0\,,$$

but this contradicts Equation (4.37). Hence, the assertion follows. $\square$

Lemma 4.24 (Monotonicity & Asymptotics of the Link Function $g_{l,i}(\theta, \theta^*, \cdot)$ for $\theta_3^* \neq \theta_3$).
Let $l \in \{1, \dots, L\}$ with $c_l > 1$. If $\theta, \theta^* \in \Theta$ such that $\theta_3 \neq \theta_3^*$, then for each $i \in \{2, \dots, c_l\}$ the function $x \mapsto g_{l,i}(\theta^*, \theta, x)$ is either strictly decreasing or strictly increasing. Moreover, it holds that either

$$g_{l,i}(\theta^*, \theta, x) \longrightarrow 0 \quad \text{or} \quad g_{l,i}(\theta^*, \theta, x) \longrightarrow 1 \quad \text{as } x \longrightarrow \infty.$$

Proof. Let $l \in \{1, \dots, L\}$ with $c_l > 1$ and let $\theta, \theta^* \in \Theta$ such that $\theta_3 \neq \theta_3^*$. If we define

$$\kappa(l, i, \theta) := \frac{\tau_l^{\theta_3}(\theta_3 + 1)}{\theta_1 B_{l,i}^{\theta_2 - 1}}, \tag{4.38}$$

then $\kappa(l, i, \theta) > 0$ for each $\theta \in \Theta$ because of $\pi_1(\Theta) \subset (0, \infty)$ and $\pi_3(\Theta) \subset (-1, \infty)$. With this notation, the function $g_{l,i}$ from Equation (4.34) can be written as

$$g_{l,i}(\theta, \theta^*, x) = \exp\left(-\frac{1}{\kappa(l, i, \theta^*)}\left[\left(\kappa(l, i, \theta)\ln(2) + x^{\theta_3+1}\right)^{\frac{\theta_3^*+1}{\theta_3+1}} - x^{\theta_3^*+1}\right]\right). \tag{4.39}$$

The rest of the proof is divided into two parts. We start by covering the asymptotic behavior of the link function, and then proceed to prove its monotonicity.
Suppose first that $\theta_3^* > \theta_3$, so that $\frac{\theta_3^*+1}{\theta_3+1} > 1$. The Bernoulli inequality (cf. Brannan 2006, p. 237) provides:

$$\begin{aligned}
\left(\kappa(l, i, \theta)\ln(2) + x^{\theta_3+1}\right)^{\frac{\theta_3^*+1}{\theta_3+1}} - x^{\theta_3^*+1} &= \left(x^{\theta_3+1}\left(1 + \frac{\kappa(l, i, \theta)\ln(2)}{x^{\theta_3+1}}\right)\right)^{\frac{\theta_3^*+1}{\theta_3+1}} - x^{\theta_3^*+1} \\
&= x^{\theta_3^*+1}\left(1 + \frac{\kappa(l, i, \theta)\ln(2)}{x^{\theta_3+1}}\right)^{\frac{\theta_3^*+1}{\theta_3+1}} - x^{\theta_3^*+1} \\
&\geq x^{\theta_3^*+1}\left(1 + \frac{\theta_3^*+1}{\theta_3+1}\frac{\kappa(l, i, \theta)\ln(2)}{x^{\theta_3+1}}\right) - x^{\theta_3^*+1} \\
&= \underbrace{\frac{\theta_3^*+1}{\theta_3+1}\kappa(l, i, \theta)\ln(2)}_{>0}\, x^{\theta_3^*-\theta_3}.
\end{aligned}$$

Due to $\theta_3^* - \theta_3 > 0$ and the positivity of the leading constant, this implies that

$$\begin{aligned}
&\left(\kappa(l, i, \theta)\ln(2) + x^{\theta_3+1}\right)^{\frac{\theta_3^*+1}{\theta_3+1}} - x^{\theta_3^*+1} \\
&\quad\geq \frac{\theta_3^*+1}{\theta_3+1}\kappa(l, i, \theta)\ln(2)x^{\theta_3^*-\theta_3} \longrightarrow \infty \quad (x \longrightarrow \infty).
\end{aligned} \tag{4.40}$$

Substituting Equation (4.40) into Equation (4.39) then yields

$$g_{l,i}(\theta, \theta^*, x) \leq \exp\left(-\frac{1}{\kappa(l, i, \theta^*)}\left[\frac{\theta_3^*+1}{\theta_3+1}\kappa(l, i, \theta)\ln(2)x^{\theta_3^*-\theta_3}\right]\right) \longrightarrow 0 \quad (x \longrightarrow \infty).$$

On the other hand, if $\theta_3^* < \theta_3$, then $0 < \frac{\theta_3^*+1}{\theta_3+1} < 1$, and instead of Equation (4.40) we obtain with exactly the same arguments:

$$\begin{aligned}
&\left(\kappa(l, i, \theta)\ln(2) + x^{\theta_3+1}\right)^{\frac{\theta_3^*+1}{\theta_3+1}} - x^{\theta_3^*+1} \\
&\quad\leq \frac{\theta_3^*+1}{\theta_3+1}\kappa(l, i, \theta)\ln(2)x^{\theta_3^*-\theta_3} \longrightarrow 0 \quad (x \longrightarrow \infty),
\end{aligned} \tag{4.41}$$

since in this case the Bernoulli inequality gives $\leq$ instead of $\geq$ and $\theta_3^* - \theta_3 < 0$. By substituting Equation (4.41) into Equation (4.39) as well, we get

$$g_{l,i}\left(\theta,\theta^*,x\right) \geq \exp\left(-\frac{1}{\kappa\left(l,i,\theta^*\right)}\left[\frac{\theta_3^*+1}{\theta_3+1}\kappa(l,i,\theta)\ln(2)x^{\theta_3^*-\theta_3}\right]\right) \longrightarrow 1 \quad (x\longrightarrow\infty)\,,$$

which proves the asserted asymptotic behavior (note that $g_{l,i}$ is bounded below by 0 and above by 1 by its definition as a conditional probability).
For the monotonicity, we define the auxiliary function

$$\gamma(x) := \left(\kappa(l,i,\theta)\ln(2) + x^{\theta_3+1}\right)^{\frac{\theta_3^*+1}{\theta_3+1}} - x^{\theta_3^*+1}\,,$$

where in the notation we suppress the dependence on θ, θ^*, l and i. As a function of x, γ is continuously differentiable. Differentiation with respect to x yields:

$$\frac{\mathrm{d}}{\mathrm{d}x}\gamma(x) = \left(\theta_3^*+1\right)x^{\theta_3}\left(\kappa(l,i,\theta)\ln(2)+x^{\theta_3+1}\right)^{\frac{\theta_3^*-\theta_3}{\theta_3+1}} - \left(\theta_3^*+1\right)x^{\theta_3^*}\,.$$

If we again assume that $\theta_3^* > \theta_3$, then due to $\kappa(l,i,\theta)\ln(2) > 0$ we further obtain:

$$\frac{\mathrm{d}}{\mathrm{d}x}\gamma(x) > \left(\theta_3^*+1\right)\left[x^{\theta_3}\left(x^{\theta_3+1}\right)^{\frac{\theta_3^*-\theta_3}{\theta_3+1}} - x^{\theta_3^*}\right] = 0\,,$$

which means that γ is strictly increasing. If we assume $\theta_3^* < \theta_3$ instead, we analogously conclude that

$$\frac{\mathrm{d}}{\mathrm{d}x}\gamma(x) < \left(\theta_3^*+1\right)\left[x^{\theta_3}\left(x^{\theta_3+1}\right)^{\frac{\theta_3^*-\theta_3}{\theta_3+1}} - x^{\theta_3^*}\right] = 0\,.$$

Therefore, γ is strictly decreasing in this case. Finally, note that

$$g_{l,i}\left(\theta,\theta^*,x\right) = \exp\left(-\frac{\gamma(x)}{\kappa\left(l,i,\theta^*\right)}\right)\,.$$

Consequently, $g_{l,i}\left(\theta,\theta^*,x\right)$ is strictly decreasing in x if γ is strictly increasing, and vice versa. The assertion then follows. ☐

Lemma 4.25 (Constantness of the Link Function $g_{l,i}\left(\theta,\theta^*,\cdot\right)$ for $\theta_3^*=\theta_3$)**.**
Let $l \in \{1,\ldots,L\}$ and $j\in\mathbb{N}$ such that $l_j = l$. The function $g_{l,1} : \Theta^2\times\mathbb{R}\to\mathbb{R}$ is defined via

$$g_{l,1}(\theta^*,\theta,x) := \mathbb{P}_{\theta^*}\left(R_{j,1}^{\theta} > 0\right)\,. \tag{4.42}$$

Let $\theta,\theta^*\in\Theta$ with $\theta\neq\theta^*$ but $\theta_3=\theta_3^*$. Then, for each $i\in\{1,\ldots,c_l\}$, the function

$$x \longmapsto g_{l,i}(\theta,\theta^*,x)$$

is constant. If $c_l > 1$, then

$$g_{l,i}(\theta,\theta^*,x) = \frac{1}{2} \quad \textit{for all } i\in\{1,\ldots,c_l\} \quad\Longleftrightarrow\quad \theta=\theta^*\,.$$

The function $g_{l,1}$ defined in Equation (4.42) is by default constant with respect to its third argument x. The idea of how to extend the definition of the functions $g_{l,i}$ to the case

$i = 1$ is motivated by Corollary 4.20. With the above notation, case (ii) of this corollary then demands that, for some $x \in \mathbb{R}$ (the particular choice obviously does not matter),

$$g_{l,i}(\theta, \theta^*, x) \neq \frac{1}{2} ,$$

which closely resembles case (i). In this way, a unified notation becomes available in Corollary 4.20. However, we deliberately avoided this in order to better distinguish the fundamentally different cases (i) and (ii).

Proof. Lemma 4.25 can be regarded as a simple corollary of Theorem 2.28. Since Equation (2.57) applies also for $i = 1$, we immediately have

$$g_{l,1}(\theta^*, \theta, x) = \mathbb{P}_{\theta^*}\left(R^{\theta}_{j,1} > 0\right) = \exp\left(-\frac{\theta_1^* s_{l_j}^{\theta_2^*-1}}{\tau_{l_j}^{\theta_3^*}(\theta_3^*+1)}\left[\left(\frac{\tau_{l_j}^{\theta_3}(\theta_3+1)}{\theta_1 s_{l_j}^{\theta_2-1}}\ln(2)\right)^{\frac{\theta_3^*+1}{\theta_3+1}}\right]\right) ,$$

as $A_{j,1} \equiv 0$ and $B_{j,1} = s_{l_j}$. Note that $u = \ln(2)$, because unlike Theorem 2.28, we are looking at the standardized hazard transforms here. If $\theta_3 = \theta_3^*$, the functions $g_{l,i}$ simplify considerably, as pointed out in the proof of Theorem 2.28, and as a result Equation (2.57) breaks down to Equation (2.58). Hence,

$$g_{l,i}(\theta^*, \theta, x) = \exp\left(-\frac{\theta_1^* B_{l,i}^{\theta_2^*}}{\theta_1 B_{l,i}^{\theta_2}}\ln(2)\right) , \quad i \in \{1, \ldots, c_l\} ,$$

which is constant with respect to x. Since $\exp(-\ln(2)) = \frac{1}{2}$ and the exponential function is injective, we observe:

$$g_{l,i}(\theta^*, \theta, x) = \frac{1}{2} \quad \Longleftrightarrow \quad \frac{\theta_1^* B_{l,i}^{\theta_2^*}}{\theta_1 B_{l,i}^{\theta_2}} = 1 \quad \Longleftrightarrow \quad \theta_1^* B_{l,i}^{\theta_2^*} = \theta_1 B_{l,i}^{\theta_2} . \tag{4.43}$$

For $\theta_2 = \theta_2^*$, dividing both sides of Equation (4.43) by $B_{l,i}^{\theta_2}$ yields

$$\theta_1^* B_{l,i}^{\theta_2^*} \overset{!}{=} \theta_1 B_{l,i}^{\theta_2} \quad \Longleftrightarrow \quad \theta_1^* = \theta_1 .$$

For $\theta_1 = \theta_1^*$, Equation (4.43) can likewise only be satisfied if $\theta_2 = \theta_2^*$ holds as well. Suppose there exists $l \in \{1, \ldots, L\}$ with $c_l > 1$ such that Equation (4.43) is fulfilled for all $i \in \{1, \ldots, c_l\}$, but $\theta \neq \theta^*$ (the assumption that $\theta_3 = \theta_3^*$ is maintained). This means that $\theta_1 \neq \theta_1^*$ and $\theta_2 \neq \theta_2^*$ must apply. However, the function

$$b \longmapsto \frac{\theta_1^*}{\theta_1} b^{\theta_2^* - \theta_2}$$

is then either strictly increasing if $\theta_2^* > \theta_2$ or strictly decreasing if $\theta_2^* < \theta_2$. In either case, the function is injective and there exists at most one $b \in (0, \infty)$ that is mapped to 1. But due to $c_l > 1$, this would imply that $B_{l,1} = B_{l,2}$, which is a contradiction. Therefore, Equation (4.43) can be satisfied for all $i \in \{1, \ldots, c_l\}$ only if $\theta = \theta^*$, proving the assertion. □

From these Lemmas, we can easily deduce that the preconditions of Theorem 4.17 or - more accurately due to step (iii) - of Corollary 4.20 are met. For this, let $\theta \in \Theta_0$ and $\theta^* \in \Theta_1$, so that $\theta \neq \theta^*$. We need to distinguish whether $\theta_3 \neq \theta_3^*$ or $\theta_3 = \theta_3^*$:

(a) $\theta_3 \neq \theta_3^*$:

Choose any $l \in \{1, \ldots, L\}$ with $c_l > 1$ and $i \in \{2, \ldots, c_l\}$. By Lemma 4.24, we have either

$$g_{l,i}(\theta^*, \theta, x) \longrightarrow 0 \quad \text{or} \quad g_{l,i}(\theta^*, \theta, x) \longrightarrow 1 \quad \text{monotonically as } x \longrightarrow \infty.$$

This implies that for sufficiently large x, the value $\frac{1}{2}$ can no longer be attained. Formally, for any $0 < \varepsilon < \frac{1}{2}$, there exists $x_0 > 0$ such that

$$\left| g_{l,i}(\theta^*, \theta, x) - \frac{1}{2} \right| > \varepsilon \quad \text{for all } x > x_0.$$

Without loss of generality, we can assume that $x_0 > 1$. If $j \in \mathbb{N}$ with $l_j = l$, then

$$\mathbb{P}_{\theta^*}(A_{j,i} > x_0) > 0$$

according to Lemma 4.23. Therefore, $\operatorname{supp}(A_{j,i}) \cap U_{l,i} = \operatorname{supp}(A_{j,i}) \cap (1, \infty) \neq \emptyset$. For any $x \in \operatorname{supp}(A_{j,i}) \cap (1, \infty)$, we conclude:

$$g_{l,i}(\theta^*, \theta, x) \neq \frac{1}{2}.$$

(b) $\theta_3 = \theta_3^*$:

Choose any $l \in \{1, \ldots, L\}$ with $c_l > 1$. Because $\theta \neq \theta^*$, Lemma 4.25 states that there exists $i \in \{1, \ldots, c_l\}$ such that

$$g_{l,i}(\theta^*, \theta, x) \neq \frac{1}{2}.$$

If such an i with $i > 1$ exists, the condition from Equation (4.16) is satisfied. Otherwise, condition (ii) from Corollary 4.20 is satisfied.

As case (b) shows, we can avoid using Corollary 4.20 by having at least one class $l \in \{1, \ldots, L\}$ with $c_l > 2$. However, the consistency is still ensured in cases where this requirement is too restrictive. We end the paragraph with a remark that will prove useful in the following subsection.

Remark 4.26 (On the Support of the Accumulated Damage $A_{j,i}$)**.**
In Lemma 4.23 we proved that the support of the accumulated damage $A_{j,i}$ is unbounded if $j \in \mathbb{N}$ with $c_{l_j} > 1$ and $i \in \{2, \ldots, c_{l_j}\}$. In fact, then $\operatorname{supp}(A_{j,i}) = [0, \infty)$ necessarily holds. We can infer this property directly from the conditional hazard functions of the model ${}^{\times}$D. To this end, we note that for all $i \in \{1, \ldots, c_{l_j}\}$ and $t > T_{i-1}^{(j)}$, it holds that

$$\frac{f_{l_j,i}\left(t \,\middle|\, T_{1:(i-1)}^{(j)}\right)}{S_{l_j,i}\left(t \,\middle|\, T_{1:(i-1)}^{(j)}\right)} = {}^{\times D}h_{l_j,i}^{\theta^*}\left(t \,\middle|\, T_{1:(i-1)}^{(j)}\right) > 0 \quad \Longleftrightarrow \quad f_{l_j,i}\left(t \,\middle|\, T_{1:(i-1)}^{(j)}\right) > 0\,,$$

which means that the conditional distribution of $T_i^{(j)}$ given $T_{i-1}^{(j)}$ has support $[T_{i-1}^{(j)}, \infty)$. Therefore, the conditional distribution of the ith interarrival time $W_i^{(j)} = T_i^{(j)} - T_{i-1}^{(j)}$ has support $[0, \infty)$, and the same is true for $B_{l_j,i} W_i^{(j)}$, as $B_{l_j,i}$ is a positive deterministic constant. Since

$$A_{j,i} = \sum_{k=1}^{i-1} B_{l_j,k} W_k^{(j)} \,,$$

we then see that the support of $A_{j,i}$ is also given by $[0, \infty)$, because

$$\sigma\Big(W_1^{(j)}, \ldots, W_k^{(j)}\Big) = \sigma\Big(T_1^{(j)}, \ldots, T_k^{(j)}\Big) \,, \quad k \in \{1, \ldots, i-1\} \,. \tag{4.44}$$

Obviously, this reasoning is less formal than the proof of Lemma 4.25, which is why we record this statement only as a remark. For the mathematically rigorous approach, use Equation (4.44) to proceed with the successive reconstruction of the joint densities of $W_{1:k}^{(j)} = (W_1^{(j)}, \ldots, W_k^{(j)})$ similar to the proof of Lemma A.37 and obtain that

$$f_{W_{1:k}^{(j)}}(w_1, \ldots, w_k) > 0 \quad \text{for all } (w_1, \ldots, w_k) \in (0, \infty)^k. \tag{4.45}$$

The probability that $A_{j,i}$ falls in a given interval $\mathrm{Int} \subset [0, \infty)$ can then be written as an integral of this joint density. Since $f_{W_{1:(i-1)}^{(j)}} > 0$ almost everywhere according to Equation (4.45), it follows that $\mathbb{P}_{\theta^*}\left(A_{j,i} \in \mathrm{Int}\right) > 0$. Finally, due to $A_{j,i} \geq 0$, this implies that $\mathrm{supp}(A_{j,i}) = [0, \infty)$.

4.3.2. Consistency of the 3-Sign Depth Test for Damage Accumulation

In Subsection 2.3.2, we introduced the Basquin load sharing model with multiplicative damage accumulation as a relative risk regression model. In all kinds of regression models, we are often less interested in the actual model parameter than in identifying which covariates influence the outcome at all. For the model $^\times$D, our primary concern is to determine whether the effect of damage accumulation is significant (i.e., $\theta_3 > 0$). If not, we can accept $\theta_3 = 0$ and use the simpler model B without damage accumulation instead. The purpose of this subsection is therefore to show the consistency of the 3-sign depth test for the hypotheses

$$\mathcal{H}_0 \,:\, \theta_3^* = 0 \quad \text{vs.} \quad \mathcal{H}_1 \,:\, \theta_3^* \neq 0 \,.$$

The subset Θ_0 of the parameter space Θ belonging to the null hypothesis is thus given by

$$\Theta_0 := \pi_3^{-1}\left(\{0\}\right) = \{\theta \in \Theta \,:\, \pi_3(\theta) = \theta_3 = 0\} \,.$$

If we choose the parameter space Θ as large as possible, that is, $\Theta = (0, \infty) \times \mathbb{R} \times (-1, \infty)$, then

$$\Theta_0 = \left\{(\theta_1, \theta_2, 0)^\top \in \mathbb{R}^3 \,:\, \theta_1 > 0\right\} \,.$$

In particular, Θ_0 is non-compact and therefore neither Theorem 4.17 nor Corollary 4.20 can be applied. We can still show the consistency of the 3-sign depth test by exploiting that $\theta_3 = 0$ holds for all $\theta \in \Theta_0$, leading to particularly simple hazard transforms. As in Subsection 4.3.1, we assume the models $\mathcal{M}_l$, $l \in \{1, \ldots, L\}$, induced by the intensities from Equation (4.28). We again consider the standardized transform, so that

$$R^{\theta}_{j,i} := {}^{\times\mathrm{D}}H^{\theta}_{l_j,i}\Big(T^{(j)}_i \,|\, T^{(j)}_{1:(i-1)}\Big) - \ln(2)\,, \quad \theta \in \Theta,\, j \in \{1,\dots,J\},\, i \in \{1,\dots,c_{l_j}\}\,.$$

Throughout this subsection, we further suppose that condition (D3) is satisfied and thus there are no asymptotically negligible classes.

Theorem 4.27 (Consistency of the 3-Sign Depth Test for Damage Accumulation)**.** *Let $\{0\} \subsetneq \Theta_3 \subset (-1,\infty)$ and let the parameter space be given by $\Theta = (0,\infty)\times\mathbb{R}\times\Theta_3$. If there exists $l \in \{1,\dots,L\}$ such that $c_l > 1$, then the 3-sign depth test for the hypotheses*

$$\mathcal{H}_0 \;:\; \theta_3^* = 0 \quad \textit{vs.} \quad \mathcal{H}_1 \;:\; \theta_3^* \neq 0$$

is consistent in the model ${}^{\times}D$.

Proof. The proof is similar to that of Theorem 4.17. We start by fixing any $\theta^* \in \Theta\setminus\Theta_0$ and choose $l \in \{1,\dots,L\}$ with $c_l > 1$. For each $\theta \in \Theta_0$ (and hence $\theta_3 = 0$), $j \in \mathbb{N}$ with $l_j = l$, and $i \in \{2,\dots,c_l\}$, we then obtain:

$$\begin{aligned}\mathbb{P}_{\theta^*}\left(R^{\theta}_{j,i} > 0\,\Big|\,T^{(j)}_{1:(i-1)}\right) &= \exp\left(-\frac{\theta_1^* B_{l_j,i}^{\theta_2^*-1}}{\tau_{l_j}^{\theta_3^*}\,(\theta_3^*+1)}\left[\left(\frac{\ln(2)}{\theta_1 B_{l_j,i}^{\theta_2-1}} + A_{j,i}\right)^{\theta_3^*+1} - A_{j,i}^{\theta_3^*+1}\right]\right)\\ &= \exp\left(-\frac{1}{\kappa\left(l_j,i,\theta^*\right)}\left[\left(\kappa\left(l_j,i,\theta\right)\ln(2) + A_{j,i}\right)^{\theta_3^*+1} - A_{j,i}^{\theta_3^*+1}\right]\right)\,,\end{aligned}$$

where as before (cf. Equation (4.38))

$$\kappa(l,i,\theta) = \frac{\tau_l^{\theta_3}(\theta_3+1)}{\theta_1 B_{l,i}^{\theta_2-1}}\,.$$

We observe that the parameter θ only affects the constant κ. With this in mind, we define

$$\tilde{g}_{l,i}(x,y) := \exp\left(-\frac{1}{\kappa\left(l,i,\theta^*\right)}\left[\left(\ln(2)x + y\right)^{\theta_3^*+1} - y^{\theta_3^*+1}\right]\right)\,,$$

so that

$$\mathbb{P}_{\theta^*}\left(R^{\theta}_{j,i} > 0\,\Big|\,T^{(j)}_{1:(i-1)}\right) = \tilde{g}_{l_j,i}\left(\kappa\left(l_j,i,\theta\right), A_{j,i}\right) \quad \text{for all } \theta \in \Theta_0.$$

Recalling how the conditional distribution of the hazard transforms is derived from an exponential distribution (compare the elaboration of Equation (2.49)), we see that it is itself continuous and it holds:

$$\mathbb{P}_{\theta^*}\left(R^{\theta}_{j,i} < 0\,\Big|\,T^{(j)}_{1:(i-1)}\right) = 1 - \tilde{g}_{l_j,i}\left(\kappa\left(l_j,i,\theta\right), A_{j,i}\right) \quad \text{for all } \theta \in \Theta_0.$$

As a function of θ only, $\kappa(l,i,\theta)$ maps Θ_0 onto the interval $(0,\infty)$. To see this, remember that $\kappa > 0$ and let $z \in (0,\infty)$. Then,

$$\kappa\Big(l,i,\underbrace{(z^{-1},1,0)}_{\in\Theta_0}\Big) = \frac{\tau_l^0\,(0+1)}{\frac{1}{z}B_{l,i}^{1-1}} = \frac{1}{\frac{1}{z}} = z\,. \tag{4.46}$$

Furthermore, for any $y \in [0,\infty)$, it holds:

$$\lim_{x\to 0} \tilde{g}_{l,i}(x,y) = 1 \quad \text{and} \quad \lim_{x\to\infty} \tilde{g}_{l,i}(x,y) = 0\,.$$

From now on, fix an arbitrary $i \in \{2,\dots,c_l\}$. Choose any $y_0 > 0$. Since $\tilde{g}_{l,i}$ is continuous, the intermediate value theorem provides the existence of $x_0 \in (0,\infty)$ satisfying

$$\tilde{g}_{l,i}(x_0, y_0) = \frac{1}{2}\,.$$

Let $\theta^0 \in \Theta_0$ such that $\kappa(l,i,\theta^0) = x_0$ (e.g., $\theta^0 = \left(x_0^{-1},1,0\right)$ according to Equation (4.46)). The proof of Lemma 4.24 shows that $\tilde{g}_{l,i}(x,y)$ is either strictly decreasing or strictly increasing in y, depending on whether $\theta_3^* > 0$ or $\theta_3^* < 0$. Without loss of generality, we consider only the case $\theta_3^* > 0$, so that $\tilde{g}_{l,i}(x,y)$ is strictly decreasing in y (the other case is completely analogous). We can thus find $\varepsilon_1 > 0$ and $0 < y_1 < y_0 < y_2 < \infty$ with

$$\begin{aligned} &\tilde{g}_{l,i}(x_0,y) > \frac{1}{2} + \varepsilon_1 \quad \text{for all } y \in [0,y_1] \\ \text{and} \qquad &\tilde{g}_{l,i}(x_0,y) < \frac{1}{2} - \varepsilon_1 \quad \text{for all } y \in [y_2,\infty). \end{aligned}$$

Moreover, Remark 4.26 shows that $\operatorname{supp}(A_{j,i}) = [0,\infty)$, so that there exists $\varepsilon_2 > 0$ with

$$\mathbb{P}_{\theta^*}\left(A_{j,i} \in [0,y_1]\right) > \varepsilon_2 \quad \text{and} \quad \mathbb{P}_{\theta^*}\left(A_{j,i} \in [y_2,\infty)\right) > \varepsilon_2\,.$$

From here, we continue by using the techniques from the proof of Theorem 4.17, so we omit some details occasionally. In analogy to Equation (4.23), we define

$$\mathcal{N}^+_{\mathrm{Int}}(\theta) := \sum_{j\in J_l} \mathbb{1}\left\{R^\theta_{j,i} > 0,\ A_{j,i} \in \mathrm{Int}\right\} \quad \text{and} \quad \mathcal{N}^-_{\mathrm{Int}}(\theta) := \sum_{j\in J_l} \mathbb{1}\left\{R^\theta_{j,i} < 0,\ A_{j,i} \in \mathrm{Int}\right\}\,,$$

where $J_l := \{j \in \{1,\dots,J\} : l_j = l\}$. We again set $j_0 = \min J_l$ as a reference index (any other choice of $j_0 \in J_l$ would be equally valid). The strong law of large numbers provides that $\mathbb{P}_{\theta^*}$-almost surely we have:

$$\begin{aligned} \frac{\mathcal{N}^+_{[0,y_1]}(\theta^0)}{\sum_{j\in J_l} \mathbb{1}\left\{A_{j,i} \in [0,y_1]\right\}} - \frac{1}{2} \xrightarrow{J\to\infty} &\frac{\mathbb{P}_{\theta^*}\left(R^{\theta^0}_{j_0,i} > 0,\ A_{j_0,i} \in [0,y_1]\right)}{\mathbb{P}_{\theta^*}\left(A_{j_0,i} \in [0,y_1]\right)} - \frac{1}{2} \\ &= \frac{\int_{\left\{A_{j_0,i}\in[0,y_1]\right\}} \tilde{g}_{l_{j_0},i}\left(\kappa\left(l_{j_0},i,\theta^0\right), A_{j_0,i}\right) \mathrm{d}\mathbb{P}_{\theta^*}}{\mathbb{P}_{\theta^*}\left(A_{j_0,i} \in [0,y_1]\right)} - \frac{1}{2} \\ &= \frac{\int_{\left\{A_{j_0,i}\in[0,y_1]\right\}} \tilde{g}_{l,i}\left(x_0, A_{j_0,i}\right) \mathrm{d}\mathbb{P}_{\theta^*}}{\mathbb{P}_{\theta^*}\left(A_{j_0,i} \in [0,y_1]\right)} - \frac{1}{2} \\ &> \frac{\left(\frac{1}{2}+\varepsilon_1\right)\mathbb{P}_{\theta^*}\left(A_{j_0,i} \in [0,y_1]\right)}{\mathbb{P}_{\theta^*}\left(A_{j_0,i} \in [0,y_1]\right)} - \frac{1}{2} = \varepsilon_1\,. \end{aligned} \tag{4.47}$$

Likewise, we obtain $\mathbb{P}_{\theta^*}$-almost surely that

$$\frac{\mathcal{N}^-_{[y_2,\infty)}(\theta^0)}{\sum_{j\in J_l} \mathbb{1}\left\{A_{j,i} \in [y_2,\infty)\right\}} - \frac{1}{2} \xrightarrow{J\to\infty} \frac{\mathbb{P}_{\theta^*}\left(R^{\theta^0}_{j_0,i} < 0,\ A_{j_0,i} \in [y_2,\infty)\right)}{\mathbb{P}_{\theta^*}\left(A_{j_0,i} \in [y_2,\infty)\right)} - \frac{1}{2}$$

$$= \frac{\int_{\left\{A_{j_0,i}\in[y_2,\infty)\right\}} 1 - \tilde{g}_{l_{j_0},i}\left(\kappa\left(l_{j_0}, i, \theta^0\right), A_{j_0,i}\right)\, \mathrm{d}\mathbb{P}_{\theta^*}}{\mathbb{P}_{\theta^*}\left(A_{j_0,i}\in[y_2,\infty)\right)} - \frac{1}{2}$$
$$= \frac{\int_{\left\{A_{j_0,i}\in[y_2,\infty)\right\}} 1 - \tilde{g}_{l,i}\left(x_0, A_{j_0,i}\right)\, \mathrm{d}\mathbb{P}_{\theta^*}}{\mathbb{P}_{\theta^*}\left(A_{j_0,i}\in[y_2,\infty)\right)} - \frac{1}{2}$$
$$> \frac{\left(1-\left(\frac{1}{2}-\varepsilon_1\right)\right)\mathbb{P}_{\theta^*}\left(A_{j_0,i}\in[y_2,\infty)\right)}{\mathbb{P}_{\theta^*}\left(A_{j_0,i}\in[y_2,\infty)\right)} - \frac{1}{2} = \varepsilon_1\,. \quad (4.48)$$

Furthermore, by setting

$$\tilde{\varepsilon}_2 := \frac{p_l}{\sum_{l=1}^{L} p_l \cdot c_l}\varepsilon_2 > 0\,,$$

we obtain $\mathbb{P}_{\theta^*}$-almost surely that

$$\frac{1}{\eta}\sum_{j\in J_l} \mathbb{1}\left\{A_{j_0,i}\in[0,y_1]\right\} \overset{\eta\to\infty}{\longrightarrow} \frac{p_l}{\sum_{l=1}^{L} p_l \cdot c_l}\mathbb{P}_{\theta^*}\left(A_{j_0,i}\in[0,y_1]\right) > \tilde{\varepsilon}_2 \quad (4.49)$$

and

$$\frac{1}{\eta}\sum_{j\in J_l} \mathbb{1}\left\{A_{j,i}\in[y_2,\infty)\right\} \overset{\eta\to\infty}{\longrightarrow} \frac{p_l}{\sum_{l=1}^{L} p_l \cdot c_l}\mathbb{P}_{\theta^*}\left(A_{j,i}\in[y_2,\infty)\right) > \tilde{\varepsilon}_2\,. \quad (4.50)$$

For $\varepsilon < \min\{\varepsilon_1, \tilde{\varepsilon}_2\}$, combining Equations (4.47), (4.48), (4.49) and (4.50) yields the existence of $\Omega_0 \subset \Omega$ with $\mathbb{P}_{\theta^*}(\Omega_0) = 1$ and the property that for any $\omega \in \Omega_0$ there exists η_ω such that for all $\eta \geq \eta_\omega$ simultaneously holds:

$$\sum_{j\in J_l} \mathbb{1}\left\{A_{j,i}(\omega)\in[0,y_1]\right\} > \varepsilon\eta\,,$$
$$\sum_{j\in J_l} \mathbb{1}\left\{A_{j,i}(\omega)\in[y_2,\infty)\right\} > \varepsilon\eta\,,$$
$$\frac{\mathcal{N}^{+}_{[0,y_1]}(\theta^0)(\omega)}{\sum_{j\in J_l}\mathbb{1}\left\{A_{j,i}(\omega)\in[0,y_1]\right\}} - \frac{1}{2} > \varepsilon\,,$$
$$\frac{\mathcal{N}^{-}_{[y_2,\infty)}(\theta^0)(\omega)}{\sum_{j\in J_l}\mathbb{1}\left\{A_{j,i}(\omega)\in[y_2,\infty)\right\}} - \frac{1}{2} > \varepsilon\,.$$

Corollary 2.26 implies that the standardized hazard transforms for $\theta \in \Theta_0$ are given by:

$$\begin{aligned} R^{\theta}_{j,i} &= \theta_1 B^{\theta_2-1}_{l_j,i}\left[A_{j,i+1}-A_{j,i}\right] - \ln(2) \\ &= \theta_1 B^{\theta_2-1}_{l_j,i}\left[B_{l_j,i}\left(T^{(j)}_i - T^{(j)}_{i-1}\right)\right] - \ln(2) \\ &= \theta_1 B^{\theta_2}_{l_j,i}\left(T^{(j)}_i - T^{(j)}_{i-1}\right) - \ln(2)\,. \end{aligned}$$

For any $\theta \in \Theta_0$, either $\theta_1 B^{\theta_2}_{l,i} \leq \theta^0_1 B^{\theta^0_2}_{l,i}$ or $\theta_1 B^{\theta_2}_{l,i} \geq \theta^0_1 B^{\theta^0_2}_{l,i}$ applies. If $\theta_1 B^{\theta_2}_{l,i} \leq \theta^0_1 B^{\theta^0_2}_{l,i}$, then

$$R^{\theta}_{j,i} \leq R^{\theta^0}_{j,i} \quad \text{for all } j \in J_l.$$

Thus, for any interval Int,

$$\mathcal{N}^{-}_{\mathrm{Int}}(\theta) = \sum_{j \in J_l} \mathbb{1}\left\{R^{\theta}_{j,i} < 0,\ A_{j,i} \in \mathrm{Int}\right\} \geq \sum_{j \in J_l} \mathbb{1}\left\{R^{\theta^0}_{j,i} < 0,\ A_{j,i} \in \mathrm{Int}\right\} = \mathcal{N}^{-}_{\mathrm{Int}}(\theta^0)\,.$$

For any $\omega \in \Omega_0$ and $\eta \geq \eta_\omega$ we can conclude that

$$\frac{\mathcal{N}^{-}_{[y_2,\infty)}(\theta)(\omega)}{\sum_{j \in J_l} \mathbb{1}\left\{A_{j,i}(\omega) \in [y_2,\infty)\right\}} - \frac{1}{2} \geq \frac{\mathcal{N}^{-}_{[y_2,\infty)}(\theta^0)(\omega)}{\sum_{j \in J_l} \mathbb{1}\left\{A_{j,i}(\omega) \in [y_2,\infty)\right\}} - \frac{1}{2} > \varepsilon\,. \tag{4.51}$$

Conversely, if $\theta_1 B^{\theta_2}_{l,i} \geq \theta^0_1 B^{\theta^0_2}_{l,i}$, then

$$R^{\theta}_{j,i} \geq R^{\theta^0}_{j,i} \quad \text{for all } j \in J_l,$$

and similar to above for any $\omega \in \Omega_0$ and $\eta \geq \eta_\omega$ it holds that

$$\frac{\mathcal{N}^{+}_{[0,y_1]}(\theta)(\omega)}{\sum_{j \in J_l} \mathbb{1}\left\{A_{j,i}(\omega) \in [0,y_1]\right\}} - \frac{1}{2} \geq \frac{\mathcal{N}^{+}_{[0,y_1]}(\theta^0)(\omega)}{\sum_{j \in J_l} \mathbb{1}\left\{A_{j,i}(\omega) \in [0,y_1]\right\}} - \frac{1}{2} > \varepsilon\,. \tag{4.52}$$

The proof now finishes completely analogous to the proof of Theorem 4.17. Since the transforms are ordered with respect to $\leq_{\mathrm{acc}}$, for $\eta \geq \eta_\omega$ we get

$$\sup_{n-m+1 \geq \varepsilon\eta} \frac{\mathcal{N}^{+}_{m,n}(\theta)(\omega)}{n-m+1} - \frac{1}{2} \geq \frac{\mathcal{N}^{+}_{[0,y_1]}(\theta)(\omega)}{\sum_{j \in J_l} \mathbb{1}\left\{A_{j,i}(\omega) \in [0,y_1]\right\}} - \frac{1}{2}$$

and

$$\sup_{n-m+1 \geq \varepsilon\eta} \frac{\mathcal{N}^{-}_{m,n}(\theta)(\omega)}{n-m+1} - \frac{1}{2} \geq \frac{\mathcal{N}^{-}_{[y_2,\infty)}(\theta)(\omega)}{\sum_{j \in J_l} \mathbb{1}\left\{A_{j,i}(\omega) \in [y_2,\infty)\right\}} - \frac{1}{2}\,.$$

Because of Equations (4.51) and (4.52), for all $\theta \in \Theta_0$ and $\eta \geq \eta_\omega$ we have either

$$\frac{\mathcal{N}^{+}_{[0,y_1]}(\theta)(\omega)}{\sum_{j \in J_l} \mathbb{1}\left\{A_{j,i}(\omega) \in [0,y_1]\right\}} - \frac{1}{2} > \varepsilon \quad \text{or} \quad \frac{\mathcal{N}^{-}_{[y_2,\infty)}(\theta)(\omega)}{\sum_{j \in J_l} \mathbb{1}\left\{A_{j,i}(\omega) \in [y_2,\infty)\right\}} - \frac{1}{2} > \varepsilon\,.$$

Accordingly, for all $\theta \in \Theta_0$ and $\eta \geq \eta_\omega$ it follows that

$$\sup_{n-m+1 \geq \varepsilon\eta} \frac{\max\left\{\mathcal{N}^{+}_{m,n}(\theta)(\omega), \mathcal{N}^{-}_{m,n}(\theta)(\omega)\right\}}{n-m+1} - \frac{1}{2} > \varepsilon\,,$$

and therefore (cf. Equation (4.13))

$$\omega \in \bigcap_{\theta \in \Theta_0} \mathbb{A}_{\eta,\varepsilon}(\theta) \quad \forall \eta \geq \eta_\omega \quad \Longrightarrow \quad \omega \in \liminf_{\eta \to \infty} \bigcap_{\theta \in \Theta_0} \mathbb{A}_{\eta,\varepsilon}(\theta)\,.$$

This implies that

$$\Omega_0 \subset \liminf_{\eta \to \infty} \bigcap_{\theta \in \Theta_0} \mathbb{A}_{\eta,\varepsilon}(\theta)\,,$$

so we can again deduce the consistency of the 3-sign depth test from Corollary 4.10. □

We conclude the chapter by highlighting a major difference between Theorem 4.17 and Theorem 4.27. Unlike the parameters θ_ν, $\nu = 1, \dots, \nu_{\max}$, in the proof of Theorem 4.17, the "reference parameter" θ^0 here is more of a hindrance in terms of consistency, since by construction for all $\delta > 0$ it admits

$$\left| \mathbb{P}_{\theta^*}\left(R_{j,i}^{\theta^0} > 0 \,|\, T_{1:(i-1)}^{(j)}\right) - \frac{1}{2} \right| < \delta$$

with positive probability. This means that there is a non-zero chance that the standardized hazard transforms at θ^0 will take on positive and negative signs with approximately equal probability. Nevertheless, the above proof shows that even for such a seemingly ill-suited parameter, there still exist regions Int_1, $\mathrm{Int}_2 \subset [0, \infty)$ so that

$$\mathbb{P}_{\theta^*}\left(R_{j,i}^{\theta^0} > 0 \,|\, A_{j,i} \in \mathrm{Int}_1\right) > \frac{1}{2} + \varepsilon \quad \text{and} \quad \mathbb{P}_{\theta^*}\left(R_{j,i}^{\theta^0} > 0 \,|\, A_{j,i} \in \mathrm{Int}_2\right) < \frac{1}{2} - \varepsilon \,.$$

Moreover, there are no "worse" parameters than θ^0, because for any $\theta \in \Theta_0$ either

$$\mathbb{P}_{\theta^*}\left(R_{j,i}^{\theta} > 0 \,|\, A_{j,i} \in \mathrm{Int}_1\right) \geq \mathbb{P}_{\theta^*}\left(R_{j,i}^{\theta^0} > 0 \,|\, A_{j,i} \in \mathrm{Int}_1\right) > \frac{1}{2} + \varepsilon \tag{4.53}$$

or

$$\mathbb{P}_{\theta^*}\left(R_{j,i}^{\theta} > 0 \,|\, A_{j,i} \in \mathrm{Int}_2\right) \leq \mathbb{P}_{\theta^*}\left(R_{j,i}^{\theta^0} > 0 \,|\, A_{j,i} \in \mathrm{Int}_2\right) < \frac{1}{2} - \varepsilon \,. \tag{4.54}$$

The two Equations (4.53) and (4.54) are essentially the counterparts of Equations (4.21) and (4.22), which explains why the proofs from there on are mostly congruent.
As a final corollary of Theorem 4.27, the consistency of the 3-sign depth test for damage accumulation can also be inferred for all smaller parameter spaces $\Theta \subset (0, \infty) \times \mathbb{R} \times (-1, \infty)$ so long as $\Theta_0 = \pi_3^{-1}(\{0\}) \neq \emptyset$ and $\Theta_1 = \Theta \setminus \Theta_0 \neq \emptyset$. The statement itself is trivial, since consistency is always inherited to subsets of parameter spaces, but it underlines that Theorem 4.27 has been formulated in the most general way.

5. Maximum Likelihood Estimation for Parametric Intensity-Based Models

As a third and final way to draw statistical inference for parametric intensity-based counting process models, we consider the established method of maximum likelihood estimation. The maximum likelihood estimator has been studied extensively, and as a result its properties - including the asymptotic distribution - are well known. Also, the approach is accompanied by a variety of useful applications such as the likelihood ratio test. They all require that we can specify the likelihood function of a counting process. Proposition A.35 from the appendix states that the probability structure of a counting process is uniquely determined by its stochastic intensity. This property enables us to express the likelihood function of a counting process in terms of its conditional intensity function. The caveat here is that this intensity function depends crucially on the underlying filtration: Generally speaking, the larger the filtration compared to the internal filtration of the counting process (i.e., the more external information is included in the stochastic intensity), the more complicated the likelihood function becomes. This is easily seen even in the framework of Section 2.1, where in the case of an intrinsic filtration, the likelihood function of the counting process N must also account for the randomness of the covariate X. As a solution, Cox proposed to discard factors of the likelihood from which “no useful information about the parameter of interest can be extracted” (Cox 1975, p. 272). Commonly, these factors arise from decomposing the full likelihood into a product of a marginal likelihood (e.g., the likelihood based only on the covariate X) and a conditional likelihood (e.g., the likelihood based on N given $X = x$), cf. Cox 1975, pp. 269–270. While in this example deleting a factor leads back to a likelihood function (either the marginal one or the conditional one), this is not true in general. Thus, to distinguish it from marginal and conditional likelihoods, Cox calls such a function, obtained by deleting certain factors of the full likelihood, a *partial likelihood*, see Cox 1975, p. 270.

What we later call the “likelihood function” of a counting process will be either an ordinary likelihood function or a partial likelihood function, depending on the chosen filtration. However, a major insight (and important contribution of Cox) is that the large sample properties are preserved when a partial likelihood function is considered instead of a likelihood function (e.g., the asymptotic distribution of the likelihood ratio test statistic), see Cox 1975, pp. 273–274. In most applications, it is therefore not necessary to draw a distinction between an ordinary likelihood and a partial likelihood. We thus follow the example of Andersen et al. 1993, pp. 403–404 and drop the word “partial” below. In the same vein, we omit further technical details on partial likelihoods, since they largely agree with the theory of ordinary likelihoods. We instead refer the interested reader to Borgan 1984, who gives sufficient conditions for the asymptotic properties of the maximum partial likelihood estimator in the multiplicative intensity model of Aalen 1978, and Andersen et al. 1993 (see in particular Sections II.7 and VI.1), where these conditions are transferred to general parametric intensity-based counting process models in Condition VI.1.1.

As we did previously for the minimum distance estimation and the K-sign depth test, we begin by specifying a suitable framework for a likelihood-based approach. As it turns out, this framework is effectively identical to that of Definition 3.2, but we restate it here for convenience.

Definition 5.1 (Framework for Maximum Likelihood Estimation).
Let $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in\mathcal{I}}, \mathbb{P}\right)$ be a filtered probability space, where $\mathcal{I} = [0, \tau]$ is a compact interval with $\tau \in (0, \infty)$. Let $N^{(1)}, \ldots, N^{(J)}$, $J \in \mathbb{N}$, denote i.i.d. copies of an adapted counting process $N = (N_t)_{t\in\mathcal{I}}$ with absolutely continuous $\left(\mathbb{P}, \{\mathcal{F}_t\}_{t\in\mathcal{I}}\right)$-compensator Λ. Let $\theta \in \Theta$ denote the parameter of interest, where $\Theta \subset \mathbb{R}^d$, $d \in \mathbb{N}$, is a bounded open set. A parametric model is given by a class $\mathcal{M}$ of cumulative intensities, that is,

$$\mathcal{M} = \{\Lambda_\theta \,:\, \theta \in \Theta\}\ .$$

Let again $\mathcal{J}_N$ denote the (random) set of time points belonging to the jumps of the counting process N, that is,

$$\mathcal{J}_N := \left\{t \,:\, N_t - \lim_{s\uparrow t} N_s \geq 1\right\}\ .$$

As before, N induces a Borel measure on $\mathcal{I}$, and for any bounded Borel function f holds:

$$\int_{\mathcal{I}} f \,\mathrm{d}\mu_N = \int_{\mathcal{I}} f(t)\,\mathrm{d}N_t = \sum_{t\in\mathcal{J}_N} f(t)\,,$$

that is, integration with respect to N is equal to summation over the function evaluations at the jump points of N.

In this framework, the likelihood function of a counting process can already be specified (compare Equation (2.7.13) of Andersen et al. 1993, p. 103). The corresponding formula features increments of the cumulative intensity Λ_θ in infinitesimal form, denoted by $\mathrm{d}\Lambda_\theta(t)$ (we touched on this in the introduction and in Remark A.48, where we used $\Lambda_\theta(\mathrm{d}t)$ instead of $\mathrm{d}\Lambda_\theta(t)$, but avoid this notation elsewhere). If Λ_θ is absolutely continuous with respect to the Lebesgue measure and admits the density λ_θ, that is,

$$\Lambda_\theta(t) = \int_0^t \lambda_\theta(u)\,\mathrm{d}u\,, \quad t \in \mathcal{I}\,,$$

then for any bounded Borel function f we observe:

$$\int_{\mathcal{I}} f(t)\,\mathrm{d}\Lambda_\theta(t) = \int_{\mathcal{I}} f(t)\lambda_\theta(t)\,\mathrm{d}t\,. \tag{5.1}$$

Because of Equation (5.1), we suggestively write $\mathrm{d}\Lambda_\theta(t) = \lambda_\theta(t)\,\mathrm{d}t$. Substituting this identity into a likelihood function given in terms of $\mathrm{d}\Lambda_\theta(t)$ will yield the expression we are aiming for. Since likelihood functions need only be specified up to proportionality, the remainder $\mathrm{d}t$ is irrelevant here and can be dropped.
The overall merit of this detour is that the likelihood function becomes substantially simpler, as it can be explicitly stated via the intensity process λ_θ that we used to define our models in the first place. The drawback, on the other hand, is that we have to assume that any cumulative intensity contained in $\mathcal{M}$ is $\mathbb{P}_{\theta^*}$-almost surely absolutely continuous, where as before θ^* is the true parameter. Obviously, we also have to require that the true compensator Λ_{θ^*} is contained in the model $\mathcal{M}$ at all, so we can usefully adopt Assumption 3.3 from Chapter 3 one-to-one. The corresponding assumptions (M1) and (M2) are as follows:

(M1) The model includes the compensator Λ. Hence, there is a true parameter $\theta^* \in \Theta$, such that
$$\Lambda = \Lambda_{\theta^*} \,.$$

(M2) Any cumulative intensity contained in $\mathcal{M}$ is $\mathbb{P}_{\theta^*}$-almost surely absolutely continuous: For each $\theta \in \Theta$, there exists a Lebesgue density λ_θ satisfying
$$\Lambda_\theta(t) = \int_0^t \lambda_\theta(u)\,\mathrm{d}u\,, \quad t \in \mathcal{I}\,.$$
Without loss of generality, we can assume λ_θ to be left-continuous. In the case $\theta = \theta^*$, λ_θ is the $(\mathbb{P}, \{\mathcal{F}_t\}_{t\in\mathcal{I}})$-intensity of N.

With these assumptions, we can now give the likelihood function of a counting process N.

Definition 5.2 (Likelihood Function of a Counting Process; Andersen et al. 1993, p. 402)**.** *In the framework of Definition 5.1 and under assumption* (M2)*, the likelihood function of a counting process N takes the form*
$$L(\theta) = L(\theta, N) = \prod_{t\in\mathcal{J}_N} \lambda_\theta(t) \exp\left(-\int_0^\tau \lambda_\theta(u)\,\mathrm{d}u\right)\,.$$

While the formal framework for the (partial) likelihood function of a counting process is more suitable in Andersen et al. 1993, the notation used here is more in line with Proposition 7.2.III. of Daley and Vere-Jones 2003, p. 232. Because of
$$\prod_{t\in\mathcal{J}_N} \lambda_\theta(t) = \exp\left(\sum_{t\in\mathcal{J}_N} \ln \lambda_\theta(t)\right) = \exp\left(\int_{\mathcal{I}} \ln \lambda_\theta(u)\,\mathrm{d}N_u\right),$$
the likelihood function is often stated in the following way (e.g., in Karr 1991 or Snyder and Miller 1991):
$$L(\theta) = \exp\left(\int_{\mathcal{I}} \ln \lambda_\theta(u)\,\mathrm{d}N_u - \int_{\mathcal{I}} \lambda_\theta(u)\,\mathrm{d}u\right).$$

If $T = (T_i)_{i\in\mathbb{N}}$ as usual denotes the point process associated with N, then $\mathcal{J}_N$ consists of the respective realizations $t_1, \dots, t_{N_\tau}$ of T in $\mathcal{I} = [0, \tau]$ and the likelihood function can be expressed as follows:
$$L(\theta) = \prod_{i=1}^{N_\tau} \lambda_\theta(t_i) \exp\left(-\int_0^\tau \lambda_\theta(u)\,\mathrm{d}u\right)\,. \tag{5.2}$$

The likelihood function of the jth counting process $N^{(j)}$ is analogously given by
$$L^{(j)}(\theta) = \prod_{i=1}^{N_\tau^{(j)}} \lambda_\theta^{(j)}\left(t_i^{(j)}\right) \exp\left(-\int_0^\tau \lambda_\theta^{(j)}(u)\,\mathrm{d}u\right)\,,$$
and hence the joint likelihood function of J independent counting processes $N^{(1)}, \dots, N^{(J)}$

is

$$L_J(\theta) = \prod_{j=1}^{J} L^{(j)}(\theta) = \prod_{j=1}^{J} \prod_{i=1}^{N_\tau^{(j)}} \lambda_\theta^{(j)}\left(t_i^{(j)}\right) \exp\left(-\int_0^\tau \lambda_\theta^{(j)}(u)\,\mathrm{d}u\right) . \tag{5.3}$$

In the case of an internal filtration, the likelihood function from Equation (5.2) actually provides the full likelihood (and not a partial likelihood) of N. A detailed derivation for this case is provided in Snyder and Miller 1991, pp. 296–302. We will settle here for the shorter but similarly instructive interpretation of Karr 1991, p. 72, which is the subject of the following remark. We point out to the reader that this remark is not intended to formally prove the form of the likelihood, but rather as an aid to its understanding.

Remark 5.3 (Interpretation of the Likelihood Function of a Counting Process)**.**
In this remark, we explain the shape of the likelihood function of a counting process N on the example of a one-point process. The interpretation can then be applied to any counting process. Let therefore N be the one-point process given via

$$N(t,\omega) = \mathbb{1}_{(-\infty,t]}\left(X(\omega)\right) ,$$

and assume that X is a continuous random variable with density function f and cumulative distribution function F. According to Lemma A.30, the intensity process with respect to the internal filtration of N is then given by

$$\lambda(t) = h(t) \cdot \mathbb{1}_{\{t<X\}} = \frac{f(t)}{1-F(t)} \cdot \mathbb{1}_{\{t<X\}} .$$

Any left-continuous modification of λ, as required in (M2), satisfies $\lambda(X) = h(X)$. Due to

$$f(x) = h(x)\exp\left(-H(x)\right) ,$$

the distribution of X is completely characterized by its hazard function h (cf. proof of Theorem A.46, particularly Equation (A.51)), and a model $\mathcal{M}$ can be defined by a parametric family of hazard functions h_θ, $\theta \in \Theta$. For the likelihood function we get:

$$L(\theta,x) = f_\theta(x) = h_\theta(x)\exp\left(-H_\theta(x)\right) = \lambda_\theta(x)\exp\left(-\int_0^x \lambda_\theta(u)\,\mathrm{d}u\right) .$$

The second factor here is equal to $1 - F_\theta(x)$ and indicates the probability that X does not fall into $[0,x)$, while the first factor corresponds to X attaining the value x.
We can extend this interpretation to any counting process. Karr 1991, p. 72 states:

> "We can interpret a point process as a dynamic, uncountable set of independent Bernoulli trials, one for each time t. Taking $\lambda(t)$ as the success probability for the trial at t we conclude that for observation over the time interval $[0,\tau]$ the probability of successes at times $t_1,\ldots,t_{N_\tau}$ is $\prod_{i=1}^{N_\tau}\lambda\left(t_i\right)$."

This heuristic explains the factor

$$\prod_{i=1}^{N_\tau} \lambda_\theta(t_i) \tag{5.4}$$

that appears in the likelihood $L(\theta,N)$. We shall see that the remainder can again be understood as the probability that the "successes" do not occur at any other time:

The probability that T_1 does not fall into $[0, t_1)$ is (compare Lemma 4.23)

$$\mathbb{P}_\theta\left(T_1 > t_1\right) = \exp\left(-H_1^\theta(t_1)\right) = \exp\left(-\int_0^{t_1} \lambda_\theta(u)\,\mathrm{d}u\right), \tag{5.5}$$

while for $i \geq 2$ the conditional probability that T_i does not fall into $[t_{i-1}, t_i)$ given $T_{1:(i-1)} = t_{1:(i-1)}$ equals

$$\begin{aligned}\mathbb{P}_\theta\left(T_i > t_i \,|\, T_{1:(i-1)} = t_{1:(i-1)}\right) &= \exp\left(-H_i^\theta\left(t_i \,|\, t_{1:(i-1)}\right)\right) \\ &= \exp\left(-\int_{t_{i-1}}^{t_i} \lambda_\theta(u)\,\mathrm{d}u\right). \end{aligned} \tag{5.6}$$

Finally, the probability that there is no further point in $[0, \tau]$ after $T_{N_\tau} = t_{N_\tau}$ can be written as

$$\mathbb{P}_\theta\left(T_{N_\tau+1} > \tau \,|\, T_{1:N_\tau} = t_{1:(N_\tau)}\right) = \exp\left(-\int_{t_{N_\tau}}^{\tau} \lambda_\theta(u)\,\mathrm{d}u\right). \tag{5.7}$$

Multiplying the terms from Equations (5.5), (5.6) and (5.7), we obtain:

$$\begin{aligned}&\exp\left(-\int_0^{t_1} \lambda_\theta(u)\,\mathrm{d}u\right) \cdot \prod_{i=2}^{N_\tau} \exp\left(-\int_{t_{i-1}}^{t_i} \lambda_\theta(u)\,\mathrm{d}u\right) \cdot \exp\left(-\int_{t_{N_\tau}}^{\tau} \lambda_\theta(u)\,\mathrm{d}u\right) \\ &\quad = \exp\left(-\int_0^{\tau} \lambda_\theta(u)\,\mathrm{d}u\right), \end{aligned} \tag{5.8}$$

and combining Equations (5.4) and (5.8) then yields the full likelihood

$$L(\theta, N) = \prod_{i=1}^{N_\tau} \lambda_\theta(t_i) \exp\left(-\int_0^{\tau} \lambda_\theta(u)\,\mathrm{d}u\right).$$

Using Equation (5.3), we can determine the likelihood functions in the Basquin load sharing models with damage accumulation (Model $^\times$D) and without damage accumulation (Model B). From these, the associated log-likelihood functions can be easily derived.

Theorem 5.4 ((Log-)Likelihood Functions in the Models $^\times$D and B; cf. Theorem II.1 of Müller and Meyer 2022, p. 3)**.**
In the framework of Definition 5.1 and under assumption (M2)*, for* $j \in \{1, \ldots, J\}$ *we set:*

$$\begin{aligned}\tilde{C}_j &:= \min\left\{N_\tau^{(j)} + 1, C_j\right\}, \qquad & \tilde{T}_i^{(j)} &:= \min\left\{T_i^{(j)}, \tau_j\right\}, \qquad & i &\in \mathbb{N}, \\ \tilde{A}_{j,1} &:= 0, & \tilde{A}_{j,i} &:= \sum_{k=1}^{i-1} B_{j,k}\left(\tilde{T}_k^{(j)} - \tilde{T}_{k-1}^{(j)}\right), & i &\in \mathbb{N} \setminus \{1\}.\end{aligned}$$

The likelihood functions $L_{^\times\mathrm{D}}$ *and* L_B *in the models* $^\times$D *and* B *are then given by*

$$L_{^\times\mathrm{D}}(\theta) = \prod_{j=1}^{J}\left[\prod_{i=1}^{N_\tau^{(j)}} \theta_1 B_{j,i}^{\theta_2}\left(\frac{A_{j,i+1}}{\tau}\right)^{\theta_3} \exp\left(-\sum_{k=1}^{\tilde{C}_j} \frac{\theta_1 B_{j,k}^{\theta_2-1}}{\tau^{\theta_3}\left(\theta_3+1\right)}\left(\tilde{A}_{j,k+1}^{\theta_3+1} - \tilde{A}_{j,k}^{\theta_3+1}\right)\right)\right],$$

$$\textit{and}\quad L_{\mathrm{B}}(\theta)=\prod_{j=1}^{J}\left[\prod_{i=1}^{N_\tau^{(j)}}\theta_1 B_{j,i}^{\theta_2}\exp\left(-\sum_{k=1}^{\tilde{C}_j}\theta_1 B_{j,k}^{\theta_2}\left(\tilde{T}_k^{(j)}-\tilde{T}_{k-1}^{(j)}\right)\right)\right],$$

respectively. The corresponding log-likelihood functions have the following form:

$$\begin{aligned}l_{\times\mathrm{D}}(\theta):=\ln L_{\times\mathrm{D}}(\theta)=\sum_{j=1}^{J}\Bigg[&\sum_{i=1}^{N_\tau^{(j)}}\left(\ln(\theta_1)+\theta_2\ln\left(B_{j,i}\right)+\theta_3\ln\left(\frac{A_{j,i+1}}{\tau}\right)\right)\\&-\frac{\theta_1}{\tau^{\theta_3}\left(\theta_3+1\right)}\sum_{k=1}^{\tilde{C}_j}B_{j,k}^{\theta_2-1}\left(\tilde{A}_{j,k+1}^{\theta_3+1}-\tilde{A}_{j,k}^{\theta_3+1}\right)\Bigg],\end{aligned}$$

$$\textit{and}\quad l_{\mathrm{B}}(\theta):=\ln L_{\mathrm{B}}(\theta)=\sum_{j=1}^{J}\left[\sum_{i=1}^{N_\tau^{(j)}}\left(\ln(\theta_1)+\theta_2\ln(B_{j,i})\right)-\theta_1\sum_{k=1}^{\tilde{C}_j}B_{j,k}^{\theta_2}\left(\tilde{T}_k^{(j)}-\tilde{T}_{k-1}^{(j)}\right)\right].$$

Proof. We only prove the assertions for the model ${}^{\times}$D. For model B, they follow immediately by setting $\theta_3=0$. By definition,

$$^{\times\mathrm{D}}\lambda_\theta^{(j)}(t)=\theta_1 B_j(t)^{\theta_2}A_j(t)^{\theta_3}\cdot\mathbb{1}_{\left\{N_{t^-}^{(j)}<C_j\right\}\cap\left\{t\le\tau_j\right\}}\,. \tag{5.9}$$

If $i\in\{1,\ldots,N_\tau^{(j)}\}$, then $i\le C_j$, because $N_\tau^{(j)}\le C_j$ by construction. Moreover, due to $^{\times\mathrm{D}}\lambda_\theta^{(j)}(t)=0$ for $t>\tau_j$, none of the $T_i^{(j)}$ can fall into $(\tau_j,\tau]$. This implies that $T_i^{(j)}\le\tau_j$, since otherwise $N_\tau^{(j)}<i$. Hence, the indicator function in Equation (5.9) is 1 at $t=T_i^{(j)}$. Because of $B_j(T_i^{(j)})=B_{j,i}$ and $\tau A_j(T_i^{(j)})=A_{j,i+1}$ (see Equation (2.25)), it thus holds:

$$^{\times\mathrm{D}}\lambda_\theta^{(j)}(T_i^{(j)})=\theta_1 B_{j,i}^{\theta_2}\left(\frac{A_{j,i+1}}{\tau}\right)^{\theta_3}.$$

Accordingly, we obtain:

$$\prod_{j=1}^{J}\prod_{i=1}^{N_\tau^{(j)}}{}^{\times\mathrm{D}}\lambda_\theta^{(j)}(T_i^{(j)})=\prod_{j=1}^{J}\prod_{i=1}^{N_\tau^{(j)}}\theta_1 B_{j,i}^{\theta_2}\left(\frac{A_{j,i+1}}{\tau}\right)^{\theta_3}. \tag{5.10}$$

For the remaining part of the likelihood we have to calculate an intensity integral. This can be done piecewise by decomposing the interval $[0,\tau]$ appropriately, that is,

$$[0,\tau]=\left[0,T_1^{(j)}\right)\cup\left[T_1^{(j)},T_2^{(j)}\right)\cup\ldots\cup\left[T_{N_\tau^{(j)}}^{(j)},\tau\right].$$

For any $i\in\{1,\ldots,N_\tau^{(j)}\}$, the intensity function $^{\times\mathrm{D}}\lambda_\theta^{(j)}$ coincides[21] with the conditional hazard function $^{\times\mathrm{D}}h_i^\theta(\,\cdot\mid T_{1:(i-1)}^{(j)},s_j)$ of Lemma 2.24 on $[T_{i-1}^{(j)},T_i^{(j)})$. Hence, we have

$$\int_{T_{i-1}^{(j)}}^{T_i^{(j)}}{}^{\times\mathrm{D}}\lambda_\theta^{(j)}(u)\,\mathrm{d}u=\int_{T_{i-1}^{(j)}}^{T_i^{(j)}}{}^{\times D}h_i^\theta\left(u\,\middle|\,T_{1:(i-1)}^{(j)},s_j\right)\mathrm{d}u={}^{\times D}H_i^\theta\left(T_i^{(j)}\,\middle|\,T_{1:(i-1)}^{(j)},s_j\right)=R_{j,i}^\theta\,,$$

[21] In Section 2.5, we derived the conditional hazard function of the model ${}^{\times}$D for a deterministic censoring scheme. On $\left\{T_i^{(j)}\le\tau_j\right\}\cap\{i\le C_j\}$, however, neither Equation (2.50) nor Equation (2.51) applies, so the formulas of Lemma 2.24 remain valid even for a random censoring scheme.

where $R^\theta_{j,i}$ is the hazard transform of $T_i^{(j)}$ in the model ${}^\times\mathrm{D}$. For the rest of the proof, we regularly write

$$n := N_\tau^{(j)}$$

in order to avoid nested indices in the subsequent calculations. We then observe:

$$\begin{aligned}\int_0^{T_n^{(j)}} {}^\times\mathrm{D}\lambda_\theta^{(j)}(u)\,\mathrm{d}u &= \sum_{i=1}^n \int_{T_{i-1}^{(j)}}^{T_i^{(j)}} {}^\times\mathrm{D}\lambda_\theta^{(j)}(u)\,\mathrm{d}u = \sum_{i=1}^n R^\theta_{j,i} \\ &= \sum_{i=1}^n \frac{\theta_1 B_{j,i}^{\theta_2-1}}{\tau^{\theta_3}\left(\theta_3+1\right)}\left(\tilde{A}_{j,i+1}^{\theta_3+1} - \tilde{A}_{j,i}^{\theta_3+1}\right), \end{aligned} \tag{5.11}$$

because $\tilde{A}_{j,i+1} = A_{j,i+1}$ as long as $T_i^{(j)} \le \tau$. To finish the proof, we need to distinguish whether $N_\tau^{(j)} = C_j$ or $N_\tau^{(j)} < C_j$.
If $n = N_\tau^{(j)} = C_j$, then $\tilde{C}_j = C_j$ and Equation (5.11) has the claimed form, that is,

$$\sum_{k=1}^{N_\tau^{(j)}} \frac{\theta_1 B_{j,k}^{\theta_2-1}}{\tau^{\theta_3}\left(\theta_3+1\right)}\left(\tilde{A}_{j,k+1}^{\theta_3+1} - \tilde{A}_{j,k}^{\theta_3+1}\right) = \sum_{k=1}^{\tilde{C}_j} \frac{\theta_1 B_{j,k}^{\theta_2-1}}{\tau^{\theta_3}\left(\theta_3+1\right)}\left(\tilde{A}_{j,k+1}^{\theta_3+1} - \tilde{A}_{j,k}^{\theta_3+1}\right). \tag{5.12}$$

On the other hand, if $n = N_\tau^{(j)} < C_j$ holds, then $N_{t-}^{(j)} < C_j$ for all $t \in \mathcal{I}$ and we get:

$$\begin{aligned}&\int_{T_n^{(j)}}^{\tau} {}^\times\mathrm{D}\lambda_\theta^{(j)}(u)\,\mathrm{d}u \\ &= \int_{T_n^{(j)}}^{\tau} \theta_1 B_{j,n+1}^{\theta_2}\left[\frac{1}{\tau}\left(B_{j,n+1}\left(u - T_n^{(j)}\right) + A_{j,n+1}\right)\right]^{\theta_3} \cdot \mathbb{1}_{\left\{u\le\tau_j\right\}}\,\mathrm{d}u && \Big|\tau_j \le \tau \\ &= \int_{T_n^{(j)}}^{\tau_j} \theta_1 B_{j,n+1}^{\theta_2}\left[\frac{1}{\tau}\left(B_{j,n+1}\left(u - T_n^{(j)}\right) + A_{j,n+1}\right)\right]^{\theta_3}\,\mathrm{d}u && \Big|\text{cf. Lemma 2.24} \\ &= \frac{\theta_1 B_{j,n+1}^{\theta_2-1}}{\tau^{\theta_3}(\theta_3+1)}\left[\left(B_{j,n+1}\left(\tau_j - T_n^{(j)}\right) + A_{j,n+1}\right)^{\theta_3+1} - A_{j,n+1}^{\theta_3+1}\right] && \Big|T_{n+1}^{(j)} > \tau_j \\ &= \frac{\theta_1 B_{j,n+1}^{\theta_2-1}}{\tau^{\theta_3}(\theta_3+1)}\left[\left(B_{j,n+1}\left(\tilde{T}_{n+1}^{(j)} - T_n^{(j)}\right) + A_{j,n+1}\right)^{\theta_3+1} - A_{j,n+1}^{\theta_3+1}\right] \\ &= \frac{\theta_1 B_{j,n+1}^{\theta_2-1}}{\tau^{\theta_3}(\theta_3+1)}\left(\tilde{A}_{j,n+2}^{\theta_3+1} - \tilde{A}_{j,n+1}^{\theta_3+1}\right), \end{aligned} \tag{5.13}$$

because $T_n^{(j)} = \tilde{T}_n^{(j)}$ and $A_{j,n+1} = \tilde{A}_{j,n+1}$. As $n+1 \le C_j$, we have $\tilde{C}_j = n+1$ and conclude by combining Equations (5.11) and (5.13):

$$\begin{aligned}\int_0^\tau {}^\times\mathrm{D}\lambda_\theta^{(j)}(u)\,\mathrm{d}u &= \int_{T_n^{(j)}}^{\tau} {}^\times\mathrm{D}\lambda_\theta^{(j)}(u)\,\mathrm{d}u + \int_0^{T_n^{(j)}} {}^\times\mathrm{D}\lambda_\theta^{(j)}(u)\,\mathrm{d}u \\ &= \frac{\theta_1 B_{j,n+1}^{\theta_2-1}}{\tau^{\theta_3}(\theta_3+1)}\left(\tilde{A}_{j,n+2}^{\theta_3+1} - \tilde{A}_{j,n+1}^{\theta_3+1}\right) + \sum_{k=1}^n \frac{\theta_1 B_{j,k}^{\theta_2-1}}{\tau^{\theta_3}\left(\theta_3+1\right)}\left(\tilde{A}_{j,k+1}^{\theta_3+1} - \tilde{A}_{j,k}^{\theta_3+1}\right) \\ &= \sum_{k=1}^{\tilde{C}_j} \frac{\theta_1 B_{j,k}^{\theta_2-1}}{\tau^{\theta_3}\left(\theta_3+1\right)}\left(\tilde{A}_{j,k+1}^{\theta_3+1} - \tilde{A}_{j,k}^{\theta_3+1}\right). \end{aligned} \tag{5.14}$$

Hence, Equations (5.12) and (5.14) yield that

$$\prod_{j=1}^{J} \exp\left(-\int_0^\tau {}^{\times\mathrm{D}}\lambda_\theta^{(j)}(u)\,\mathrm{d}u\right) = \prod_{j=1}^{J} \exp\left(-\sum_{k=1}^{\tilde{C}_j} \frac{\theta_1 B_{j,k}^{\theta_2-1}}{\tau^{\theta_3}\left(\theta_3+1\right)}\left(\tilde{A}_{j,k+1}^{\theta_3+1} - \tilde{A}_{j,k}^{\theta_3+1}\right)\right), \quad (5.15)$$

and we obtain the claimed form of the likelihood $L_{\times\mathrm{D}}$ by multiplying Equations (5.10) and (5.15). For the corresponding log-likelihood, taking the natural logarithm immediately leads to the desired result. □

The remainder of this chapter is devoted to some comments on the practical implementation of maximum likelihood estimation in the model ${}^{\times}$D. They address

(i) how the dimensions of the optimization problem can be reduced, and

(ii) how the normalizing constant τ and type I censoring affect the likelihood.

We obtain a maximum likelihood estimator in the model ${}^{\times}$D by maximizing the likelihood $L_{\times\mathrm{D}}$ or, equivalently, the log-likelihood $l_{\times\mathrm{D}}$. For this, we require that there exists $j \in \{1, \ldots, J\}$ with $N_\tau^{(j)} > 0$. Otherwise, the likelihood is strictly decreasing in θ_1, so that there is no permissible maximum likelihood estimate due to $\theta_1 > 0$. This is perfectly plausible, since $N_\tau^{(j)} = 0$ for all $j \in \{1, \ldots, J\}$ means that no points were realized within $[0, \tau]$, which reflects the behavior under the trivial intensity $\lambda \equiv 0$.
If a maximum likelihood estimate can be given on the open domain Θ, then it is located at a critical point of the (log-)likelihood function. At such a critical point, θ_1 can be written as a function of θ_2 and θ_3. We can take advantage of this to reduce the optimization problem from three to two dimensions. To do so, we define two auxiliary functions:

$$G_1\left(\theta_2, \theta_3\right) := \sum_{j=1}^{J} \sum_{i=1}^{N_\tau^{(j)}} \left(\theta_2 \ln\left(B_{j,i}\right) + \theta_3 \ln\left(\frac{A_{j,i+1}}{\tau}\right)\right),$$

$$G_2\left(\theta_2, \theta_3\right) := \frac{1}{\tau^{\theta_3}\left(\theta_3+1\right)} \sum_{j=1}^{J} \sum_{k=1}^{\tilde{C}_j} B_{j,k}^{\theta_2-1}\left(\tilde{A}_{j,k+1}^{\theta_3+1} - \tilde{A}_{j,k}^{\theta_3+1}\right).$$

Then, the log-likelihood $l_{\times\mathrm{D}}$ from Theorem 5.4 can be written as

$$l_{\times\mathrm{D}}(\theta) = \ln(\theta_1) \sum_{j=1}^{J} N_\tau^{(j)} + G_1\left(\theta_2, \theta_3\right) - \theta_1 G_2\left(\theta_2, \theta_3\right). \quad (5.16)$$

For the parameter $\theta = (\theta_1, \theta_2, \theta_3)^\top$ to be a critical point of $l_{\times\mathrm{D}}$, it has to satisfy

$$\frac{\mathrm{d}}{\mathrm{d}\theta} l_{\times\mathrm{D}}(\theta) \stackrel{!}{=} 0,$$

which means that the gradient of $l_{\times\mathrm{D}}$ at θ is equal to 0. In particular, for the partial derivative with respect to θ_1 we get

$$0 \stackrel{!}{=} \frac{\partial}{\partial\theta_1} l_{\times\mathrm{D}}(\theta) = \frac{1}{\theta_1} \sum_{j=1}^{J} N_\tau^{(j)} - G_2\left(\theta_2, \theta_3\right), \quad (5.17)$$

and solving Equation (5.17) for θ_1 yields:

$$\theta_1 \stackrel{!}{=} \frac{\sum_{j=1}^{J} N_\tau^{(j)}}{G_2\left(\theta_2, \theta_3\right)} =: \hat{\theta}_1\left(\theta_2, \theta_3\right) . \tag{5.18}$$

Because $\hat{\theta}_1\left(\theta_2, \theta_3\right) > 0$ as long as $N_\tau^{(j)} > 0$ for some $j \in \{1, \ldots, J\}$, Equation (5.18) provides a valid estimate for θ_1. Substituting this identity into Equation (5.16) leads to the following function of two variables, namely θ_2 and θ_3:

$$l_{\times\mathrm{D}}\Big(\left(\hat{\theta}_1\left(\theta_2, \theta_3\right), \theta_2, \theta_3\right)^\top\Big) = \left[\ln\left(\frac{\sum_{j=1}^{J} N_\tau^{(j)}}{G_2\left(\theta_2, \theta_3\right)}\right) - 1\right] \sum_{j=1}^{J} N_\tau^{(j)} + G_1\left(\theta_2, \theta_3\right) . \tag{5.19}$$

By setting $\theta_3 = 0$, the log-likelihood in the model B can then be similarly expressed as a function of only θ_2 via

$$l_{\mathrm{B}}\Big(\left(\hat{\theta}_1\left(\theta_2, 0\right), \theta_2\right)^\top\Big) = l_{\times\mathrm{D}}\Big(\left(\hat{\theta}_1\left(\theta_2, 0\right), \theta_2, 0\right)^\top\Big) . \tag{5.20}$$

Upon further observation of Equations (5.19) and (5.20), we can also conclude that the maximum likelihood estimator for θ does not directly depend on the normalizing constant τ, neither in model $^\times$D nor B. While the choice of τ affects the random covariates τ_j due to $\tau_j \leq \tau$, choosing any larger $\tilde{\tau} > \tau$ has no further influence on the log-likelihood. To see this, note that

$$N_\tau^{(j)} = N_{\tilde{\tau}}^{(j)} \quad \text{for all } \tau_j \leq \tau < \tilde{\tau}. \tag{5.21}$$

Moreover, we can show that τ cancels out in Equation (5.19). We write:

$$\begin{aligned} G_1(\theta_2, \theta_3) &= \sum_{j=1}^{J} \sum_{i=1}^{N_\tau^{(j)}} \left(\theta_2 \ln\left(B_{j,i}\right) + \theta_3 \ln\left(\frac{A_{j,i+1}}{\tau}\right)\right) \\ &= \sum_{j=1}^{J} \sum_{i=1}^{N_\tau^{(j)}} \left(\theta_2 \ln\left(B_{j,i}\right) + \theta_3 \ln\left(A_{j,i+1}\right) - \theta_3 \ln(\tau)\right) \\ &= \underbrace{\sum_{j=1}^{J} \sum_{i=1}^{N_\tau^{(j)}} \left(\theta_2 \ln\left(B_{j,i}\right) + \theta_3 \ln\left(A_{j,i+1}\right)\right)}_{\text{does not depend on } \tau \text{ due to Equation (5.21)}} - \theta_3 \ln(\tau) \sum_{j=1}^{J} N_\tau^{(j)} , \end{aligned} \tag{5.22}$$

and

$$\begin{aligned} -\ln\left(G_2(\theta_2, \theta_3)\right) &= -\ln\left(\frac{1}{\tau^{\theta_3}\left(\theta_3 + 1\right)} \sum_{j=1}^{J} \sum_{k=1}^{\tilde{C}_j} B_{j,k}^{\theta_2 - 1}\left(\tilde{A}_{j,k+1}^{\theta_3+1} - \tilde{A}_{j,k}^{\theta_3+1}\right)\right) \\ &= -\left[\ln\left(\frac{1}{\theta_3 + 1} \sum_{j=1}^{J} \sum_{k=1}^{\tilde{C}_j} B_{j,k}^{\theta_2 - 1}\left(\tilde{A}_{j,k+1}^{\theta_3+1} - \tilde{A}_{j,k}^{\theta_3+1}\right)\right) - \ln\left(\tau^{\theta_3}\right)\right] \\ &= \theta_3 \ln(\tau) - \underbrace{\ln\left(\frac{1}{\theta_3 + 1} \sum_{j=1}^{J} \sum_{k=1}^{\tilde{C}_j} B_{j,k}^{\theta_2 - 1}\left(\tilde{A}_{j,k+1}^{\theta_3+1} - \tilde{A}_{j,k}^{\theta_3+1}\right)\right)}_{\text{does not depend on } \tau} . \end{aligned} \tag{5.23}$$

Since $-\ln(G_2(\theta_2, \theta_3))$ is multiplied by $\sum_{j=1}^{J} N_\tau^{(j)}$ in Equation (5.19), both Equation (5.22) and Equation (5.23) provide the term

$$\theta_3 \ln(\tau) \sum_{j=1}^{J} N_\tau^{(j)} \,,$$

which is then canceled out due to the opposite signs. Because the remaining terms no longer depend on τ due to Equation (5.21), this shows that the maximum likelihood estimator is not affected by the specific choice of τ as long as τ is sufficiently large. Consequently, we may allow $\tau \to \infty$ and find ourselves in the situation of an unbounded observation horizon. If we further let $\mathbb{P}^{\tau_j} = \delta_\tau$ (i.e., $\tau_j = \tau$ with probability one), we can also remove the random type I censoring altogether. In that case, $N_\tau^{(j)} \to C_j$ with probability one as $\tau \to \infty$, and the likelihood function $L_{\times\mathrm{D}}$ can be reduced to

$$\begin{aligned} L_{\times\mathrm{D}}(\theta) &= \prod_{j=1}^{J} \left[\prod_{i=1}^{C_j} {}^{\times\mathrm{D}}\lambda_\theta^{(j)} \left(t_i^{(j)} \right) \exp\left(-\int_0^{T_{C_j}^{(j)}} {}^{\times\mathrm{D}}\lambda_\theta^{(j)}(u)\,\mathrm{d}u \right) \right] \\ &= \prod_{j=1}^{J} \left[\prod_{i=1}^{C_j} \theta_1 B_{j,i}^{\theta_2} A_{j,i+1}^{\theta_3} \exp\left(-\sum_{k=1}^{C_j} \frac{\theta_1 B_{j,k}^{\theta_2 - 1}}{\theta_3 + 1} \left(A_{j,k+1}^{\theta_3+1} - A_{j,k}^{\theta_3+1} \right) \right) \right] \,, \end{aligned} \tag{5.24}$$

where the second layer of random censoring via $C_j \le I < \infty$ ensures that we still only deal with finitely many observations. Note that in the absence of type I censoring, $\mathbb{P}_{\theta^*}(T_i^{(j)} > \tau) > 0$ holds regardless of the value of $\tau < \infty$ if $\mathbb{P}_{\theta^*}(C_j \ge i) > 0$ for $i \in \mathbb{N}$.

The practical implications of this are twofold: On the one hand, an unbounded observation period cannot be implemented, but on the other hand, τ can also never be identified as "sufficiently large" in advance. Technically, this means that calculating the formula of Equation (5.24) is often not feasible in applications. However, if τ has not been specified before conducting the experiment, one can deliberately choose τ to be arbitrarily large as soon as every observable event has occurred, that is, $N_\tau^{(j)} = C_j$ for all $j \in \{1, \ldots, J\}$.

6. Simulation Studies and Robustness of the Proposed Methods

In the last major chapter of this thesis, we compare the statistical methods of the previous chapters in a simulation study. To this end, in Section 6.1 we first describe an algorithm for the simulation of point processes whose cumulative intensity can be specified in terms of invertible cumulative conditional hazard functions, as is the case for our main models B and $^{\times}$D. By simulating processes with the given intensity ${}^{\times\mathrm{D}}\lambda_{\theta^*}$ of the Basquin load sharing model with damage accumulation, we can evaluate the competing methods in terms of coverage rate and size of their confidence regions for the true parameter θ^*. Both the construction and comparison of these confidence regions can be found in Section 6.2. We then conduct hypothesis tests in Section 6.3 to decide whether the damage accumulation term that extends model B to model $^{\times}$D is statistically significant. In the final Section 6.4, we assess the robustness of our methods by applying them to contaminated data.

All simulations and computations were implemented by the author of this thesis and executed in `R` (R Core Team 2023). The package `xtable` (Dahl et al. 2019) was helpful in exporting tables to LaTeX. In addition, the packages `ggplot2` (Wickham 2016) and `tikzDevice` (Sharpsteen and Bracken 2023) were used for most of the visualizations.

6.1. Simulation of a Point Process With Given Cumulative Intensity

In order to run a simulation study, it must be possible to generate random variables or processes with certain properties, like a predefined distribution. In our case, the distribution of the counting processes is determined by the underlying conditional intensity function, given with respect to an intrinsic filtration. We have seen as early as Equation (2.5) that such a conditional intensity function can be expressed through a family of conditional hazard functions. These conditional hazard functions in turn define the conditional cumulative hazard functions by integration, which laid the foundation for Section 2.5 on the hazard transformation of a point process. It is precisely this transformation that allows us to easily simulate counting processes with a specific cumulative intensity.

Algorithm 6.1 Simulation of a point process with given cumulative conditional hazard functions (*cchf*) $H_i(\,\cdot \mid t_{1:(i-1)}, x)$ by the inversion method, cf. Daley and Vere-Jones 2003, p. 260. Requires that the inverse cchf can be stated explicitly.

Input:

$n \in \mathbb{N}$	number of points to simulate,
$\mathbb{P}^X$	distribution of the random covariates,
$t_0 \in \mathbb{R}_+$	value of T_0, defaults to $t_0 = 0$,
$H_i^{-1}(\,\cdot \mid t_{i-1}, \dots, t_0, x)$	inverse cchf, $i = 1, \dots, n$.

Output:

$t_{1:n} \in \mathbb{R}_+^n$	vector with realizations of the points $T_1, \dots, T_n$.

draw sample x of the covariate distribution $\mathbb{P}^X$
draw i.i.d. samples $y_1, \dots, y_n$ of the unit exponential distribution $\mathcal{E}(1)$
for $i = 1, \dots, n$ **do**
  $t_i \leftarrow H_i^{-1}(y_i \mid t_{i-1}, \dots, t_0, x)$
end for

The idea is to combine Theorem 2.20 with Equation (2.44): If the hazard transformation takes any process to a standard Poisson process, then the inverse hazard transformation can generate arbitrary processes from a standard Poisson process. In the case that the inverse cumulative conditional hazard functions $H_i^{-1}(\,\cdot \mid t_{i-1}, \dots, t_0, x)$ can be stated explicitly, this simple procedure is explained in Algorithm 6.1. In this algorithm, we have not taken into account that due to random censoring schemes, the inverses of the cumulative conditional hazard functions often cannot be given at all. This complication is closely related to the compatibility of hazard transformation and censoring schemes, which we addressed in Remark 2.22. Unlike previously, however, in a simulation study this proves to be unproblematic, because the simulated process does not differentiate between whether the censoring covariates C_j and τ_j are known in advance or whether the censoring takes place retrospectively. In practice, this allows that we first simulate a full realization

$$T_{1:I}^{(j)}(\omega) = (t_1^{(j)}, \dots, t_I^{(j)})$$

of the jth point process $T^{(j)}$ using Algorithm 6.1. We then draw samples $C_j(\omega)$ and $\tau_j(\omega)$ of the random censoring covariates. The number of observations with $t_i^{(j)} \le \tau_j(\omega)$ is given by

$$N_{\tau_j(\omega)}^{(j)}(\omega) = \max\left\{i \,:\, t_i^{(j)} \le \tau_j(\omega)\right\}.$$

This means that any point with an index less than or equal to

$$\tilde{C}_j := \min\left\{C_j(\omega), N_{\tau_j(\omega)}^{(j)}(\omega)\right\}$$

is not affected by censoring, so as an actual realization of the jth point process $T^{(j)}$ we obtain

$$\left(t_1^{(j)}, \dots, t_{\tilde{C}_j}^{(j)}\right).$$

Note that we have implicitly used here that in our models each point process consists of at most I points. If this cannot be assumed, the C_j should be sampled in advance so that $I = \max_{j=1,\dots,J} C_j(\omega)$ can be chosen.

6.2. Comparison of Confidence Sets for the True Parameter

Each of the methods presented in this thesis lends itself to the construction of confidence regions for the true parameter θ^* of a parametric intensity-based point process model. Within the next three paragraphs, we discuss the following approaches one after another:

(i) Wald-type confidence sets based on the minimum distance estimator of Chapter 3,

(ii) Confidence sets obtained from the consistent 3-sign depth test of Chapter 4,

(iii) Confidence sets constructed from the likelihood-ratio given in Chapter 5.

In order to compare these confidence regions, we need to ensure that their frameworks are compatible. The chosen parameters in the model $^{\times}$D as well as the values of the covariates used for the simulation study are given in Table 2 on the next page.
Since we have $\tau_j = 1$ and $C_j = 10$ for all $j \in \{1, \dots, J\}$, we can assume $\mathbb{P}^{\tau_0} = \delta_{\{1\}}$ and $\mathbb{P}^{C_0} = \delta_{\{10\}}$, which fits the frameworks of the minimum distance estimator (Definition 3.2) and the maximum likelihood estimator (Definition 5.1). Moreover, C_j can be regarded as deterministic, while τ_j is chosen large enough that no type I censoring occurs during

model $^{\times}$D	model parameters			random covariates			repetitions
param./covariate	θ^*	I	τ	τ_j	C_j	s_j	J
chosen value(s)	$(10^{-4},3,1)^\top$	25	1	1	10	$80,120,200$	$9,18,30,90,180$

Table 2: The selected model parameters and values of covariates for the simulation study.

the study, both of which are requirements for the hazard transformation framework of Definition 2.23. Finally, the three initial stress levels $80, 120, 200$ are repeated collectively, which is why the number J of simulated processes is always a multiple of three. Within the framework of Definition 4.11, this corresponds to $L = 3$ classes that are successively repeated, see also the design in Table 1. In the other frameworks, we can take the s_j as i.i.d. realizations of the discrete uniform distribution with support $\{80, 120, 200\}$, that is,

$$\mathbb{P}^{s_0} = \frac{1}{3}\left(\delta_{\{80\}} + \delta_{\{120\}} + \delta_{\{200\}}\right) .$$

Note that the true parameter value $\theta^* = (10^{-4}, 3, 1)^\top$ is vaguely based on the estimates obtained by Müller and Meyer 2022 for a real data experiment. We adjusted the scaling parameter θ_1^* so that the range of observations better matches our choice of τ. The parameter θ_3^* was also increased to achieve a more pronounced damage accumulation effect in the simulation studies.

(i) Confidence Sets Based on the Minimum Distance Estimator

As our first confidence set for the true parameter θ^*, we consider Wald-type confidence regions constructed from the asymptotic distribution of the minimum distance estimator. Recall that by virtue of Corollary 3.26 we have (as usual, d is the dimension of the parameter space Θ)

$$\sqrt{J}\left(\hat{\theta}_J - \theta^*\right) \xrightarrow{\mathrm{d}} \mathcal{N}_d\left(0, \Phi_0(\theta^*)^{-1}\Sigma(\theta^*)\Phi_0(\theta^*)^{-1}\right) \quad (J \to \infty) .$$

To shorten the notation, for $\theta \in \mathrm{B}_\varepsilon(\theta^*)$ with sufficiently small[22] $\varepsilon > 0$ we may write

$$\Sigma_\Phi(\theta) := \Phi_0(\theta)^{-1}\Sigma(\theta)\Phi_0(\theta)^{-1} .$$

For $0 < \alpha < 1$, an asymptotic Wald-type $(1-\alpha)$-confidence region for θ^* is then given by

$$\mathcal{C}_{J,1-\alpha}^{(\mathrm{wald})} := \left\{\theta \in \Theta \,:\, J(\hat{\theta}_J - \theta)^\top \Sigma_\Phi(\theta^*)^{-1}(\hat{\theta}_J - \theta) \le \chi^2_{d,1-\alpha}\right\},$$

where $\chi^2_{d,1-\alpha}$ is the $(1-\alpha)$-quantile of the χ^2-distribution with d degrees of freedom. Since the covariance matrix $\Sigma_\Phi(\theta^*)$ is unknown, it must be estimated. Kopperschmidt and Stute 2013 suggest to replace the unknown standardizing matrix $\Phi_0(\theta^*)$ with $\Phi_J(\hat{\theta}_J)$ (compare Equations (3.23) and (3.24)) and $\Sigma(\theta^*)$ with an appropriate sample analogue. While the results from Section 3.3 on the asymptotics of the minimum distance estimator (see in particular Proposition 3.13) show that this approach is valid in theory, a closer look at the

[22]We must ensure that $\Phi_0(\theta)$ is invertible, which is not guaranteed for all of Θ, but in a sufficiently small neighborhood of θ^*. Compare Lemma 3.18 from Chapter 3 for details.

formulas involved reveals the practical difficulties. As emphasized by Kopperschmidt 2005, p. 159, the analytical calculation of $\Phi_J(\hat{\theta}_J)$ is possible in principle, but highly demanding already for simpler models such as the model B. The same applies to a sample analogue of the asymptotic covariance matrix $\Sigma(\theta^*)$, to such an extent that Kopperschmidt developed further approximations of the occurring integrands (cf. Kopperschmidt 2005, pp. 160–161). He then resorts to numerical integration and (partial) differentiation in order to estimate $\Sigma_\Phi(\theta^*)$. Each step in the calculation therefore adds another layer of approximation. In our adaptation of Kopperschmidt's procedure, these accumulating approximations ultimately led to unusable results.

We consequently need an alternative method for estimating the covariance matrix $\Sigma_\Phi(\theta^*)$. Our approach relies on Proposition 3.13 and Theorem 3.25, which justify the approximation

$$\Sigma_\Phi(\hat{\theta}_J) \approx \Sigma_\Phi(\theta^*) \,. \tag{6.1}$$

Note that we replace the true parameter θ^* with its minimum distance estimation $\hat{\theta}_J$, but otherwise do not switch to the sample analogues from before. We can then simulate $J_{\text{sim}} \in \mathbb{N}$ new observations based on the "true" parameter $\hat{\theta}_J$ and calculate the minimum distance estimator for $\hat{\theta}_J$, which we denote here by $\tilde{\theta}_{J_{\text{sim}}}$. Then, approximately,

$$\sqrt{J_{\text{sim}}}\left(\tilde{\theta}_{J_{\text{sim}}} - \hat{\theta}_J\right) \sim \mathcal{N}_d\left(0, \Sigma_\Phi(\hat{\theta}_J)\right) \,. \tag{6.2}$$

We do not have to choose $J_{\text{sim}} = J$, but we will usually do so in order to reuse the realized covariates for the new simulations, as their distribution is generally unknown in practice. For small J, however, it is advisable to choose a multiple $J_{\text{sim}} = J\eta$ ($\eta \in \mathbb{N}$) of J and repeat the realized covariates accordingly, because otherwise the distribution assumption from Equation (6.2) may not apply. In order to estimate the covariance matrix $\Sigma_\Phi(\hat{\theta}_J)$, we simulate $n_{\text{rep}} \in \mathbb{N}$ realizations of the minimum distance estimator $\tilde{\theta}_{J_{\text{sim}}}$ for $\hat{\theta}_J$, denoted by

$$\tilde{\theta}^{(n)}_{J_{\text{sim}}}\,, \quad n = 1, \dots, n_{\text{rep}} \,.$$

We can employ a standard estimator such as the sample covariance to estimate $\Sigma_\Phi(\hat{\theta}_J)$. This estimation then also serves as an approximation of $\Sigma_\Phi(\theta^*)$ due to Equation (6.1). Since under the distribution assumption from Equation (6.2) the expected value is known, we can alternatively use the following estimator:

$$\hat{\Sigma}_\Phi(\hat{\theta}_J) := \frac{J_{\text{sim}}}{n_{\text{rep}}} \sum_{n=1}^{n_{\text{rep}}} (\tilde{\theta}_{J_{\text{sim}}} - \hat{\theta}_J)^\top (\tilde{\theta}_{J_{\text{sim}}} - \hat{\theta}_J) \,. \tag{6.3}$$

The Algorithm 6.2 summarizes this procedure, enabling us to estimate the covariance matrix $\Sigma_\Phi(\hat{\theta}_J)$ with any desired accuracy. An approximative $(1-\alpha)$-confidence region for θ^* is then given by

$$\mathcal{C}^{(\text{dist})}_{J,1-\alpha} := \left\{\theta \in \Theta \,:\, J(\hat{\theta}_J - \theta)^\top \hat{\Sigma}_\Phi(\hat{\theta}_J)^{-1}(\hat{\theta}_J - \theta) \le \chi^2_{d,1-\alpha}\right\} . \tag{6.4}$$

Similar to Kopperschmidt's approach, however, our method also reaches its computational limits when we increase both J_{sim} and n_{rep}. This becomes evident in a simulation study where we aim to generate a large number of confidence regions. If the covariance matrix $\Sigma_\Phi(\hat{\theta}_J)$ for each repetition should be estimated with high accuracy, such a study is

Algorithm 6.2 Estimation of the unknown covariance matrix $\Sigma_\Phi(\theta^*)$.

Input:

$\hat{\theta}_J$	previous minimum distant estimation for the true parameter θ^*,
J_{sim}	number of simulated processes to compute the MDE $\tilde{\theta}_{J_{\text{sim}}}$ for $\hat{\theta}_J$,
n_{rep}	number of realizations of the MDE $\tilde{\theta}_{J_{\text{sim}}}$ from which $\Sigma_\Phi(\hat{\theta}_J)$ is estimated,
...	further inputs required by Algorithm 6.1.

Output:

$\hat{\Sigma}_\Phi(\hat{\theta}_J)$	estimation of the covariance matrix $\Sigma_\Phi(\hat{\theta}_J) \approx \Sigma_\Phi(\theta^*)$

```
for n = 1, ..., n_rep do
  for j = 1, ..., J_sim do
    simulate realization t^(j) of the point process T^(j) via Algorithm 6.1
    // (random) censoring may take place here if required
  end for
  calculate MDE θ~^(n)_{J_sim} from t^(1), ..., t^(J_sim)
end for
compute Σ^_Φ(θ^_J) from θ~^(1)_{J_sim}, ..., θ~^(n_rep)_{J_sim} via Equation (6.3)
```

computationally not feasible. If, on the other hand, we choose an insufficient J_{sim} or n_{rep}, then the inaccurate estimation of $\Sigma_\Phi(\hat{\theta}_J)$ and thus $\Sigma_\Phi(\theta^*)$ results in poor confidence regions. In the scope of our simulation study, we decide on the following compromise:

- For the visualizations in this paragraph, we calculate the confidence region $\mathcal{C}^{(\text{dist})}_{J,1-\alpha}$ from Equation (6.4) with $J_{\text{sim}} = J$ and $n_{\text{rep}} = 3000$.
- For the comparison of the confidence regions from paragraphs (i), (ii) and (iii), we estimate $\Sigma_\Phi(\theta^*)$ in advance. To do this, we exploit that in a simulation study the true parameter is known and obtain $\hat{\Sigma}_\Phi(\theta^*)$ from Algorithm 6.2 by passing θ^* instead of $\hat{\theta}_J$ as the input. We then replace each instance of $\hat{\Sigma}_\Phi(\hat{\theta}_J)$ with $\hat{\Sigma}_\Phi(\theta^*)$.

Even if not applicable in practice, this compromise is reasonable because of Equation (6.1). The runtime for determining the confidence region is mainly driven by the duration needed to calculate the minimum distance estimations. A single confidence interval requires

$$\underbrace{1}_{\text{for the actual MDE } \hat{\theta}_J} + \underbrace{n_{\text{rep}}}_{\text{for the covariance matrix estimation } \hat{\Sigma}_\Phi(\hat{\theta}_J)}$$

calculations of the MDE. So if n_{conf} denotes the number of realized confidence regions, the above compromise reduces the number of MDE calculations from

$$n_{\text{conf}} \cdot (1 + n_{\text{rep}}) = n_{\text{conf}} + n_{\text{conf}} \cdot n_{\text{rep}}$$

to just

$$\underbrace{n_{\text{rep}}}_{\text{for preliminary covariance matrix estimation } \hat{\Sigma}_\Phi(\theta^*)} + \underbrace{n_{\text{conf}}}_{\text{for the MDE corresponding to each confidence region}}$$

and therefore saves $(n_{\text{conf}} - 1) \cdot n_{\text{rep}}$ MDE calculations, which represents the majority of the computational effort. The running time of each evaluation of the MDE should nevertheless be reduced as much as possible. Similar to the maximum likelihood estimator, the minimum distance estimator is obtained by solving an optimization problem whose dimension can be reduced from three to two. We revisit Lemma 3.10, where we have seen that the gradient of the Cramér-von Mises distance disappears at the MDE $\hat{\theta}_J$, that is,

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}\theta}\left\|\overline{N}^{(J)} - \overline{\Lambda}_\theta^{(J)}\right\|_{\overline{N}^{(J)}}^2\Bigg|_{\theta=\hat{\theta}_J} = 0\,.$$

The counting processes $N^{(1)}, \ldots, N^{(J)}$ almost surely have no common discontinuities according to Lemma A.37. By Equation (3.4), we can therefore write the Cramér-von Mises distance as

$$\left\|\overline{N}^{(J)} - \overline{\Lambda}_\theta^{(J)}\right\|_{\overline{N}^{(J)}}^2 = \frac{1}{J}\sum_{t\in\mathcal{J}_{\overline{N}^{(J)}}}\left(\overline{N}_t^{(J)} - \overline{\Lambda}_\theta^{(J)}(t)\right)^2.$$

The partial derivative with respect to θ_1 is given by

$$\frac{\partial}{\partial\theta_1}\left\|\overline{N}^{(J)} - \overline{\Lambda}_\theta^{(J)}\right\|_{\overline{N}^{(J)}}^2 = -\frac{2}{J}\sum_{t\in\mathcal{J}_{\overline{N}^{(J)}}}\left[\left(\overline{N}_t^{(J)} - \overline{\Lambda}_\theta^{(J)}(t)\right)\frac{\partial}{\partial\theta_1}\overline{\Lambda}_\theta^{(J)}(t)\right].$$

For the gradient of the Cramér-von Mises distance to vanish, all partial derivatives must be equal to 0. Hence,

$$0 \overset{!}{=} \frac{1}{2}\frac{\partial}{\partial\theta_1}\left\|\overline{N}^{(J)} - \overline{\Lambda}_\theta^{(J)}\right\|_{\overline{N}^{(J)}}^2 \quad\Longleftrightarrow\quad 0 \overset{!}{=} \sum_{t\in\mathcal{J}_{\overline{N}^{(J)}}}\left[\left(\overline{N}_t^{(J)} - \overline{\Lambda}_\theta^{(J)}(t)\right)\frac{\partial}{\partial\theta_1}\overline{\Lambda}_\theta^{(J)}(t)\right]. \tag{6.5}$$

For the model ${}^\times\mathrm{D}$, we can immediately see that the following relationship applies:

$${}^{\times\mathrm{D}}\Lambda_{(\theta_1,\theta_2,\theta_3)^\top} \equiv \theta_1\,{}^{\times\mathrm{D}}\Lambda_{(1,\theta_2,\theta_3)^\top}\,. \tag{6.6}$$

Substituting Equation (6.6) into Equation (6.5) and solving for θ_1 yields (we omit the model indicator ${}^\times\mathrm{D}$ for readability):

$$\begin{aligned}
&& 0 &\overset{!}{=} \sum_{t\in\mathcal{J}_{\overline{N}^{(J)}}}\left[\left(\overline{N}_t^{(J)} - \overline{\Lambda}_\theta^{(J)}(t)\right)\frac{\partial}{\partial\theta_1}\overline{\Lambda}_\theta^{(J)}(t)\right]\\
&& &= \sum_{t\in\mathcal{J}_{\overline{N}^{(J)}}}\left[\left(\overline{N}_t^{(J)} - \theta_1\overline{\Lambda}_{(1,\theta_2,\theta_3)^\top}^{(J)}(t)\right)\cdot\overline{\Lambda}_{(1,\theta_2,\theta_3)^\top}^{(J)}(t)\right]\\
\Longleftrightarrow && \theta_1\sum_{t\in\mathcal{J}_{\overline{N}^{(J)}}}\left(\overline{\Lambda}_{(1,\theta_2,\theta_3)^\top}^{(J)}(t)\right)^2 &\overset{!}{=} \sum_{t\in\mathcal{J}_{\overline{N}^{(J)}}}\left(\overline{N}_t^{(J)}\cdot\overline{\Lambda}_{(1,\theta_2,\theta_3)^\top}^{(J)}(t)\right)\\
\Longleftrightarrow && \theta_1 &\overset{!}{=} \frac{\sum_{t\in\mathcal{J}_{\overline{N}^{(J)}}}\left(\overline{N}_t^{(J)}\cdot\overline{\Lambda}_{(1,\theta_2,\theta_3)^\top}^{(J)}(t)\right)}{\sum_{t\in\mathcal{J}_{\overline{N}^{(J)}}}\left(\overline{\Lambda}_{(1,\theta_2,\theta_3)^\top}^{(J)}(t)\right)^2} =: \hat{\theta}_1(\theta_2,\theta_3)\,.
\end{aligned}$$

Analogous to the (log-)likelihood from Equation (5.19), the Cramér-von Mises distance can then be expressed as a function of the two parameters θ_2 and θ_3 if we replace θ_1 with $\hat{\theta}_1(\theta_2,\theta_3)$. This simplification significantly accelerates numerical optimization. Moreover, numerical methods occasionally fail to find the minimum of the Cramér-von Mises distance in three dimensions, even if the initial value is set to the true parameter. As this never occurred for the two-dimensional optimization in any of the simulation runs, the reduction of the dimension also enhances the overall reliability of the method. However, we do not want to conceal the fact that the minimum distance estimator still occasionally deviates drastically from the true value, especially for smaller $J \leq 9$. This is the sole reason why the case $J = 3$ is not considered in some aspects of the simulation study, because the exceptionally high variance of the minimum distance estimate for θ_1 leads to numerically singular covariance matrices, rendering the computation of Wald-type confidence regions impossible. Note that this is not a matter of scale, as the results are similar at $\theta^* = (1, 3, 1)$, for example. We will see shortly that this complication is not unique to the minimum distance estimator, but is an inherent property of the model $^\times$D.

We are now able to plot a confidence region $\mathcal{C}_{J,1-\alpha}^{(\text{dist})}$ for the true parameter θ^* based on the counting process realizations obtained from Algorithm 6.1. All confidence regions visualized in this dissertation are generated from the same seed to ensure comparability. For each of the initial stress levels $80, 120, 200$, we simulate 60 counting processes with the parameters and covariates given in Table 2 for a total of $J = 180$ counting process realizations. We then obtain the realizations in the cases $J < 180$ by looking at subsets of this data. This means, for example, that all the realizations used to calculate the confidence regions $\mathcal{C}_{30,1-\alpha}^{(\text{dist})}$ (i.e., $J = 30$) are also included in the confidence regions $\mathcal{C}_{90,1-\alpha}^{(\text{dist})}$ and $\mathcal{C}_{180,1-\alpha}^{(\text{dist})}$, which in practice equates to adding further data from follow-up experiments. As pointed out earlier, we choose $J_{\text{sim}} = J$ and $n_{\text{rep}} = 3000$ for the estimation of $\Sigma_\Phi(\hat{\theta}_J)$.

Since the parameter space for the model $^\times$D is 3-dimensional, the confidence regions shown here are difficult to visualize as subsets of $\mathbb{R}^3$. So instead of plotting a confidence region itself, we show the intersections of the confidence region with suitable hyperplanes:

$$\begin{aligned}
&\text{In } \theta_1\text{-}\theta_2\text{-direction:} && \mathcal{C}_{J,1-\alpha} \cap \{\theta \in \Theta \,:\, \theta_3 = \theta_3^*\}\ ,\\
&\text{In } \theta_1\text{-}\theta_3\text{-direction:} && \mathcal{C}_{J,1-\alpha} \cap \{\theta \in \Theta \,:\, \theta_2 = \theta_2^*\}\ ,\\
&\text{In } \theta_2\text{-}\theta_3\text{-direction:} && \mathcal{C}_{J,1-\alpha} \cap \{\theta \in \Theta \,:\, \theta_1 = \theta_1^*\}\ .
\end{aligned}$$

We have omitted the indicator (dist) here, as this procedure is universal for the other confidence regions, too. The hyperplanes are sometimes abbreviated as $\pi_i^{-1}(\{\theta_i^*\})$ or just $\{\theta_i = \theta_i^*\}$. Two points of criticism should be mentioned in advance:
First, a single seed does not provide a representative image of the confidence region, but it gives us an idea of its overall shape. Second, this approach generally does not provide two-dimensional $(1-\alpha)$-confidence regions, which is why the following figures should be treated with caution.
All plots are displayed in Appendix C in Figures 22 (for the θ_1-θ_2-direction), 23 (for the θ_1-θ_3-direction) and 24 (for the θ_2-θ_3-direction). Since the confidence regions $\mathcal{C}_{J,1-\alpha}^{(\text{dist})}$ are ellipsoids by construction, the intersection with a hyperplane is always an ellipse. We can see this directly from the plots or at least assume it due to the size of the ellipse. Figure 5 shows the panel corresponding to $J = 180$ from Figure 22. The large variance of the estimate for θ_1 elongates the ellipse and gives the impression of a confidence "band". The same pattern can be observed for the intersection with the hyperplane $\{\theta_2 = \theta_2^*\}$ (see Figure 23), while for the intersection with the hyperplane $\{\theta_1 = \theta_1^*\}$ the ellipses are clearly recognizable (see Figure 24). We might be tempted to assume that the large deviations

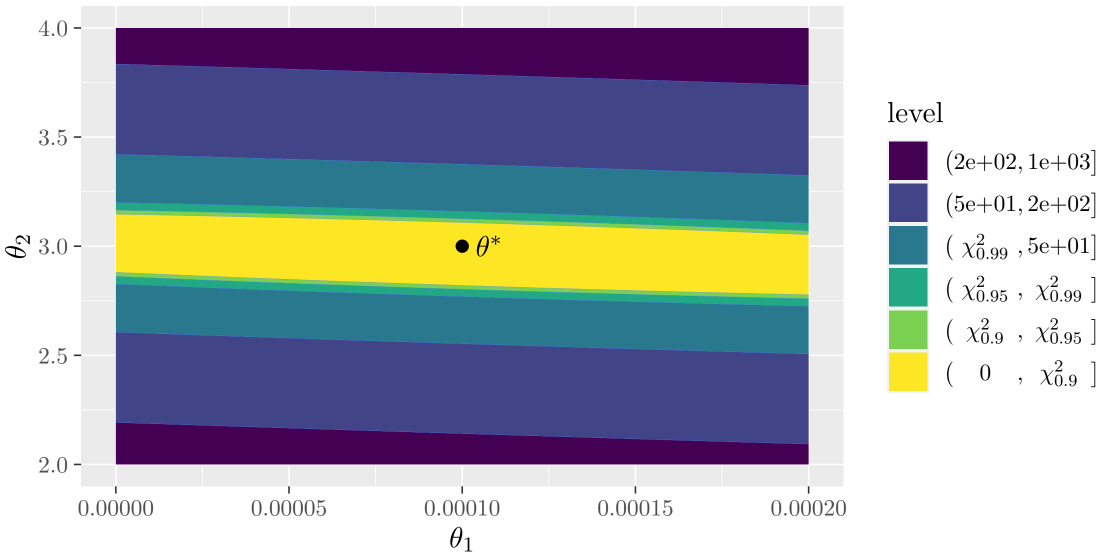


Figure 5: Visualization of $\mathcal{C}^{(\text{dist})}_{J,1-\alpha} \cap \{\theta_3 = \theta_3^*\}$ at $J = 180$. This plot is taken from one of the panels of Figure 22 in Appendix C.

regarding θ_1 are due to numerical optimization. In an experimental approach, we have therefore removed all those observations $\tilde{\theta}^{(n)}_{J_{\text{sim}}}$, $n \in \{1, \ldots, n_{\text{rep}}\}$, for the estimation of $\Sigma_\Phi(\hat{\theta}_J)$ whose first component would be classified as an outlier in a classical boxplot. Thus, writing

$$\tilde{\theta}^{(n)}_{J_{\text{sim}}} = \left(\tilde{\theta}^{(n)}_{J_{\text{sim}},1}, \tilde{\theta}^{(n)}_{J_{\text{sim}},2}, \tilde{\theta}^{(n)}_{J_{\text{sim}},3}\right)^\top, \quad n \in \{1, \ldots, n_{\text{rep}}\},$$

we calculated the first and third quartile of the observations' first component, denoted by

$$\tilde{q}_{0.25}\left(\tilde{\theta}^{(1)}_{J_{\text{sim}},1}, \ldots, \tilde{\theta}^{(n_{\text{rep}})}_{J_{\text{sim}},1}\right) \quad \text{and} \quad \tilde{q}_{0.75}\left(\tilde{\theta}^{(1)}_{J_{\text{sim}},1}, \ldots, \tilde{\theta}^{(n_{\text{rep}})}_{J_{\text{sim}},1}\right),$$

respectively. The interquartile range is given by the difference between these quartiles,

$$\mathrm{IQR}\left(\tilde{\theta}^{(1)}_{J_{\text{sim}},1}, \ldots, \tilde{\theta}^{(n_{\text{rep}})}_{J_{\text{sim}},1}\right) = \tilde{q}_{0.75}\left(\tilde{\theta}^{(1)}_{J_{\text{sim}},1}, \ldots, \tilde{\theta}^{(n_{\text{rep}})}_{J_{\text{sim}},1}\right) - \tilde{q}_{0.25}\left(\tilde{\theta}^{(1)}_{J_{\text{sim}},1}, \ldots, \tilde{\theta}^{(n_{\text{rep}})}_{J_{\text{sim}},1}\right).$$

For the calculation of $\hat{\Sigma}_\Phi(\hat{\theta}_J)$, we then removed all observations $n \in \{1, \ldots, n_{\text{rep}}\}$ with

$$\begin{aligned} &\tilde{\theta}^{(n)}_{J_{\text{sim}},1} > \tilde{q}_{0.75}\left(\tilde{\theta}^{(1)}_{J_{\text{sim}},1}, \ldots, \tilde{\theta}^{(n_{\text{rep}})}_{J_{\text{sim}},1}\right) + 1.5 \cdot \mathrm{IQR}\left(\tilde{\theta}^{(1)}_{J_{\text{sim}},1}, \ldots, \tilde{\theta}^{(n_{\text{rep}})}_{J_{\text{sim}},1}\right) \\ \text{or} \quad &\tilde{\theta}^{(n)}_{J_{\text{sim}},1} < \tilde{q}_{0.25}\left(\tilde{\theta}^{(1)}_{J_{\text{sim}},1}, \ldots, \tilde{\theta}^{(n_{\text{rep}})}_{J_{\text{sim}},1}\right) - 1.5 \cdot \mathrm{IQR}\left(\tilde{\theta}^{(1)}_{J_{\text{sim}},1}, \ldots, \tilde{\theta}^{(n_{\text{rep}})}_{J_{\text{sim}},1}\right), \end{aligned}$$

and adjusted n_{rep} accordingly. Figure 6 shows how this procedure affects the plot in Figure 5 (the same seed was used for these confidence regions). While the elliptical shape of the intersection becomes more apparent, the overall impact is moderate, especially when we look at the confidence regions of the following paragraphs. This suggests that the shape of the confidence regions may not be caused by inaccuracies in the numerical

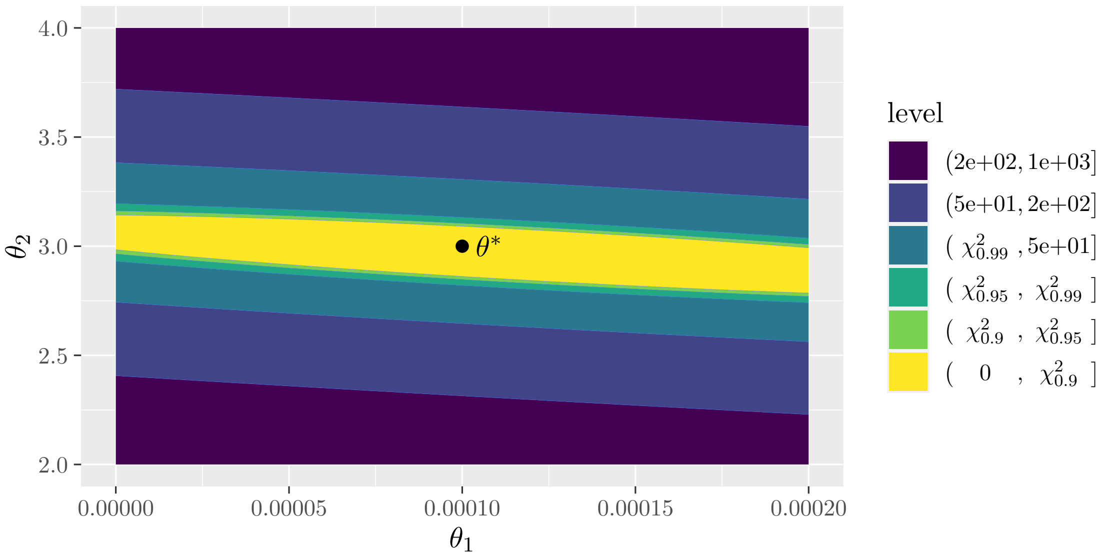


Figure 6: Visualization of $\mathcal{C}^{(\text{dist})}_{J,1-\alpha} \cap \{\theta_3 = \theta_3^*\}$ at $J = 180$, where the outliers relating to θ_1 were removed for the estimation of $\Sigma_\Phi(\hat{\theta}_J)$.

optimization, but by the inherent characteristics of the model $^\times$D. We will come to a better understanding of this by the end of the following paragraph, once we have examined the confidence regions constructed from the 3-sign depth test.

(ii) Confidence Sets Based on the 3-Sign Depth Test

For the remaining statistical methods, the construction of confidence regions is more straightforward than for the minimum distance estimator. In both cases, we utilize the fact that confidence regions can easily be obtained from hypothesis tests: If $\varphi_{\alpha,\theta_0}$ denotes a level $\alpha \in (0,1)$ test for the one-point hypothesis $\mathcal{H}_0 : \theta^* = \theta_0$, then a $(1-\alpha)$-confidence region for the true parameter θ^* is given by the set

$$\{\theta_0 \in \Theta : \varphi_{\alpha,\theta_0} = 0\} . \tag{6.7}$$

In the case of the K-sign depth test from Definition 4.7, this region becomes

$$\{\theta_0 \in \Theta : \Psi_K(\mathcal{W}^\eta(\theta_0)) \geq q_\alpha(\Psi_K)\}$$

because of $\Theta_0 = \{\theta_0\}$, so we do not have to take the supremum here. In the model $^\times$D, $\Psi_K(\mathcal{W}^\eta(\theta_0))$ is the normalized K-sign depth based on η standardized hazard transforms at θ_0 ordered with respect to $\leq_{\text{acc}}$ (cf. Definitions 2.29 and 4.15 as well as Theorem 4.5). The total number $\eta = \eta(J)$ (cf. Assumption 4.12) of observed points depends on the random covariates τ_j and C_j. In the context of this simulation study, we always have $\eta = 10J$ due to $C_j = 10$ and sufficiently large τ_j for each $j \in \{1, \ldots, J\}$. In the following, the confidence region for the true parameter θ^* based on the 3-sign depth test is denoted by

$$\mathcal{C}^{(\text{depth})}_{J,1-\alpha} := \left\{\theta_0 \in \Theta : \Psi_3(\mathcal{W}^{\eta(J)}(\theta_0)) \geq q_\alpha(\Psi_3)\right\} .$$

α	0.01	0.025	0.05	0.1
$q_\alpha(\Psi_3)$	-2.1947	-1.6398	-1.2182	-0.7976

Table 3: Computed values for the quantiles $q_\alpha(\Psi_3)$ of the asymptotic distribution $\Psi_3(W)$.

In order to compute $\mathcal{C}^{(\text{depth})}_{J,1-\alpha}$, we require the quantiles $q_\alpha(\Psi_3)$ of the asymptotic distribution $\Psi_3(W)$ of the 3-sign depth. This distribution is explicitly given in Theorem 4.4, enabling us to approximate its quantiles through a simulation. We describe this procedure in Algorithm 6.3. The calculated quantiles using $\texttt{rep} = 10^6$ and $\texttt{step} = 10^5$ are listed in Table 3.

Algorithm 6.3 Approximation of the quantile $q_\alpha(\Psi_3)$ via the empirical α-quantile of i.i.d. samples from the asymptotic distribution $\Psi_3(W)$. The Brownian motion W on $[0,1]$ is realized by forming cumulative sums of i.i.d. normally distributed random variables.

Input:

$\alpha \in (0,1)$ which quantile is to be approximated,
$\texttt{rep} \in \mathbb{N}$ number of i.i.d. samples from $\Psi_3(W)$ to be simulated,
$\texttt{step} \in \mathbb{N}$ number of i.i.d. random variables used to simulate W.

Output:

$\tilde{q}_\alpha \in \mathbb{R}$ empirical α-quantile of the $\texttt{rep}$ i.i.d. samples from $\Psi_3(W)$.

for $r = 1, \dots, \texttt{rep}$ **do**
  draw i.i.d. samples $x_1, \dots, x_{\texttt{step}}$ of the standard normal distribution $\mathcal{N}(0,1)$
  $W \leftarrow \frac{1}{\sqrt{\texttt{step}}} \texttt{cumsum}\left(x_1, \dots, x_{\texttt{step}}\right)$ // simulate Brownian motion on $[0,1]$
  $\texttt{to\_Int} \leftarrow 0$
  for $i = 1, \dots, \texttt{step}$ **do**
    $\texttt{to\_Int} \leftarrow \texttt{to\_Int} + W_i^2 - W_{\texttt{step}} \cdot W_i$ // compute simplified integral for $\Psi_3(W)$
  end for
  $\psi_r \leftarrow \frac{3}{4}(1 - W_{\texttt{step}}^2) - 3\frac{\texttt{to\_Int}}{\texttt{step}}$ // get rth realization of $\Psi_3(W)$
end for
compute empirical α-quantile $\tilde{q}_\alpha = \tilde{q}_\alpha\left(\psi_1, \dots, \psi_{\texttt{rep}}\right)$

From here we can proceed exactly as in paragraph (i). We again visualize intersections of the 3-dimensional confidence regions with selected hyperplanes. To obtain $(1-\alpha)$-confidence regions for different levels $\alpha \in \{0.01, 0.05, 0.1\}$ at once, we show in each case a contour plot of the normalized 3-sign depth as a function of the parameter θ. We provide the full plots in Appendix C, see Figures 25 (for the θ_1-θ_2-direction), 26 (for the θ_1-θ_3-direction) and 27 (for the θ_2-θ_3-direction). Figure 7 shows the intersection of $\mathcal{C}^{(\text{depth})}_{30,1-\alpha}$ with the hyperplane $\{\theta_2 = \theta_2^*\}$. We immediately notice that the confidence regions are less smooth compared to the ellipses we obtained from $\mathcal{C}^{(\text{dist})}_{J,1-\alpha}$. Nevertheless, the intersections with the hyperplanes $\{\theta_2 = \theta_2^*\}$ as well as $\{\theta_1 = \theta_1^*\}$ are still nearly elliptical, especially when we increase J. The plot in θ_1-θ_2-direction is of particular interest to us, since we still owe an explanation for the large deviation with regard to θ_1 that we observed in paragraph (i) above. We give this plot in Figure 8 at $J = 180$, which can therefore be compared directly to Figure 5 as it was calculated from the exact same data. Similar to the Wald-type confidence regions based on the minimum distance estimator, the confidence regions here again appear to be stretched in the θ_1-direction.

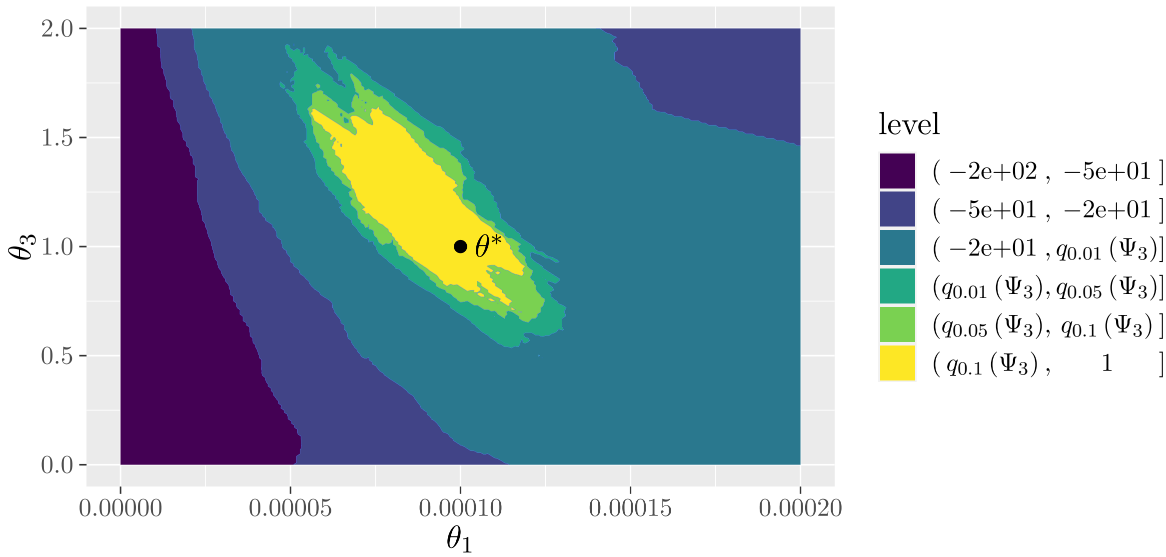


Figure 7: Visualization of $\mathcal{C}^{(\text{depth})}_{J,1-\alpha} \cap \{\theta_2 = \theta_2^*\}$ at $J = 30$. This plot is taken from one of the panels of Figure 26 in Appendix C.

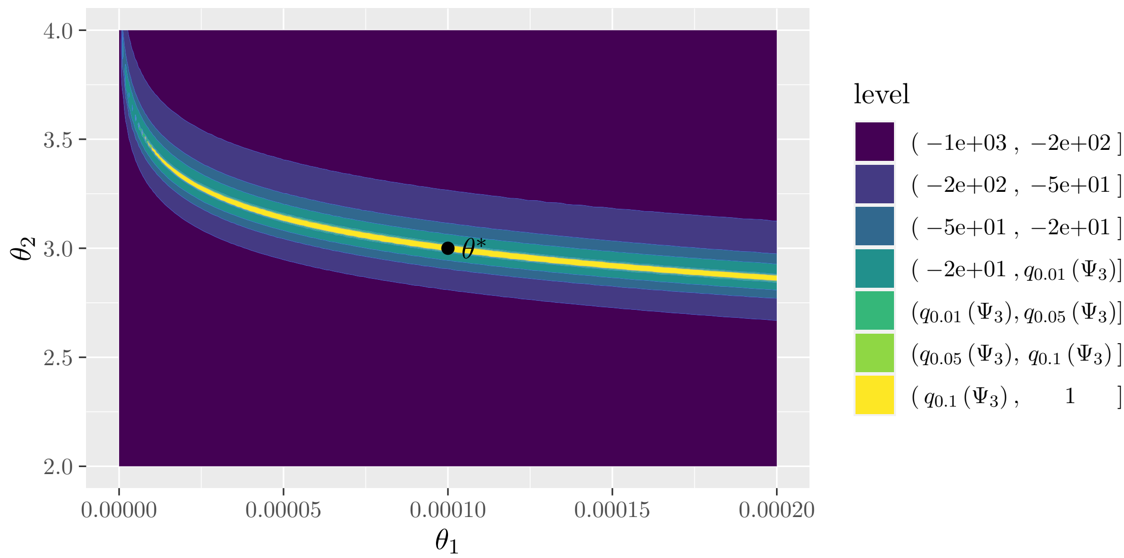


Figure 8: Visualization of $\mathcal{C}^{(\text{depth})}_{J,1-\alpha} \cap \{\theta_3 = \theta_3^*\}$ at $J = 180$. This plot is taken from one of the panels of Figure 25 in Appendix C.

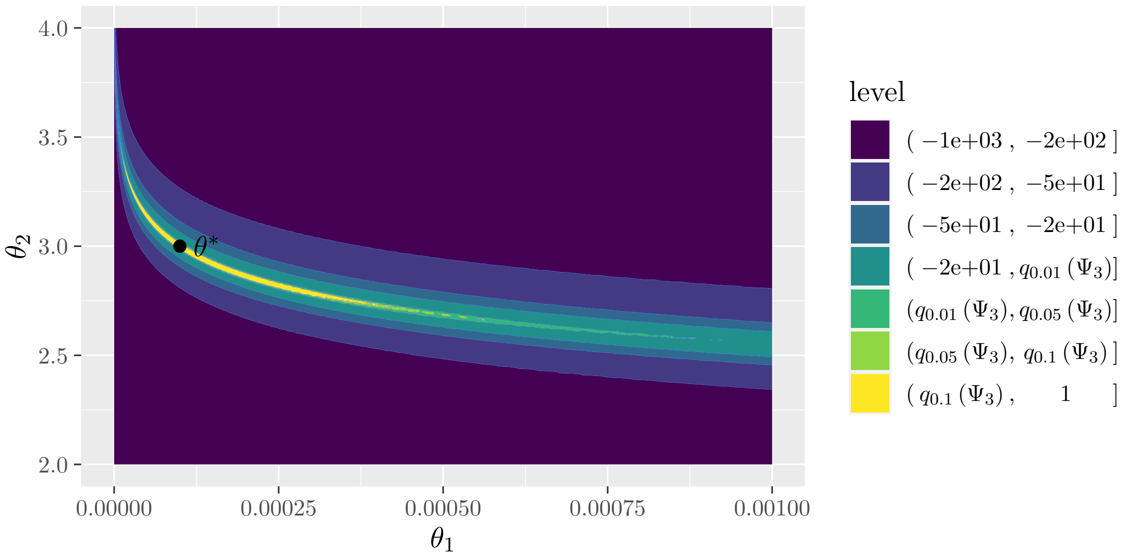


Figure 9: Extended visualization of $\mathcal{C}^{(\text{depth})}_{J,1-\alpha} \cap \{\theta_3 = \theta_3^*\}$ at $J = 180$. The plot is functionally identical to Figure 8, but the displayed area has been extended in θ_1-direction.

From the consistency of the 3-sign depth test, we know that this "band" cannot extend indefinitely as we increase J, and instead contracts to the true parameter θ^*. We can already see this for $J = 180$, but we have to look at an area five times as large and extend the contour plot up to $\theta_1 = 0.001$, see Figure 9. What we could not yet see from the confidence region $\mathcal{C}^{(\text{dist})}_{J,1-\alpha}$ now becomes evident: The visible "band" is characteristic of an exponential decay, so that a logarithmic scale is more suitable for the parameter θ_1.
In the panel "linear" of Figure 10, we find that the minimum distance estimates of the true parameter are essentially located along this band. If these estimates (and thus virtually the entire band) are to be contained in an ellipse, it must be disproportionately large. This is exactly the behavior we noticed in Figure 5.
By instead plotting $\ln(\theta_1)$ against θ_2, these estimates form the anticipated ellipses, as shown in the panel "logarithmic" of Figure 10. The alternative parametrization of Equation (2.28) is hence the better choice for minimum distance estimation.
For the remainder of this thesis, however, we will stick to the original parametrization of model $^\times$D. The important finding here is that the confidence regions $\mathcal{C}^{(\text{depth})}_{J,1-\alpha}$ perform reasonably even under suboptimal modelling decisions. As we will discover shortly, this also applies to the confidence regions based on the likelihood ratio.

(iii) Confidence Sets Based on the Likelihood Ratio

Our final confidence region is constructed directly from the likelihood ratio test statistic. For the hypotheses

$$\mathcal{H}_0 \,:\, \theta^* \in \Theta_0 \quad \text{vs.} \quad \mathcal{H}_1 \,:\, \theta^* \in \Theta \setminus \Theta_0$$

and a likelihood function L, this test statistic is given by (cf. Serfling 1980, p. 157):

$$\mathrm{LR}(\Theta_0, \Theta) := -2 \ln \left(\frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)} \right) . \tag{6.8}$$

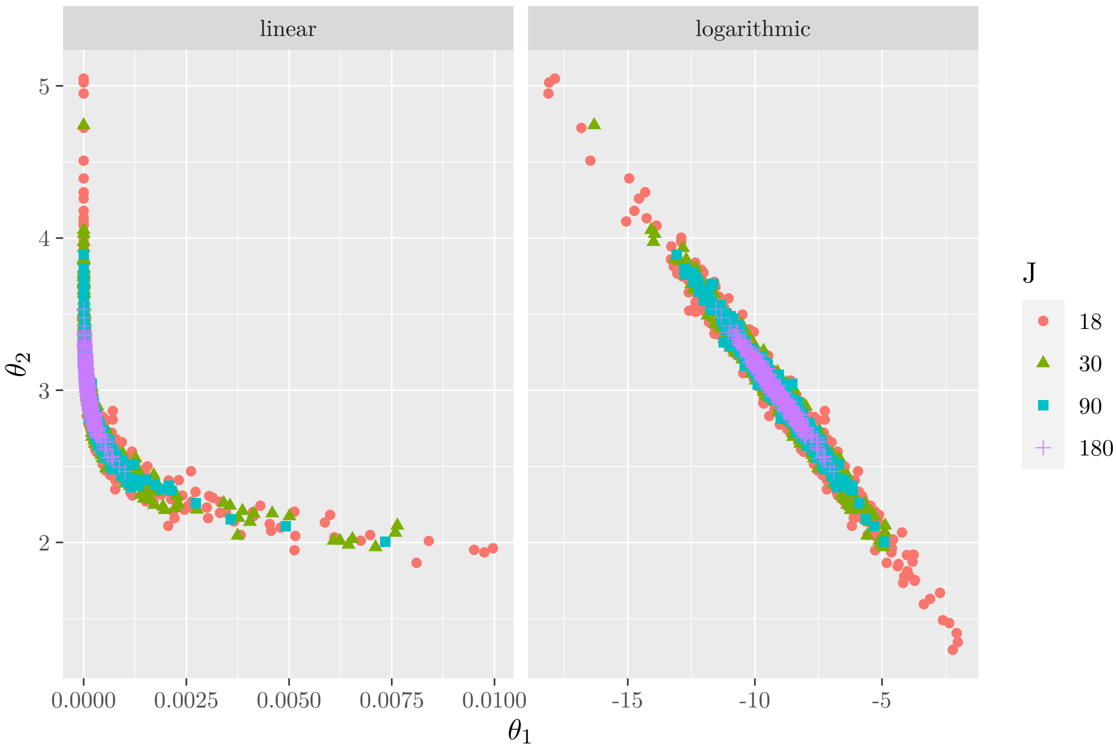


Figure 10: Scatterplots of θ_1 vs. θ_2 (panel "linear") and $\ln(\theta_1)$ vs. θ_2 (panel "logarithmic") for 500 simulated values of the minimum distance estimator at each $J \in \{18, 30, 90, 180\}$. As usual, the parameters & covariates from Table 2 were used.

The likelihood ratio converges in distribution to a χ^2-distribution, whereby the degrees of freedom depend on Θ and Θ_0. In the case of a one-point hypothesis (i.e., $\Theta_0 = \{\theta_0\}$ for some $\theta_0 \in \Theta$), the degrees of freedom correspond to the dimension d of the parameter space Θ (cf. Andersen et al. 1993, p. 403). In general, if Θ_0 is determined by a collection of $r \leq d$ restrictions

$$R_i(\theta) = 0\,, \quad i = 1, \ldots, r\,,$$

so that

$$\Theta_0 = \{\theta \in \Theta \,:\, R_i(\theta) = 0 \text{ for all } i = 1, \ldots, r\}\,, \tag{6.9}$$

then - under certain regularity conditions - the degrees of freedom are given by the number r of restrictions (cf. Serfling 1980, pp. 152, 156–158). If $r < d$, $\mathcal{H}_0 \,:\, \theta^* \in \Theta_0$ is called a composite hypothesis, see also Andersen et al. 1993, p. 426. As an asymptotic test, the likelihood ratio test utilizes the asymptotic distribution of $\mathrm{LR}(\Theta_0, \Theta)$. The likelihood ratio test at level $\alpha \in (0, 1)$ for the one-point hypothesis $\mathcal{H}_0 \,:\, \theta^* = \theta_0$ is defined as

$$\varphi^{(\mathrm{lr})}_{\alpha,\theta_0} = \mathbb{1}\left\{\mathrm{LR}(\,\{\theta_0\}\,, \Theta) > \chi^2_{d,1-\alpha}\right\}.$$

From Equation (6.7), we then immediately obtain the corresponding confidence region for the true parameter θ^*, where $J \in \mathbb{N}$ implicitly affects $\mathrm{LR}(\,\{\theta_0\}\,, \Theta)$ via the likelihood

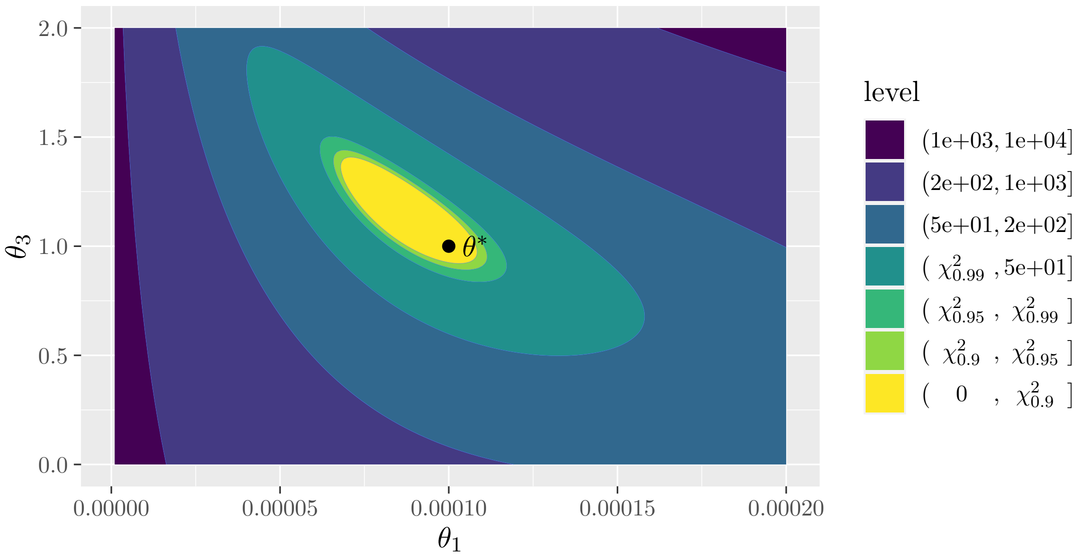


Figure 11: Visualization of $\mathcal{C}^{(\mathrm{lr})}_{J,1-\alpha} \cap \{\theta_2 = \theta_2^*\}$ at $J = 30$. This plot is taken from one of the panels of Figure 29 in Appendix C.

function L:

$$\mathcal{C}^{(\mathrm{lr})}_{J,1-\alpha} := \left\{\theta_0 \in \Theta \, : \, \mathrm{LR}(\{\theta_0\}, \Theta) \leq \chi^2_{d,1-\alpha}\right\} .$$

In the model $^\times$D, we have $d = 3$. Since the quantiles are already known here, we can immediately move on to the visualization, which is analogous to the preceding paragraphs. The plots show the contour lines of $\mathrm{LR}(\{\theta\}, \Theta)$ as a function of θ on the hyperplanes $\{\theta_3 = \theta_3^*\}$ (Figure 28), $\{\theta_2 = \theta_2^*\}$ (Figure 29) and $\{\theta_1 = \theta_1^*\}$ (Figure 30) for $J \in \{18, 30, 90, 180\}$ and the parameters & covariates given in Table 2. As before, they can be found in Appendix C.

The likelihood ratio confidence regions appear similar in shape to the confidence regions based on the 3-sign depth, but with smoother boundaries, especially for smaller sample sizes. As an example, in Figure 11 we look at the intersection of the confidence region $\mathcal{C}^{(\mathrm{lr})}_{30,1-\alpha}$ with the hyperplane $\{\theta_2 = \theta_2^*\}$. We can compare this plot directly with the earlier Figure 7, where the depth-based confidence region was computed from the same data. The smoother boundaries are clearly noticeable here.

Finally, we remark that the likelihood ratio test essentially builds on the asymptotic normality of the maximum likelihood estimator, so that a Wald-type confidence region similar to $\mathcal{C}^{(\mathrm{dist})}_{J,1-\alpha}$ could also be considered. However, we want to emphasize that the confidence region $\mathcal{C}^{(\mathrm{lr})}_{J,1-\alpha}$ does not suffer from the same problems as the Wald-type confidence region constructed from the minimum distance estimator and thus seems less reliant on an appropriate parametrization. Consequently, the confidence region $\mathcal{C}^{(\mathrm{lr})}_{J,1-\alpha}$ shares the advantages of $\mathcal{C}^{(\mathrm{dist})}_{J,1-\alpha}$ and $\mathcal{C}^{(\mathrm{depth})}_{J,1-\alpha}$ without adopting their drawbacks, which makes it stand out in direct comparison. We also observe this in the following study of the size and coverage rate between the three different confidence regions for the true parameter.

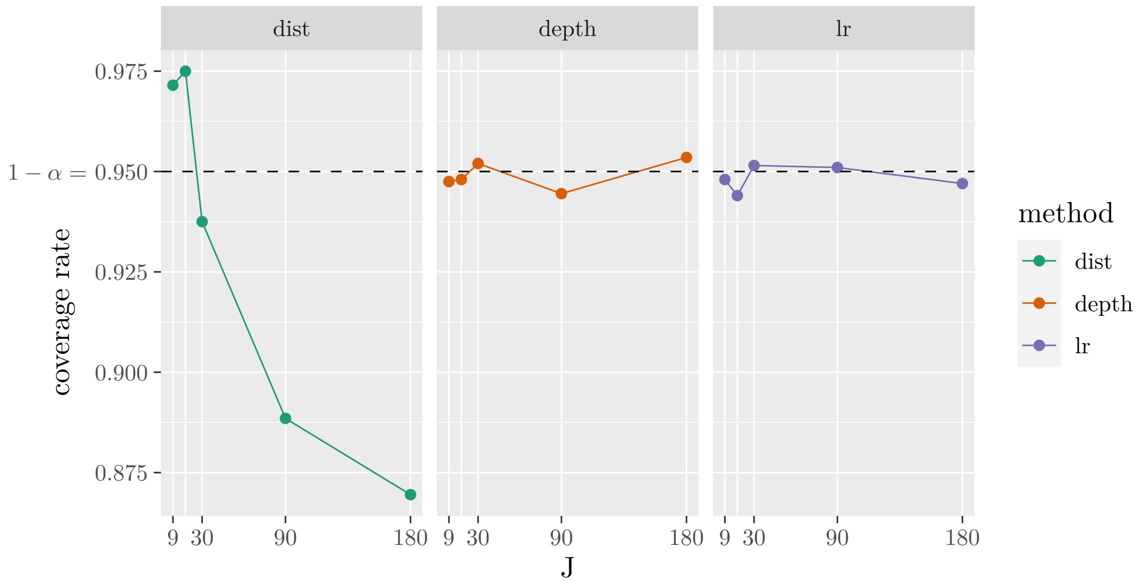


Figure 12: Comparison of the coverage rates of the true parameter θ^* between the confidence regions based on the minimum distance estimator (method “dist”), the 3-sign depth test (method “depth”) and the likelihood ratio (method “lr”) in a simulation study with 2000 simulated $(1-\alpha)$-confidence regions at $\alpha = 0.05$ for each method and each $J \in \{9, 18, 30, 90, 180\}$.

Simulation Study on Size and Coverage Rate of the Confidence Regions

So far, we have visualized only a single realization of the proposed confidence regions. To compare them in a statistically sound manner, we compute 2000 confidence regions at each $J \in \{9, 18, 30, 90, 180\}$ for all of the methods and evaluate both their size and coverage rate of the true parameter. We compare the size by placing a $41 \times 41 \times 41$ grid centred around θ^* in the parameter space Θ. For each method and $J \in \{9, 18, 30, 90, 180\}$, we then determine how many grid points lie on average in the generated confidence regions. The grid is chosen in such a way that the parameter ranges shown in Figures 22 through 30 are covered, but we use a logarithmic scale in θ_1-direction to account for the larger confidence regions based on the minimum distance estimator. In total, we checked all the $41 \times 41 \times 41 = 68921$ parameter combinations $(\theta_1, \theta_2, \theta_3)^\top \in \Theta$ where

$$\begin{aligned}
\log(\theta_1) &\in \{ \; -6 \;\; , -5.90, -5.80, \;\; \ldots \;\; , \;\; -2 \;\; \}, \\
\theta_2 &\in \{ \;\;\; 2 \;\;\; , \;\; 2.05, \;\;\; 2.10, \;\;\; \ldots \;\; , \;\;\;\; 4 \;\;\; \}, \\
\theta_3 &\in \{ \;\;\; 0 \;\;\; , \;\; 0.05, \;\;\; 0.10, \;\;\; \ldots \;\; , \;\;\;\; 2 \;\;\; \}.
\end{aligned}$$

To estimate the coverage rate, we examined how often on average the true parameter θ^* was located within the generated confidence region. A summary of all results is given in Table 8 in Appendix C. Figure 12 shows the coverage rates for the true parameter θ^* obtained from 2000 simulated $(1-\alpha)$-confidence regions at $\alpha = 0.05$ for each of the three methods and each $J \in \{9, 18, 30, 90, 180\}$. While the confidence regions based on the 3-sign depth and the likelihood ratio meet the $(1-\alpha)$-level, the Wald-type confidence region obtained from the minimum distance estimator fails to do so for larger J. This is

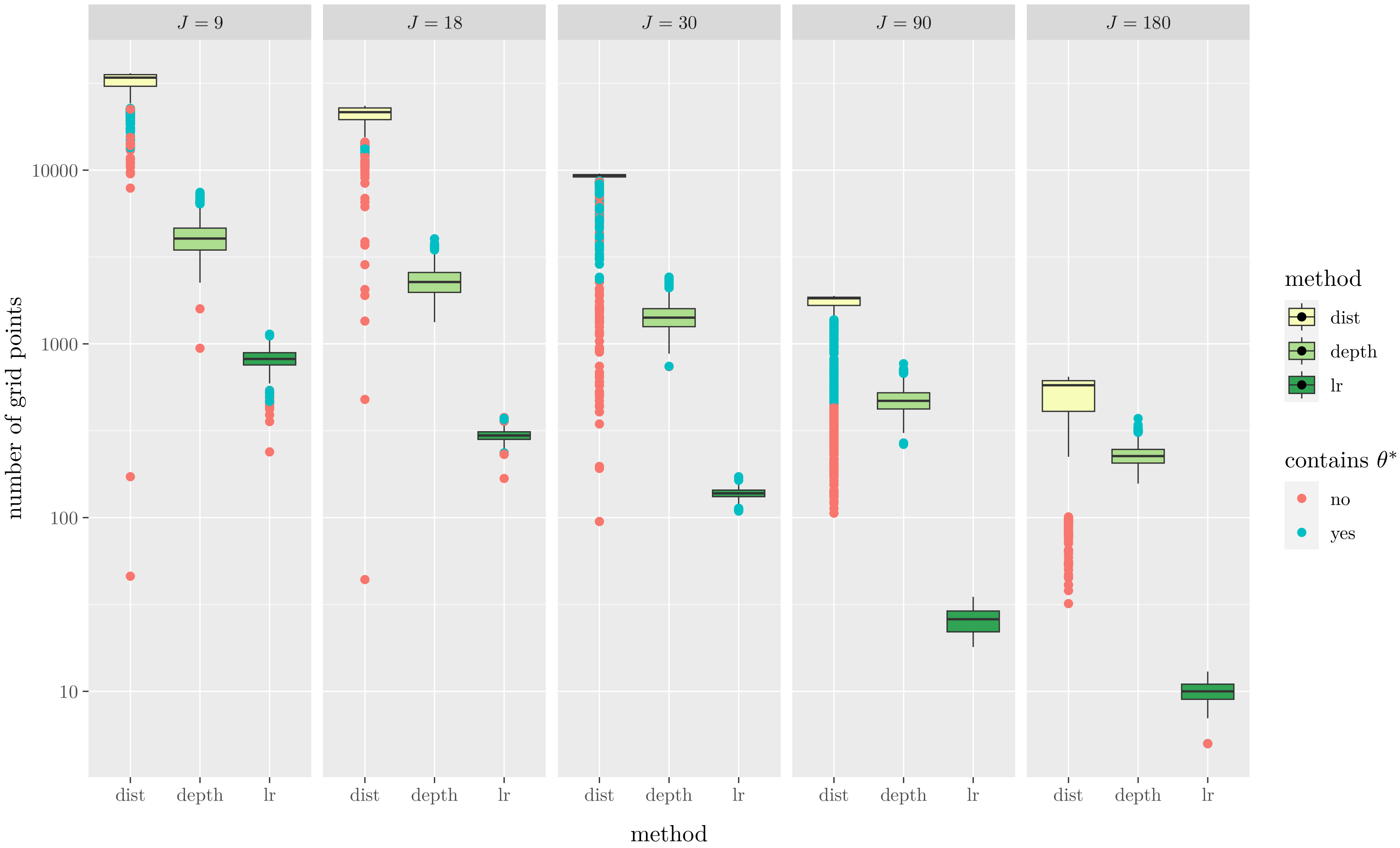


Figure 13: Comparison of the sizes between the confidence regions based on the minimum distance estimator (method “dist”), the 3-sign depth test (method “depth”) and the likelihood ratio (method “lr”) through box plots. The size of a simulated confidence region is measured by counting the grid points of a $41 \times 41 \times 41$ grid on $[10^{-6}, 10^{-2}] \times [2, 4] \times [0, 2] \subset \Theta$ it contains. The simulation study features 2000 simulated $(1 - \alpha)$-confidence regions at $\alpha = 0.05$ for each method and each $J \in \{9, 18, 30, 90, 180\}$. For each outlier, the color indicates whether the associated confidence region contained the true parameter θ^* (blue) or not (red).

unexpected, as it seems to contradict the consistency of that estimator. For comparison of size, in Figure 13 we provide box plots for the numbers of grid points contained in the simulated confidence regions. The results are consistent across all sample sizes $J \in \{9, 18, 30, 90, 180\}$ and show that the method “lr” (likelihood ratio) produces the smallest confidence regions, followed by the method “depth” (3-sign depth test). The confidence regions given by the method “dist” (minimum distance estimator), on the other hand, are so large that a logarithmic axis had to be used to allow visual comparison. For this, a total of 20 observations, where the number of grid points was 0, had to be removed from the simulated data set. In these cases, the entire calculated confidence region was located somewhere outside the grid. Despite the size of its confidence regions, this only happened for the method “dist” at $J = 9$ (nine times), $J = 18$ (three times) and $J = 30$ (eight times). Figure 13 also shows that atypically small confidence regions obtained from this method - classified as outliers in the box plots - generally do not contain the true parameter θ^*. It can therefore be assumed that these extend beyond the observed grid and are consequently evaluated as being smaller than they actually are.
This further demonstrates that a comparison based on the contained grid points only allows for rough estimates of the confidence regions’ sizes. The fact that for fixed J the Wald-type confidence regions based on the minimum distance estimator should have the same size, since they share the same estimate for the asymptotic covariance matrix, reinforces this criticism. It is the reason why the box collapses at $J = 30$, while fluctuations in size can occur for larger J due to shifts within the grid. This happens because the confidence regions become comparatively small in relation to the resolution of the grid. Moreover, non-convex confidence regions may be located unfavourably between the grid points, making them appear considerably smaller. Because of that, a size comparison based on grid points only works well if the confidence regions are convex and the resolution of the grid is high enough, both of which can be questioned in our scenario (cf. Figures 25 and 28). Nevertheless, the boxes in Figure 13 are rather short and never overlap regardless of the sample size J. This lends a certain credibility to our conclusions, as we would expect larger fluctuations for either an insufficient resolution or confidence regions whose shape does not fit well with the given grid.

6.3. Comparison of Hypothesis Tests for Damage Accumulation

In the last section, we studied confidence sets for the true parameter θ^*. As we have seen in the construction of the confidence regions based on either the 3-sign depth or the likelihood ratio, these correspond directly to statistical tests for one-point hypotheses. We next turn to the composite hypothesis $\mathcal{H}_0 : \theta_3^* = 0$, which is of particular interest as it can be used to test the significance of the damage accumulation term. The aim of this section is to conduct tests for this hypothesis based on the minimum distance estimator, the 3-sign depth and the likelihood ratio, and to evaluate their power.
Since the 3-sign depth test can by definition also be performed for composite hypotheses, for this we can again apply the hypothesis test from Definition 4.7 by setting

$$\Theta_0 = \left\{\theta = (\theta_1, \theta_2, \theta_3)^\top \in \Theta : \theta_3 = 0\right\} .$$

We obtain the same set from Equation (6.9) if we define $R_1(\theta) := \theta_3$. We can therefore continue to use the likelihood ratio test statistic from Equation (6.8), where the degrees of freedom of its asymptotic χ^2-distribution now amount to $r = 1$. This leaves the minimum

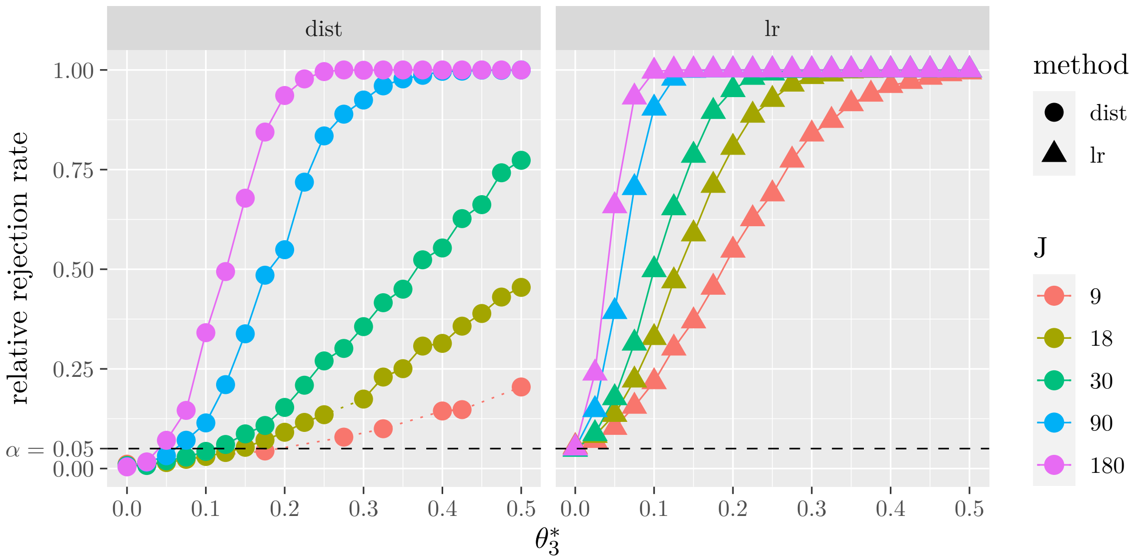


Figure 14: Rejection rates of the level $\alpha = 0.05$ tests for damage accumulation, i.e., for the hypothesis $\mathcal{H}_0 : \theta_3^* = 0$, based on the minimum distance estimator (method "dist") and the likelihood ratio (method "lr") at different values of θ_3^* and $J \in \{9, 18, 30, 90, 180\}$. The remaining parameter values were fixed at $\theta_1^* = 10^{-4}$ and $\theta_2^* = 3$. Missing values indicate that the estimated covariance matrix was computationally singular, so no test could be performed for method "dist". Dotted lines represent a linear interpolation between available data.

distance approach, for which a suitable test must first be constructed. A basic test for composite hypotheses can be derived from a confidence region for the true parameter: If $\mathcal{C}_{1-\alpha}$ is a $(1-\alpha)$-confidence region for θ^* and $\Theta_0 \subset \Theta$ is an arbitrary subset, then a level α test for $\mathcal{H}_0 : \theta^* \in \Theta_0$ is given by

$$\varphi = \begin{cases} 1\,, & \text{if } \mathcal{C}_{1-\alpha} \cap \Theta_0 = \emptyset\,, \\ 0\,, & \text{otherwise.} \end{cases} \tag{6.10}$$

This directly follows from the definition of a confidence set, as for any $\theta^* \in \Theta_0$ we have:

$$\mathbb{P}_{\theta^*}(\varphi = 1) = \mathbb{P}_{\theta^*}(\mathcal{C}_{1-\alpha} \cap \Theta_0 = \emptyset) \leq \mathbb{P}_{\theta^*}(\mathcal{C}_{1-\alpha} \cap \{\theta^*\} = \emptyset) = \mathbb{P}_{\theta^*}(\theta^* \notin \mathcal{C}_{1-\alpha}) \leq \alpha\,.$$

The test φ is rooted in a simple reasoning: The confidence region $\mathcal{C}_{1-\alpha}$ contains the plausible values for the true parameter based on the observed data. If this set does not include a single parameter from Θ_0, then no parameter from Θ_0 is a suitable candidate for θ^*. We therefore reject $\mathcal{H}_0 : \theta^* \in \Theta_0$. The test also has an easy visual interpretation. In the case $\Theta_0 = \{\theta_3 = 0\}$, we only need to check whether the computed confidence region cuts through the θ_1-θ_2-plane. The test for damage accumulation based on the Wald-type confidence region $\mathcal{C}_{J,1-\alpha}^{(\mathrm{dist})}$ is now given by

$$\mathbb{1}\left\{\sup_{\theta \in \Theta_0} J(\hat{\theta}_J - \theta)^\top \hat{\Sigma}_\Phi(\hat{\theta}_J)^{-1}(\hat{\theta}_J - \theta) > \chi^2_{d,1-\alpha}\right\}.$$

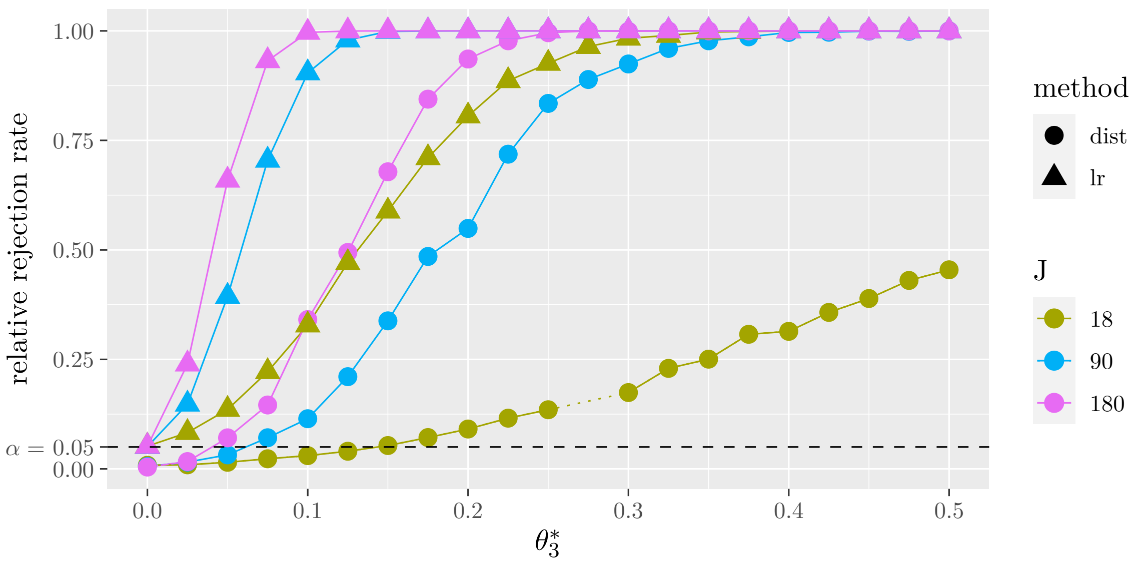


Figure 15: Combined power plot from Figure 14. For better readability, only $J \in \{18, 90, 180\}$ were considered.

Remarkably, the 3-sign depth test from Definition 4.7 was derived according to the same principle. Both tests therefore share the shortcoming that they are usually rather conservative and fall well below the targeted level α.
We can now compare the available tests for damage accumulation in a simulation study. To this end, we simulate $J = 180$ point process realizations at $\theta^* = (\theta_1^*, \theta_2^*, \theta_3^*)^\top$, where $\theta_1^* = 10^{-4}$ and $\theta_2^* = 3$ are fixed, while θ_3^* ranges from 0 to 0.5. We then apply the three tests to this data (for $J \in \{9, 18, 30, 90\}$, a subset of the data is used), and repeat this 5000 times for each possible value of θ_3^*. Depending on θ_3^*, the relative rejection rates of these tests can then be determined. For the tests based on the minimum distance estimator (method "dist") and the likelihood ratio (method "lr"), they are plotted in Figure 14. As expected, for both methods we see an increase in power as the sample size J grows. We find that the test based on the minimum distance estimator is too conservative, which is common for a test constructed via Equation (6.10). This does benefit the method to some extent, as we saw in the previous section that the corresponding confidence regions struggled to keep their level. Moreover, we again observe that for $J = 9$ (and here even once for $J = 18$) the estimated covariance matrix is often computationally singular. Since its inverse cannot be determined in these cases, the test for damage accumulation cannot be carried out, leading to missing values in the power plot. These missing values can be imputed by (e.g., linear) interpolation, which is indicated by a dotted line. In Figure 15, we combined both panels of Figure 14 to directly compare the competing methods. Apparently, the method "dist" requires around ten times as many observations in order to match the method "lr", compare "lr" for $J = 18$ and "dist" for $J = 180$. Nevertheless, we find that even the basic test based on the minimum distance estimator achieves acceptable power. However, this does not apply to the 3-sign depth test for damage accumulation: Across all repetitions and values for $\theta_3^* \in [0, 0.5]$, this test could never reject the null hypothesis. We have therefore omitted it from Figures 14 and 15,

J	360	450	540	630
rejection rate	0.004	0.051	0.322	0.752

Table 4: Rejection rates of the 3-sign depth test for damage accumulation, i.e., for the hypothesis $\mathcal{H}_0 : \theta_3^* = 0$, at $\theta^* = \left(10^{-4}, 3, 0.5\right)^\top$ and $J \in \{360, 450, 540, 630\}$.

because the relative rejection rate is a constant 0. As it turns out, the 3-sign depth test requires even larger sample sizes for consistency to take effect: For $J = 180$, the confidence region still extended beyond the θ_1-θ_2-plane for each of the 5000 realizations. For significantly larger J, the confidence set narrows to the true parameter θ^* and at some point no longer intersects the θ_1-θ_2-plane. At $\theta^* = \left(10^{-4}, 3, 0.5\right)^\top$, we started to observe this for $J \geq 360$. The rejection rates for $J \in \{360, 450, 540, 630\}$ are given in Table 4. Although the power further increases as J becomes larger, it is still unacceptably low in relation to the sample size, so we discontinued our investigations at this point.

6.4. Study on the Robustness of the Methods for Contaminated Data

Kopperschmidt and Stute 2013, p. 1278 claim that the minimum distance procedure yields robust consistent estimates of the true parameter θ^*. Furthermore, the 3-sign depth test as a generalization of the classical sign test is naturally outlier robust, because its test statistic only accounts for the signs of the (transformed) observations. We want to assess the robustness of these methods in the presence of contaminated data and evaluate whether the hitherto superior likelihood methods fall short in this setting.
To generate contaminated data, we modify the “raw data”, that is, the i.i.d. samples $y_1, \ldots, y_n$ of the unit exponential distribution needed to simulate a point process realization (see Algorithm 6.1). We consider two types of modification:

(1) **Depth-specific contamination.** Contaminate the raw data by increasing the deviation from the median $\ln(2)$ of the $\mathcal{E}(1)$-distribution. For a contamination at index i, replace the ith sample y_i with

$$\tilde{y}_i = \max\left\{2\left(y_i - \ln(2)\right) + \ln(2),\, q_{0.0001}\left(\mathcal{E}(1)\right)\right\},$$

where $q_{0.0001}\left(\mathcal{E}(1)\right)$ is the 0.0001-quantile of the unit exponential distribution. The maximum is required to prevent negative observations. With this modification, the signs of the standardized hazard transforms at θ^* are retained. The 3-sign depth test is therefore virtually unaffected by this contamination, since only the order of the hazard transforms with respect to $\leq_{\text{acc}}$ changes slightly. As it is tailored to depth-based methods, we refer to this type of contamination as depth-specific.

(2) **Quantile-based contamination.** Contaminate the raw data by replacing it with atypically small or large values in terms of the $\mathcal{E}(1)$-distribution. For a contamination at index i, randomly replace the ith sample y_i with either

$$\underbrace{\tilde{y}_i = q_{0.0001}\left(\mathcal{E}(1)\right)}_{\text{atypically small}} \quad \text{or} \quad \underbrace{\tilde{y}_i = q_{0.9999}\left(\mathcal{E}(1)\right)}_{\text{atypically large}}.$$

Since this type of contamination involves the quantiles of the unit exponential distribution, we call it quantile-based.

type of contamination	contaminated fraction 20%	40%
depth-specific	d20	d40
quantile-based	q20	q40

Table 5: Labels for the considered combinations of fraction and type of contamination.

Algorithm 6.4 shows how the contamination of a specified proportion of the raw data is carried out. In our robustness study, we compare the methods by the respective tests for the one-point hypothesis

$$\mathcal{H}_0 \,:\, \theta^* = \left(10^{-4}, 3, 1\right)^\top .$$

For this purpose, we simulate 5000 times $J = 180$ point process realizations at each of 9 parameter vectors along a line segment through $\theta_0 = \left(10^{-4}, 3, 1\right)^\top$. All these parameter vectors are listed in Table 6. The raw data used for these $9 \times 5000 \times 180$ point process realizations are then contaminated. We apply Algorithm 6.4 with `contam_type` $= 1, 2$ (both types of contamination) and `contam_frac` $= 0.2, 0.4$ (20% or 40% contamination).

Algorithm 6.4 Generating the contaminated raw data $\tilde{y}_{1:n}$ from the raw data $y_{1:n}$ used in Algorithm 6.1. The contaminated point process realization $\tilde{t}_{1:n}$ is then obtained by substituting $\tilde{y}_i$ instead of y_i into the inverse cchf.

Input:

$y_1, \ldots, y_n \in \mathbb{R}_+$	i.id. samples of $\mathcal{E}(1)$ used to simulate $t_{1:n}$,
`contam_type` $\in \{1, 2\}$	depth-specific (1) or quantile-based (2) contamination,
`contam_frac` $\in [0, 1]$	fraction of the data to be contaminated.

Output:

$\tilde{y}_1, \ldots, \tilde{y}_n \in \mathbb{R}_+$	contaminated sample.

```
contam_numb ← ⌊contam_frac · n⌋ // number of samples to contaminate
contam_ind ← (contam_numb randomly selected distinct indices from {1, ..., n})
for i = 1, ..., n do
  if i in contam_ind then
    if contam_type = 1 then // depth-specific contamination
      ỹ_i ← max {2 (y_i − ln(2)) + ln(2), q_0.0001 (E(1))}
    else // quantile-based contamination
      draw sample x from Bin (1, 0.5) // direction of the contamination
      if x = 0 then
        ỹ_i ← q_0.0001 (E(1))
      else
        ỹ_i ← q_0.9999 (E(1))
      end if
    end if
  else
    ỹ_i ← y_i
  end if
end for
```

number of parameter vector	θ_1^*	θ_2^*	θ_2^*
1	$9.2 \cdot 10^{-5}$	2.92	0.92
2	$9.4 \cdot 10^{-5}$	2.96	0.96
3	$9.8 \cdot 10^{-5}$	2.98	0.98
4	$9.9 \cdot 10^{-5}$	2.99	0.99
5	$10.0 \cdot 10^{-5}$	3.00	1.00
6	$10.1 \cdot 10^{-5}$	3.01	1.01
7	$10.2 \cdot 10^{-5}$	3.02	1.02
8	$10.4 \cdot 10^{-5}$	3.04	1.04
9	$10.8 \cdot 10^{-5}$	3.08	1.08

Table 6: Parameter vectors used in the robustness study. The parameter $\theta_0 = (10^{-4}, 3, 1)^\top$ is highlighted.

This leads to four contaminated data sets, which we label d20, d40, q20, q40, see Table 5. For each of the total of five data set (no contamination, d20, d40, q20, q40) and each $J \in \{9, 18, 30, 90, 180\}$, we can then perform the three tests for the null hypothesis $\mathcal{H}_0 : \theta^* = \theta_0$ and compute the relative rejection rates at each of the 9 parameters based on the 5000 repetitions. The detailed results are given in Tables 9 (for the test based on the minimum distance estimator), 10 (for the 3-sign depth test) and 11 (for the likelihood ratio test) in Appendix C. We first compare the results for the uncontaminated data set as a starting point for the robustness study. In Figure 16 we show for all three methods and $J \in \{9, 30, 180\}$ the relative rejection rates at the 9 parameter vectors from Table 6.

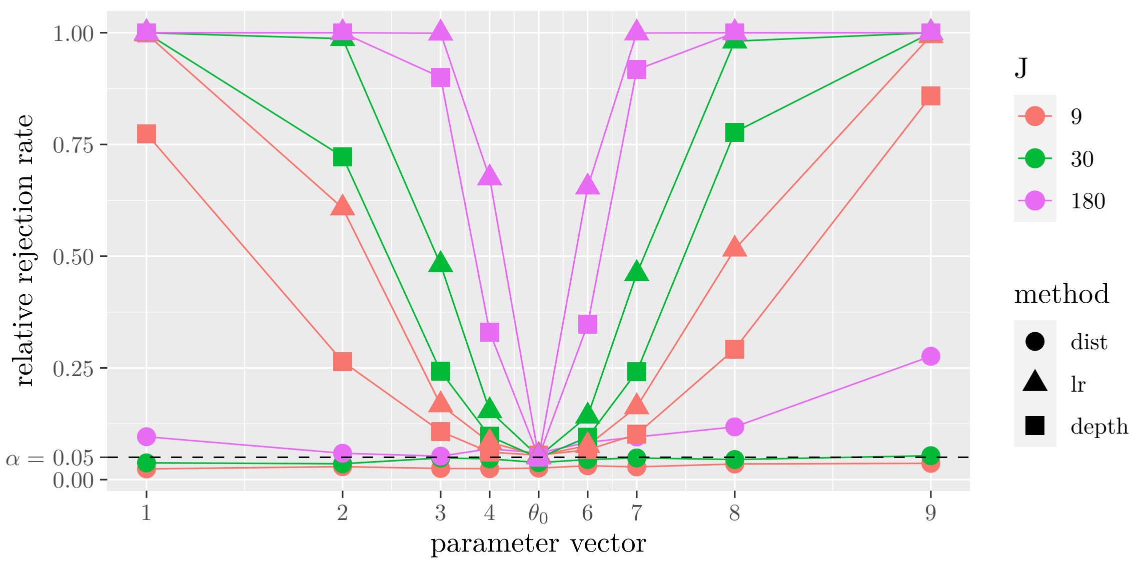


Figure 16: Rejection rates of the level $\alpha = 0.05$ tests for $\mathcal{H}_0 : \theta^* = \theta_0$ based on the minimum distance estimator (method "dist"), the likelihood ratio (method "lr") and the 3-sign depth (method "depth") at the 9 different parameter vectors of Table 6 and $J \in \{9, 30, 180\}$. The scaling of the x-axis was adjusted to reflect the actual distances of the parameter vectors to θ_0.

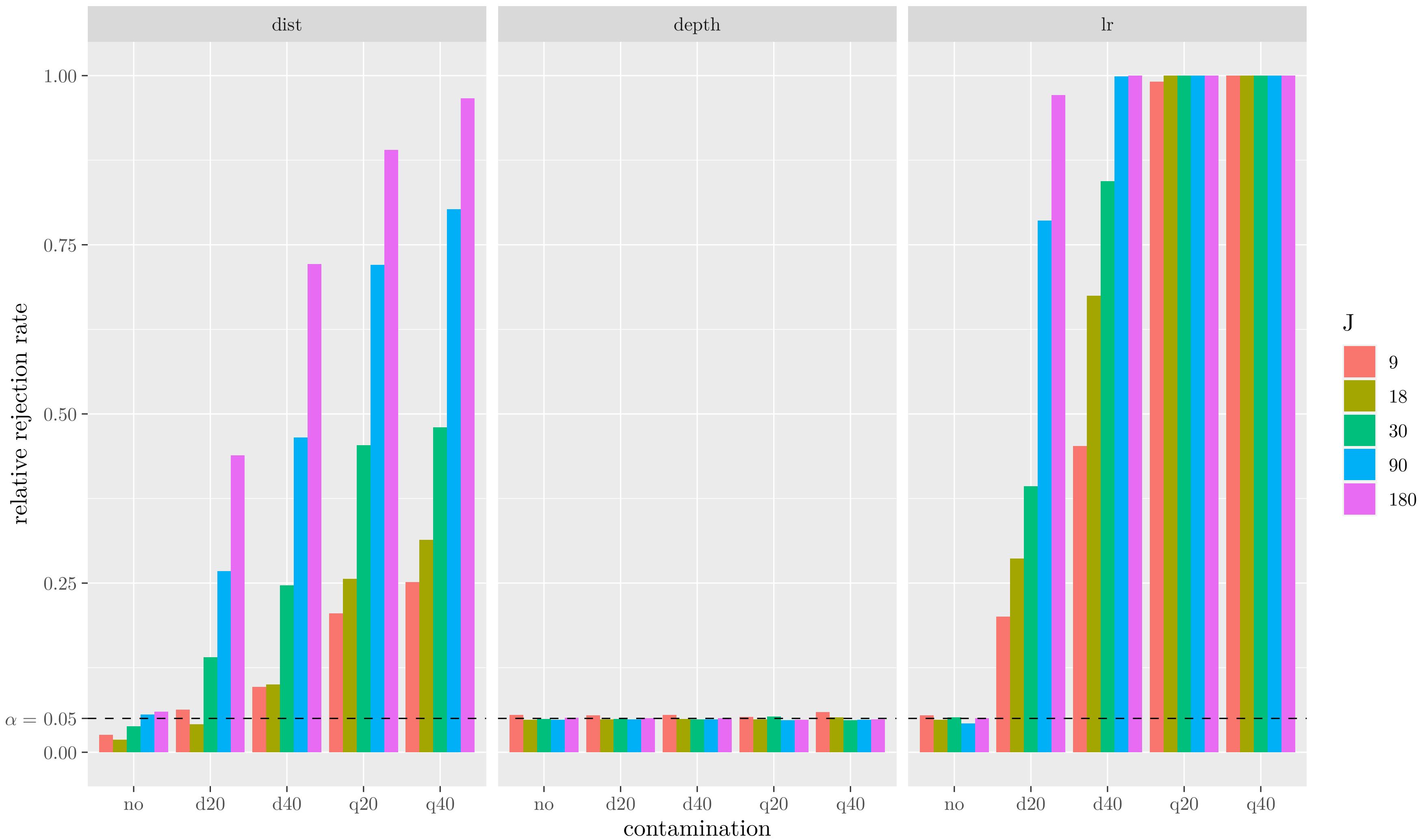


Figure 17: Type I errors of the level $\alpha = 0.05$ tests for $\mathcal{H}_0 : \theta^* = \theta_0$ based on the minimum distance estimator (method “dist”), the 3-sign depth (method “depth”) and the likelihood ratio (method “lr”), grouped by contamination and the sample size $J \in \{9, 18, 30, 90, 180\}$. The bar plot shows the relative rejection rates under $\mathcal{H}_0$ calculated from 5000 repetitions.

All tests nearly meet the level α at θ_0 (the exact values can be found in Tables 9 to 11 in Appendix C). With regard to Section 6.2, this is particularly surprising for the method "dist" and could be attributable to the higher number of repetitions (5000 vs. 2000). Once again, the likelihood ratio test consistently achieves the highest power among the three tests. However, while the 3-sign depth test performed poorly in detecting damage accumulation effects, where the test based on the minimum distance estimator provided acceptable results, the roles are now reversed. This is because the confidence regions $\mathcal{C}_{J,1-\alpha}^{(\text{dist})}$ do not extend far in the θ_3-direction (which makes them suitable for damage accumulation testing), but are comparatively large (which is why the test cannot identify minor deviations from θ_0). The corresponding test only reaches a decent power if the true parameter deviates even further from θ_0 (especially in θ_2- or θ_3-direction) or the sample size J is increased, which - as with the 3-sign depth test for damage accumulation - is beyond the scope of this thesis.
After gaining an impression of the performance of the tests, we move on to the contaminated data. To qualify as robust for contaminated data, a test should meet two criteria:

(i) The test maintains the α level, i.e., the type I error does not increase substantially.

(ii) The test still achieves a reasonable power against the alternative.

We understand these criteria as a qualitative measure of robustness; we do not quantify the concept of robustness in the context of this thesis. To check the first criterion, we look at the type I errors of the tests when applied to the contaminated data, see Figure 17. At any sample size $J \in \{9, 18, 30, 90, 180\}$, the 3-sign depth test preserves its level $\alpha = 0.05$ not only for the depth-specific contamination, but also for the quantile-based

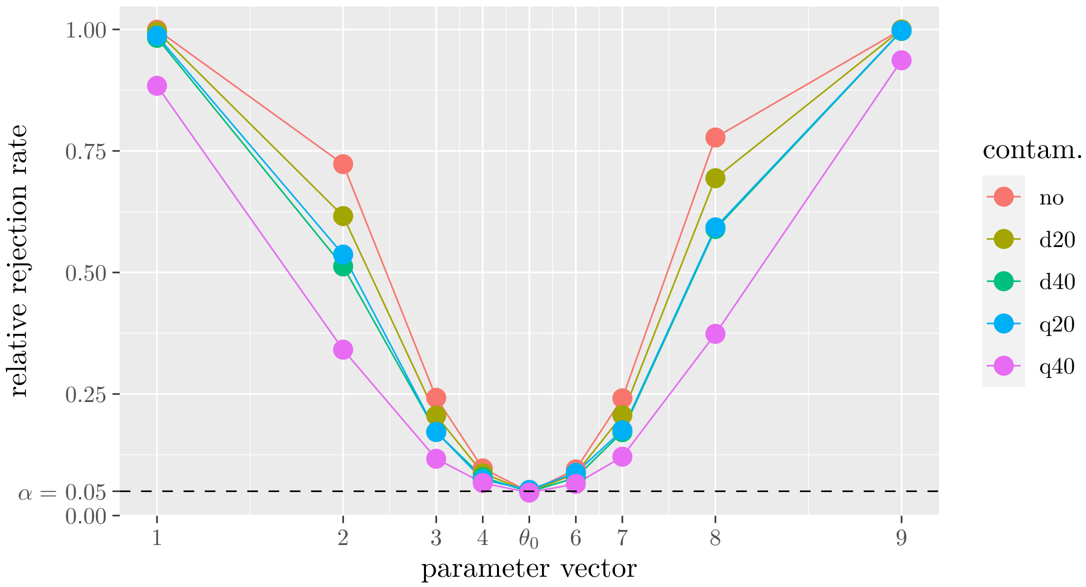


Figure 18: Rejection rates of the level $\alpha = 0.05$ 3-sign depth test with $J = 30$ for $\mathcal{H}_0 : \theta^* = \theta_0$ at the 9 different parameter vectors of Table 6 and for the different types and proportions of contamination given in Table 5. The scaling of the x-axis was adjusted to reflect the actual distances of the parameter vectors to θ_0.

contamination. Meanwhile, the methods "dist" and "lr" cannot keep the level regardless of the type and proportion of contamination. The exceptions are the cases $J \in \{9, 18\}$ at 20% depth-specific contamination in the method "dist", but the full results in Table 9 suggest that this is likely due to the overall low power of the test. Consequently, the 3-sign depth test is the only method to fulfill the first robustness criterion. For the second criterion, we need to assess the extent to which the contamination of the data affects the power of this test. In Figure 18, we plot the power of the 3-sign depth test with $J = 30$ for both the uncontaminated data set and the contaminated data sets d20, d40, q20 and q40 (cf. Table 5). As expected, the sharpest drop in power occurs at 40% quantile-based contamination, whereas the loss in power is the lowest at 20% depth-specific contamination. In general, quantile-based contamination plausibly leads to lower rejection rates than depth-specific contamination, with the test showing similar behavior at d40 and q20. This observation also applies to the other sample sizes $J \in \{9, 18, 90, 180\}$, as demonstrated in Figure 31 in Appendix C.
Despite the lower power, the performance of the test can still be considered satisfactory: With 40% contaminated data, the power of the 3-sign depth test with $J \geq 90$ still shows the desired V-shape, whereas 20% contamination already renders the other methods useless (e.g., the likelihood ratio test with $J = 180$ always rejects the null hypothesis in the data set q20). Overall, we conclude that the 3-sign depth test also satisfies the second robustness criterion and can therefore be deemed robust in the presence of contaminated data.

This finding marks the end of our simulation studies, and also closes the last major chapter of the dissertation. We have seen that the minimum distance estimator is sensitive to the chosen parametrization and does not display the claimed robustness, but is capable of detecting damage accumulation. The 3-sign depth test, on the other hand, proves to be ineffective when testing for damage accumulation, while being robust against contaminated data and providing comparatively small confidence regions for the true parameter. Finally, the likelihood-based approaches consistently achieved the best results as long as we did not deviate from the model assumptions. They therefore perform well in simulation studies, but their practical applicability is hampered by their lack of robustness.

7. Outlook for Future Research

This dissertation significantly contributes to the statistical analysis of intensity-based point process models, especially in the context of load sharing systems with accumulating damage, in two ways: First, we prove that the minimum distance estimator is indeed asymptotically normal distributed as claimed by Kopperschmidt and Stute 2013. Second, we introduce a new procedure to implement the robust 3-sign depth test into a point process framework. In both directions, however, the research is far from complete with the conclusion of this thesis. In this outlook, we would like to point out a few potential paths that future studies on these methods may pursue.

While we consider the theoretical foundations of the minimum distance estimator to be finalized, the derived tests provide much room for improvement. In line with the likelihood ratio test, a test statistic of the form

$$\frac{\inf_{\theta \in \Theta_0} \left\| \overline{N}^{(n)} - \overline{\Lambda}_\theta^{(n)} \right\|_{\overline{N}^{(n)}}}{\inf_{\theta \in \Theta} \left\| \overline{N}^{(n)} - \overline{\Lambda}_\theta^{(n)} \right\|_{\overline{N}^{(n)}}}$$

would be conceivable, where the minimum Cramér-von Mises distance under the null hypothesis $\mathcal{H}_0 : \theta \in \Theta_0$ is related to the minimum distance on the entire parameter space Θ. A simulation study could be the first step towards determining the asymptotic distribution of this test statistic. It is accompanied by the search for an accurate computation of the asymptotic covariance matrix of the minimum distance estimator, which turned out to be infeasible for load sharing models with damage accumulation. To this end, the intensity integrals involved need to be calculated explicitly. This remains an open problem, but the suboptimal performance so far raises doubts as to whether the benefits justify this effort.

The novelty of our depth-related approach to point process inference opens up a multitude of possible research questions. On the example of the *unknown load sharing model* of Kvam and Peña 2005, we established the consistency of the 3-sign depth test for a broad class of intensity-based models. An interesting topic is the extent to which the requirements for consistency can be further weakened. For instance, exploratory simulation studies suggest that ordering the hazard transforms with respect to the essential past of the process - although crucial for the proof of consistency - barely affects the performance of the test. We further conjecture that the consistency of the 3-sign depth test for the significance of damage accumulation (i.e., $\mathcal{H}_0 : \theta_3^* = 0$) extends to any hypothesis of the form

$$\mathcal{H}_0 \, : \, \theta_3^* \in \mathrm{Int} \, ,$$

where Int is a compact interval. Nevertheless, our simulation studies do indicate that the power of this test is likely to be unsatisfactory.

Beyond the methodological groundwork, intensity-based load sharing models are themselves a subject of future research. We introduced a generalization of the Basquin load sharing model with multiplicative damage accumulation proposed by Müller and Meyer 2022, that can be represented as a relative risk regression model. Since we translate any isotonic transformation of the damage accumulation term into a parametric model, we enable the construction of an entire class of intensity-based models including both load sharing and damage accumulation. One objective is to impose conditions on these isotonic transformations under which the properties of the Basquin load sharing model with multiplicative damage accumulation transfer to this class. Another intriguing prospect is to incorporate the isotonic transformation into the intensity-based model as a non-parametric part. Although this adds immense flexibility, it calls for a paradigm shift, as the methods presented in this thesis can no longer be applied in such a semi-parametric framework.

Due to the more theoretical nature of this thesis, we have only scratched the surface as far as simulation studies are concerned. Among others, the following extensions are envisaged:

- Investigate into the influence of the true parameter. How does it affect the size and coverage rate of the studied confidence regions? How do they behave in special cases such as $\theta_3^* = 0$ (i.e., in the absence of damage accumulation) or $\theta_3^* < 0$?
- Consider different values and distributions for the (random) covariates. Do the methods still produce acceptable results if the model assumptions for these covariates are violated?
- Visualize the confidence regions in three dimensions. How are the depth-based regions shaped? Can this knowledge improve the performance of the 3-sign depth test for the significance of damage accumulation?
- Extend the robustness study by looking at further types and fractions of contaminated data. What level of contamination does the 3-sign depth test withstand? Are there other types of contamination where it is less robust?
- Benchmark with a wider range of methods. For example, the Kolmogorov-Smirnov test can be applied to the hazard transforms. How do minimum distance estimator and 3-sign depth test compare to the established methods for point process inference?

Once the methods have been thoroughly evaluated in simulation studies, the next step is to apply them to real data. The broken tension wires in Figure 1 originate from a large-scale test series carried out by the Faculty of Architecture and Civil Engineering at TU Dortmund University, see Szugat et al. 2016 for details. One of the most exciting questions in the wake of this dissertation is whether the effect of damage accumulation is significant in this real data. As the study of Szugat et al. 2016 shows, we are also usually less interested in the unknown model parameter θ^* than in predicting the failure of a load sharing system or its components. Such prediction intervals can be determined from the confidence regions presented in this thesis. However, this is a task for future research, and we look forward to contributing to it by laying the theoretical foundations.

Appendices

A. Comprehensive Introduction to Intensity Theory

This part of the thesis provides the mathematical background for the statistical analysis of intensity-based models. We acquaint ourselves with the concept of compensators, which we will identify as conditional (cumulative) hazard rates or "intensities". Furthermore, the foundation for the depth-related methods of statistical inference is laid by incorporating the hazard transformation of a point process.

The textbook style of this broad introduction is intended to appeal to those who have not yet encountered statistical inference via intensity-based models. Conversely, we encourage readers familiar with stochastic intensities to use this overview as a convenient reference in their study of this thesis. It is divided into five sections: In Section A.1, we learn about simple point processes and their associated counting processes before moving on to filtrations, martingales, and compensators in Section A.2. This in particular covers the intensity theory, which is the topic of Subsection A.2.4. We discuss the hazard transformation in Section A.3, and close this introduction with complementary proofs and explanatory remarks in the Sections A.4 and A.5, respectively.

A.1. Simple Point Processes and Their Associated Counting Processes

In the following we focus on a special type of stochastic processes and its characteristics: the counting process. Therefore, this section is dedicated to the introduction of general stochastic processes and the definition of counting processes and their relatives, the point processes, as well as giving some basic properties. We furthermore discover the duality between simple point processes and their associated counting processes, which justifies the interchangeability of these terms. First we want to state the definition of a general stochastic process.

Definition A.1 (Stochastic Process; Ethier and Kurtz 1986, p. 49)**.**
Let $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and* $(E, \mathcal{E})$ *a measurable space. A* stochastic process X *with* index set $\mathcal{I}$ *and* state space $(E, \mathcal{E})$ *defined on* $(\Omega, \mathcal{F}, \mathbb{P})$ *is a function defined on* $\mathcal{I} \times \Omega$ *with values in* E *such that for each* $t \in \mathcal{I}$,

$$X(t, \cdot) \,:\, \Omega \longrightarrow E$$

is an E*-valued random variable, that is:*

$$\{\omega \in \Omega \,:\, X(t,\omega) \in \mathcal{A}\} \in \mathcal{F}\,, \quad \textit{for all } A \in \mathcal{E}\,.$$

In the context of this work and throughout many applications the abbreviated notation $X = (X_t)_{t\in\mathcal{I}}$ is commonly used. Hereinafter, we will mostly consider as index sets the natural numbers $\mathbb{N}$, compact intervals or - more generally - subsets of the d-dimensional space $\mathbb{R}^d$ for some positive integer $d \in \mathbb{N}$. For $\mathcal{I} \subset \mathbb{R}^d$, the indexing parameter $t \in \mathcal{I}$ often represents time, space, or a combination of these in most of our considerations. The stochastic process X is then called a *discrete-parameter process*, if $\mathcal{I}$ is a countable set, and it is called a *continuous-parameter process* otherwise (Snyder and Miller 1991, p. 24). In the case $d = 1$, we are primarily interested in viewing the stochastic process X as a "random" function of time. Consequently, it is natural to put further restrictions on X

(Ethier and Kurtz 1986, p. 50). If in particular $(E, \mathcal{T})$ is a topological space endowed with its Borel σ-algebra $\mathcal{E} := \mathcal{B}(E) := \sigma(\mathcal{T})$, we can discuss the continuity properties of the process's sample paths (Brémaud 2020, p. 208).
Since we will assume throughout that E is a metric space, which is naturally equipped with the topology induced by the associated metric, from here on $\mathcal{E} = \mathcal{B}(E)$ always denotes the corresponding Borel σ-algebra.

Definition A.2 (Measurability and Continuity of Continuous-Parameter Processes; Ethier and Kurtz 1986, p. 50)**.**
Let $\mathcal{I} \subset \mathbb{R}$ be an interval and $X = (X_t)_{t\in\mathcal{I}}$ a continuous-parameter process. The process X is called measurable *if $X : \mathcal{I} \times \Omega \to E$ is $(\mathcal{B}(\mathcal{I}) \otimes \mathcal{F})$-$\mathcal{E}$-measurable, where $\mathcal{B}(\mathcal{I})$ denotes the Borel σ-algebra on $\mathcal{I}$. We say that X is (almost surely)* continuous *(*right-continuous, left-continuous*), if for (almost) every $\omega \in \Omega$, the* sample path

$$\begin{aligned} X(\cdot, \omega) \,:\, \mathcal{I} &\longrightarrow E \\ t &\longmapsto X(t, \omega) = X_t(\omega) \end{aligned}$$

is continuous (right-continuous, left-continuous)[23]*. Analogously, we say that X is (almost surely)* increasing*, if for (almost) every $\omega \in \Omega$ the sample path $X(\cdot, \omega)$ is increasing.*

We will see later that each counting process has the above properties, in that it is measurable, right-continuous and increasing. For the definition of a counting process we want to follow the approach of Jacobsen 2006 that is also consistent with Daley and Vere-Jones 2003. The sample paths of such a counting process can be described by sequences of points (cf. Daley and Vere-Jones 2003, p. 41), thereby motivating the following definition of so-called simple point processes.

Definition A.3 (Simple Point Process; cf. Jacobsen 2006, p. 9)**.**
Let $t_0 \in \mathbb{R}$. A simple point process *is a stochastic process $T = (T_i)_{i\in\mathbb{N}}$ with index set $\mathbb{N}$ and state space $([t_0, \infty], \mathcal{B}([t_0, \infty]))$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ such that*

(i) $\mathbb{P}(t_0 < T_1 \leq T_2 \leq \ldots) = 1\,,$

(ii) $\mathbb{P}(T_i < T_{i+1}, T_i < \infty) = \mathbb{P}(T_i < \infty)\,, \quad$ *for all* $i \in \mathbb{N}\,,$

(iii) $\mathbb{P}\left(\lim_{i\to\infty} T_i = \infty\right) = 1\,.$

The definition given in Jacobsen 2006 only considers the case $t_0 = 0$, but for many applications this restriction is not required. Accordingly, a less rigorous definition can be found in Daley and Vere-Jones 2003 that encompasses the above notion. Remark that for each i with $\mathbb{P}(T_i < \infty) \neq 0$ the second property of Definition A.3 can equivalently be stated as

$$\mathbb{P}(T_i < T_{i+1} \,|\, T_i < \infty) = 1\,, \quad \text{for all } i \in \mathbb{N}\,.$$

Thus, a simple point process is an almost surely increasing sequence of possibly infinite random variables (property (i)), strictly increasing as long as they are finite (property (ii)) and with almost sure limit ∞ (property (iii)), see Jacobsen 2006, p. 10.

[23] For the definition of the continuity property it is required that E is a topological space, whereas this requirement can be omitted for the definition of measurability.

Remark A.4. Every simple point process can be represented by a step function. To see this we note that for each realization $(t_i)_{i\in\mathbb{N}}$ of the point process $(T_i)_{i\in\mathbb{N}}$ we naturally obtain a measure $\mu_{(t_i)_{i\in\mathbb{N}}}$ by counting the number of points falling into subsets of $[t_0,\infty)$:

$$\mu_{(t_i)_{i\in\mathbb{N}}}(A) = \#\left\{i \in \mathbb{N} \,:\, t_i \in A\right\} , \quad A \subset [t_0,\infty) \ .$$

One usually imposes the restriction that $A \in \mathcal{B}([t_0,\infty))$ in order that we may operate conveniently on the measure induced by the underlying point process (cf. Daley and Vere-Jones 2003, p. 42). Accordingly, $\mu_{(t_i)_{i\in\mathbb{N}}}$ is a counting measure on the σ-algebra $\mathcal{B}([t_0,\infty))$ and therefore completely characterized[24] by its values on the compact intervals $[t_0,t]$ for $t_0 \leq t < \infty$. If we define the function $N \,:\, [t_0,\infty) \to \mathbb{N}_0$ by setting

$$N(t) := \mu_{(t_i)_{i\in\mathbb{N}}}([t_0,t]) = \#\left\{i \in \mathbb{N} \,:\, t_0 \leq t_i \leq t\right\} , \tag{A.1}$$

the resulting step function determines $\mu_{(t_i)_{i\in\mathbb{N}}}(A)$ for all Borel sets $A \in \mathcal{B}([t_0,\infty))$ and hence describes the realization $(t_i)_{i\in\mathbb{N}}$ of the point process T.

The observation of Remark A.4 is incorporated in the upcoming definition, utilizing the former Definition A.3 of a simple point process.

Definition A.5 (Counting Process Associated With a Simple Point Process; cf. Jacobsen 2006, pp. 11–12)**.**
Let $t_0 \in \mathbb{R}$ and $T = (T_i)_{i\in\mathbb{N}}$ be a simple point process with state space $([t_0,\infty], \mathcal{B}([t_0,\infty]))$ defined on $(\Omega,\mathcal{F},\mathbb{P})$. The counting process *associated with T is the $\mathbb{N}_0$-valued continuous-parameter process $N = (N_t)_{t\in[t_0,\infty)}$ with state space[25] $(\mathbb{R},\mathcal{B}(\mathbb{R}))$, where*

$$N_t := \sum_{i=1}^{\infty} \mathbb{1}_{(-\infty,t]}(T_i) = \#\{i \in \mathbb{N} \,:\, T_i \leq t\} \,. \tag{A.2}$$

For any set $A \in \mathcal{F}$ the indicator function $\mathbb{1}_A$ *is defined as*

$$\mathbb{1}_A \,:\, \Omega \longrightarrow \{0,1\} \;\; \omega \longmapsto \mathbb{1}_A(\omega) = \begin{cases} 1\,, & \omega \in A\,, \\ 0\,, & \omega \notin A\,. \end{cases}$$

Thus N_t counts the number of events until time t and is therefore by definition an increasing, right-continuous stochastic process, as the right-continuity of the process's sample paths is inherited from the indicator function. Occasionally the notation $N(t)$ is used instead of N_t, reflecting Remark A.4, where the duality between a point process and a step function was first indicated. In fact, the sequence $(T_i)_{i\in\mathbb{N}}$ is easily recovered from $(N_t)_{t\in[t_0,\infty)}$ since $\mathbb{P}$-almost surely it holds:

$$T_i = \inf\{t \geq t_0 \,:\, N_t \geq i\}\,, \tag{A.3}$$

where here as elsewhere we define $\inf\emptyset = \infty$ (Jacobsen 2006, p. 12).

[24] This works in the same manner as a cumulative distribution function determines a probability measure on Borel sets (Daley and Vere-Jones 2003, p. 51).

[25] To discuss the continuity properties of counting processes we refrain from using $\left(\mathbb{N}_0, 2^{\mathbb{N}_0}\right)$ as state space, where $2^{\mathbb{N}_0}$ denotes the power set of $\mathbb{N}_0$. Moreover, we choose $E = \mathbb{R}$ instead of $E = [0,\infty]$ for simplicity.

In particular, the aforementioned duality is expressed through the equivalence

$$T_i \leq t \quad \Longleftrightarrow \quad N_t \geq i \,. \tag{A.4}$$

By virtue of Equations (A.3) and (A.4), the notions of *point process* and *counting process* are often used interchangeably. Accordingly, the counting process is commonly referred to as point process (e.g., in Daley and Vere-Jones 2003 and Kopperschmidt and Stute 2013).

Remark A.6. We can easily recognize how the properties that define a simple point process are transferred to the associated counting process. Let again $T = (T_i)_{i\in\mathbb{N}}$ denote a simple point process and $N = (N_t)_{t\in[t_0,\infty)}$ its associated counting process.

(i) The first condition provides $N_{t_0} = 0$ almost surely. However, the order condition is negligible for the definition of the associated counting process.
This is especially relevant in case of multidimensional generalizations of simple point processes, where there might be no natural order on the point process's state space, e.g. for $E \subset \mathbb{R}^d$ with $d > 1$. Here the associated counting process can once again be defined in a similar fashion as the cumulative distribution function, compare Equation (A.1) from Remark A.4.

(ii) The second condition ensures that almost surely the sample paths of the process increase only in jumps of size 1.
Accordingly, for $\mathbb{P}$-almost all $\omega \in \Omega$ it holds that $\Delta N_t(\omega) \in \{0, 1\}$ for all t, where for any *càdlàg*[26] function f the function Δf is defined via (cf. Jacobsen 2006, p. 12)

$$t \longmapsto \Delta f(t) := f(t) - f(t-) := f(t) - \lim_{s\uparrow t} f\left(s\right) \,.$$

A formal proof of this statement can be found in Appendix A.4.

(iii) The third condition is needed to prevent a so-called explosion of the process, as retaining only conditions (i) and (ii) from Definition A.3 allows $N_t = \infty$ to occur with probability > 0 (Jacobsen 2006, pp. 10–12). The inclusion of condition (iii) thus ensures that, for $\mathbb{P}$-almost all sample paths, only finitely many jumps can occur in finite time.
Note that in order to drop condition (iii), we would have to consider the completion $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$ of $\mathbb{R}$ as state space, whereby the definition of a counting process could easily be extended to the index set $[t_0, \infty]$. This extension is widely discussed in Jacobsen 2006, but does not bring any additional benefit for our purposes.

To summarize the above remarks, one could equivalently define a counting process as a $\mathbb{N}_0$-valued continuous-parameter process on $[t_0, \infty)$ whose sample paths $t \mapsto N_t(\omega)$ are required to be $\mathbb{P}$-almost surely (i.e., for $\mathbb{P}$-almost all $\omega \in \Omega$) right-continuous and satisfy $N_{t_0}(\omega) = 0$ as well as $\Delta N_t(\omega) \in \{0, 1\}$ for all t.
Nevertheless, this approach appears less understandable than the supposed detour via the definition of simple point processes. Moreover, the above duality would remain unmentioned, so that the interchangeability of the terms *counting process* and *(simple) point process* could not be motivated. Another term that should not remain unmentioned in conjunction with point processes is that of "interarrival times". We illuminate this

[26] Abbreviation of the French term *continue à droite, limite à gauche*, i.e. right-continuous with left limits.

notion in the following example, which at the same time introduces perhaps the best known representative of point processes: the (homogeneous) Poisson process.

Example A.7 (Homogeneous Poisson Process; cf. Snyder and Miller 1991, pp. 41, 57). Let $N = (N_t)_{t \geq 0}$ be a counting process with the following properties[27]:

(i) For each $0 \leq s < t$, the increment $N(s,t) := N_t - N_s$ is Poisson distributed with parameter $\lambda(t-s)$, where $\lambda \in (0, \infty)$ is a positive constant:

$$\mathbb{P}\left(N(s,t) = n\right) = \frac{1}{n!}\left(\lambda(t-s)\right)^n \mathrm{e}^{-\lambda(t-s)}\,.$$

(ii) N has indepent increments, that is, for any finite collection of times $0 = t_0 \leq t_1 < t_2 < \ldots < t_j$, $j \in \mathbb{N}$, the increments

$$N\left(t_{i-1}, t_i\right) = N_{t_i} - N_{t_{i-1}}\,, \quad 1 \leq i \leq j\,,$$

are stochastically independent.

Then N is called a *homogeneous Poisson process with intensity* λ. Since $\mathbb{E}\left(N(s,t)\right) = \lambda(t-s)$, λ equals the average density of points.

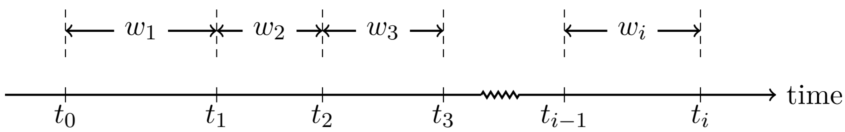


Figure 19: Illustration of realized interarrival times $w_i = W_i(\omega)$.

If $T = (T_i)_{i \in \mathbb{N}}$ denotes the simple point process associated with N, by setting $T_0 \equiv t_0$ we can define the *interarrival times* through

$$W_i := T_i - T_{i-1}\,, \quad i \in \mathbb{N}\,, \tag{A.5}$$

see Figure 19 for a visualization. The interarrival times are independent and identically distributed with the common distribution being exponential with parameter λ (see Snyder and Miller 1991, pp. 53–58 for a proof of this statement and further properties),

$$W_1, W_2, \ldots \overset{\text{i.i.d}}{\sim} \mathcal{E}(\lambda)\,. \tag{A.6}$$

Although the Poisson process serves merely as an introductory example at this point, we will recognize that it plays a crucial in our methods of statistical inference[28]. Note that the Poisson process is covered only superficially here, as it is more of a tool for us

[27] Snyder and Miller 1991 also require that $\mathbb{P}\left(N_0 = 0\right) = 1$ holds, whereas this property is part of our definition of a counting process, see Remark A.6 (i) on this issue.

[28] As discussed in Protter 2005, pp. 14–16, this is mainly due to the Poisson process being the *only* counting process with "stationary increments indepent of the past" within our framework.

than an actual subject of consideration. For a comprehensive account of Poisson processes and their generalizations, we encourage the reader to study Chapters 2 and 3 of Daley and Vere-Jones 2003.
In the following, we will occasionally deal with counting processes observed only on a compact subinterval $\left[t_0, t^0\right] \subset [t_0, \infty)$ of the index set. This can be achieved by restricting a counting process; conversely, any $\mathbb{N}_0$-valued continuous-parameter process on $\left[t_0, t^0\right]$ whose sample paths are almost surely right-continuous, start in 0 and increase only by jumps of size 1 can be extended to a counting process on $[t_0, \infty)$. We record this basic property in a remark.

Remark A.8 (Restriction and Extension of a Counting Process).
A stochastic process $N = (N_t)_{t \in I}$ is considered a counting process on an interval $\mathcal{I} \subset [t_0, \infty)$ if there exists a counting process $\tilde{N} = (\tilde{N}_t)_{t \in [t_0, \infty)}$ such that

$$N = \tilde{N}\big|_{\mathcal{I}} . \tag{A.7}$$

Conversely, if N is a counting process on the compact interval $\left[t_0, t^0\right]$, then one can always obtain a counting process $\tilde{N}$ with index set $[t_0, \infty)$ satisfying Equation (A.7) by setting

$$\tilde{N}_t := N_{\min\{t, t^0\}} , \quad t \geq t_0 ,$$

that is, through a constant extension of the process.

We close the current section with a proposition stating the measurability of right-continuous processes, including counting processes associated with a simple point process.

Proposition A.9 (Measurability of Right-Continuous Stochastic Processes).
Let $X = (X_t)_{t \in \mathcal{I}}$ be a right-continuous stochastic process, where $\mathcal{I} \subset \mathbb{R}$ is an interval and the state space is $\left(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right)\right)$ for some integer d. Then, X is measurable.

Proof. The proof is based on Brémaud 2020, where a more general result on progressive measurability - a negligible property related to the filtrations introduced in Section A.2 - is shown (cf. Theorem 5.3.8, Brémaud 2020, p. 214). The adapted proof can be found in Appendix A.4. □

In the following section, we dive into the theory of filtrations, martingales, and compensators, but return regularly to counting processes by showing their connections to the subsequently introduced concepts.

A.2. Filtrations, Martingales and Compensators

Besides stating the basic definitions, we present a concise overview of important results from martingale and intensity theory. The intimate link between these fields of stochastics is embodied in the *Doob-Meyer decomposition for non-negative submartingales*: This central theorem allows for the decomposition of a counting process N into a predictable part - the compensator of N - and a martingale, which serves as an innovation process conveying all information about N not derivable from the strict past (cf. Karr 1991). The compensator proves to be vital for the intensity-based modeling of counting processes, since it can be represented as an integrated intensity process.
The following major theorems are taught in courses on probability theory and are therefore

not proved here. However, we always provide comprehensive bibliographical references to facilitate the understanding of the technically demanding proofs.
Note that we will not elaborate on the theory of stopping times and therefore may not capture the full extent of some results regarding local martingales. Nevertheless, the concepts covered are quite sufficient for providing the framework of our statistical inference, hence this compromise of completeness and conciseness is made.

A.2.1. Filtrations and the Usual Conditions

We start by introducing the concept of a *filtration* and explain why it is commonly referred to as *history*. Subsequently, the *usual conditions* that are often implicitly imposed on the filtrations under consideration are discussed in the context of counting processes.
Since the following definitions can be found in practically every textbook on probability theory, the cited references serve only as a rough orientation on where to find a notation consistent with the one used here.

Definition A.10 (Filtration; cf. Bauer 1996, p. 133 and Jacobsen 2006, p. 301)**.**
Let $\mathcal{F}$ be a σ-algebra and $\mathcal{I} \subset \mathbb{R}$. A collection $\{\mathcal{F}_t\}_{t\in\mathcal{I}}$ of sub-σ-algebras of $\mathcal{F}$ is called a filtration of $\mathcal{F}$*, if*

$$\mathcal{F}_s \subset \mathcal{F}_t\,, \quad \textit{for all } s,t \in \mathcal{I} \textit{ with } s \leq t\,. \tag{A.8}$$

If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $\{\mathcal{F}_t\}_{t\in\mathcal{I}}$ is a filtration of $\mathcal{F}$, then $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in\mathcal{I}}, \mathbb{P})$ is called a filtered probability space*.*

Definition A.11 (Natural Filtration and Adapted Process; cf. Ethier and Kurtz 1986, p. 50)**.**
Let $\mathcal{I} \subset \mathbb{R}$ and $X = (X_t)_{t\in\mathcal{I}}$ be a continuous-parameter process with state space $(E, \mathcal{E})$ defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in\mathcal{I}}, \mathbb{P})$. X is adapted *with respect to the filtration $\{\mathcal{F}_t\}_{t\in\mathcal{I}}$, if X_t is $\mathcal{F}_t$-$\mathcal{E}$-measurable for each $t \in \mathcal{I}$. The* natural filtration *of X is defined by*

$$\mathcal{F}_t^X := \sigma\left(\{X_s \,:\, s \in \mathcal{I}, s \leq t\}\right)\,, \quad t \in \mathcal{I}\,.$$

Remark A.12 (Internal History and General Remarks on Filtrations)**.**

(i) Each stochastic process is adapted with respect to its natural filtration.

(ii) We say abbreviatively that "X is adapted" when it is clear from the context which filtration is being referred to. In the case of adapted processes on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in\mathcal{I}}, \mathbb{P})$, this is always the underlying filtration $\{\mathcal{F}_t\}_{t\in\mathcal{I}}$, unless otherwise specified.

(iii) Note that to state Equation (A.8) it is only required that $\mathcal{I}$ is an ordered set (cf. Bauer 1996, p. 133). Nevertheless, throughout many textbooks only the more restrictive case $\mathcal{I} = [0, \infty)$ is considered (e.g., Ethier and Kurtz 1986, p. 50 and Jacobsen 2006, p. 301). However, this coincides with our choice of $\mathcal{I} \subset \mathbb{R}$, since we will mainly consider counting processes - and their associated natural filtrations - with index sets $[t_0, \infty)$, where frequently $t_0 = 0$ is assumed.

(iv) The natural filtration $\left\{\mathcal{F}_t^X\right\}_{t\in\mathcal{I}}$ consists of the history of X but does not encompass any further external information. As a consequence, the natural filtration is commonly referred to as *internal history* of the stochastic process. For X to be adapted w.r.t. a filtration $\{\mathcal{F}_t\}_{t\in\mathcal{I}}$ it is required that

$$\mathcal{F}_t \supset \mathcal{F}_t^X$$

is satisfied for each $t \in \mathcal{I}$. Therefore, each such filtration is called a *history* of X in that it comprises the internal history of X but may contain additional information. While we refrain from using this term, it is synonymous with filtrations in many textbooks, which may be motivated by the above context (e.g., Brémaud 2020, p. 213 or Daley and Vere-Jones 2008, p. 357).

By virtue of Remark A.12 (iv), a filtration $\{\mathcal{F}_t\}_{t\geq t_0}$ is often associated with the information available at time t. Since we are particularly interested in adapted counting processes, the natural filtration $\{\mathcal{F}_t^N\}_{t\geq t_0}$ of such a process N plays a crucial role for us. In this case, however, $N_{t_0} = 0$ implies[29]

$$\mathcal{F}_{t_0}^N = \{\emptyset, \Omega\} \ , \tag{A.9}$$

so no information is available at time t_0, which does not accurately reflect reality:
For example, when conducting an experiment, one can expect to already have preliminary information on events that can influence its outcome, even if they remain unchanged over the course of the experiment. This information can be accounted for by initial conditioning on a non-trivial sub-σ-algebra of $\mathcal{F}$, with the most prominent case being randomized experimental conditions that are specified in advance. However, any information not initially present at time t_0 should be limited to arise from observing the experiment (i.e., the stochastic process). The above perception of external prior information leads to the consideration of intrinsic filtrations, which provide sensible augmentations of the natural filtration. They are usually referred to as intrinsic histories to emphasize their relationship to the internal history. We are already giving their definition here, although we will not experience their benefits until later.

Definition A.13 (Intrinsic Filtration; cf. Daley and Vere-Jones 2003, p. 357)**.**
Let $X = (X_t)_{t\geq t_0}$ be a continuous-parameter process defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ *and let $\{\mathcal{F}_t^X\}_{t\geq t_0}$ denote its natural filtration. Any filtration $\{\mathcal{F}_t\}_{t\geq t_0}$ satisfying*

$$\mathcal{F}_t = \mathcal{F}_t^X \vee \mathcal{G}_0 := \sigma\left(\mathcal{F}_t^X \cup \mathcal{G}_0\right) \ , \quad \textit{for all } t \geq t_0 \, ,$$

is called an intrinsic filtration *of X, where $\mathcal{G}_0 \subset \mathcal{F}$ is a (generally non-trivial) sub-σ-algebra of $\mathcal{F}$. Note that we introduced the $\vee$-operator for the smallest σ-algebra containing the union of the given σ-algebras, which normally is not a σ-algebra itself.*

Although sometimes embedded in the definition of a filtration itself, we now want to impose further constraints on filtered probability spaces that allow for convenient modifications of the stochastic processes under consideration, namely martingales and sub-martingales. In particular, the *usual conditions* - also known as *"les conditions habituelles*

[29]More generally, we only require $\mathbb{P}(N_{t_0} = 0) = 1$, so the σ-algebra generated by N_{t_0} consists only of events with probability 0 or 1. Nevertheless, the interpretation remains that there is no essential information available at time t_0.

(de la théorie générale du processus)"[30] - will ensure that, under mild assumptions, any (sub)martingale permits a càdlàg modification (Liptser and Shiryaev 2001, pp. 57–60 and Karatzas and Shreve 1988, pp. 16–17). As an essential precondition of the Doob-Meyer decomposition for non-negative submartingales, these constraints thereby provide the right-continuity of the predictable compensator (see Theorem A.23 further below for details). Hence, we will encounter the usual conditions primarily in the more theoretical parts of this thesis, since in practice the continuity of the compensator will be presumed for application-oriented modeling.

Definition A.14 (Usual Conditions; Jacobsen 2006, p. 301)**.**
Let $t_0 \in \mathbb{R}$. A filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq t_0}, \mathbb{P})$ satisfies the usual conditions*, if the following requirements are fulfilled:*

(i) The probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is complete*, that is,*

$$A_0 \in \mathcal{F}\,, \mathbb{P}\left(A_0\right) = 0\,, A \subset A_0 \quad \Longrightarrow \quad A \in \mathcal{F}\,.$$

(ii) For $\mathcal{N} = \{A \in \mathcal{F} : \mathbb{P}(A) = 0\}$ (i.e., $\mathcal{N}$ is the collection of $\mathbb{P}$-null sets in $\mathcal{F}$) it holds[31]

$$\mathcal{N} \subset \mathcal{F}_{t_0}\,.$$

(iii) The filtration $\{\mathcal{F}_t\}_{t \geq t_0}$ is right-continuous*, that is,*

$$\mathcal{F}_t = \mathcal{F}_{t+} := \bigcap_{s>t} \mathcal{F}_s\,, \quad \textit{for all } t \geq t_0\,.$$

The conditions (i) and (ii) from Definition A.14 are closely related and often combined into one condition: Condition (i) postulates that subsets of $\mathbb{P}$-null sets from $\mathcal{F}$ are included in $\mathcal{F}$ as well, while condition (ii) ensures that condition (i) also holds for $(\Omega, \mathcal{F}_t, \mathbb{P})$ given any $t \geq t_0$. Furthermore, condition (i) proves to be a minor technical requirement that can always be met without affecting the underlying probabilistic structure.

Lemma A.15 (Completion of a Probability Space; Jacobsen 2006, p. 301)**.**
Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Then there exists a complete probability space $\left(\Omega, \overline{\mathcal{F}}, \overline{\mathbb{P}}\right)$ with $\mathcal{F} \subset \overline{\mathcal{F}}$ and

$$\overline{\mathbb{P}}(A) = \mathbb{P}(A)\,, \quad \textit{for all } A \in \mathcal{F}. \tag{A.10}$$

$\left(\Omega, \overline{\mathcal{F}}, \overline{\mathbb{P}}\right)$ is called the completion *of $(\Omega, \mathcal{F}, \mathbb{P})$.*

Proof. The essence of the proof is given in Jacobsen 2006, p. 301. A slightly more detailed elaboration can be found in Appendix A.4, albeit the completion of the elementary yet lengthy proof is left to the reader. Note that while the completion in the above sense is not uniquely determined, this can be achieved by restricting oneself to the smallest possible sigma algebra $\overline{\mathcal{F}}$. ☐

[30] This designation goes back to Dellacherie 1972, but is widely used throughout the literature, see Andersen et al. 1993, p. 60 and Karr 1991, pp. 59,415.
[31] While $\mathcal{N}$ depends on $\mathcal{F}$ and $\mathbb{P}$, this dependency is neglected in favor of a shorter notation.

In the previous section, the right-continuity and therefore measurability of counting processes followed directly from their inherent characteristics. Consequently, these properties could instead be incorporated in the definition of counting processes, as stated in the brief summary of Remark A.6. Similarly, the usual condition formulated in Definition A.14 (iii) arises naturally from the characteristics of a counting process's internal history.

Lemma A.16 (Right-Continuity of the Internal History of a Counting Process; Protter 2005, p. 16)**.**
Let $N = (N_t)_{t \in [t_0, \infty)}$ *be a counting process. Then the internal history* $\left\{ \mathcal{F}_t^N \right\}_{t \geq t_0}$ *of* N *is a right-continuous filtration.*

Proof. Even though the proof stated in Protter 2005, p. 16 is basic, it broadens our understanding of the interrelation between the properties of a continuous-parameter process and its natural filtration. For the sake of completeness, it can be found in Appendix A.4. □

While conditions (i) and (iii) of Definition A.14 are easily satisfiable in the context of counting processes according to Lemmata A.15 and A.16, the same is not necessarily true for the remaining condition (ii). Although we can plainly transition to the completed filtration $\{\overline{\mathcal{F}}_t\}_{t \geq t_0}$ by successive use of Lemma A.15, this does not guarantee that $\mathcal{N} \subset \overline{\mathcal{F}}_{t_0}$ holds. Additionally, if one mimics the proof of Lemma A.15 and thereby enforces $\mathcal{N} \subset \overline{\mathcal{F}}_{t_0}$, then Lemma A.16 no longer applies[32]. Therefore, a trade-off is made between showing the validity of condition (ii) or condition (iii), so that the remaining condition must be assumed to be valid. Heuristically speaking, accepting the inclusion of $\mathbb{P}$-null sets does not provide any additional information, which justifies our approach to assume condition (ii) is satisfied. It will turn out that the usual conditions are seemingly redundant for practical use in intensity-based modeling, since stronger (albeit natural) assumptions will be employed, see Remarks A.24 and A.36 below. However, this first requires knowledge about martingales and compensators, which will be acquired in the course of this section.

A.2.2. Martingales and Predictable Processes

Martingales are well known in mathematics for formalizing the notion of a fair game, but the scope of martingale theory extends beyond that to describing the purely random part of a compensated process obtained by subtracting its systematic part (Brémaud 2020, p. 495 and Andersen et al. 1993, p. 46). Indeed, the standard examples of continuous-time martingales are the Brownian motion and the compensated Poisson process, which is a recurring example that accompanies us throughout this thesis (cf. Karatzas and Shreve 1988, p. 11). Nevertheless, martingales are naturally linked to increasing information patterns, namely filtrations, through the concept of conditional expectations, which becomes evident in studying Definition A.17 below (cf. Brémaud 1981, pp. 3–4). After extending our view to Sub- and Supermartingales, we will eventually shed some light on a class of processes complementary to martingales, the predictable processes (Andersen et al. 1993, p. 65). Along the way, we will always emphasize the relevance of these notations in terms of counting processes.

[32]Deploying the concept of the *usual augmentation*, this operation indeed yields an intrinsic filtration satisfying the usual conditions, see Dellacherie and Meyer 1978, p. 115 for details. Yet, the above proof does not suffice to prove this statement.

Definition A.17 (Continuous-Time Martingale, Brémaud 1981, p. 4)**.**
Let $X = (X_t)_{t\in\mathcal{I}}$ *be a continuous-parameter process defined on a filtered probability space* $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in\mathcal{I}}, \mathbb{P})$ *for some interval* $\mathcal{I} \subset \mathbb{R}$*. The process*[33] X *is said to be a* $(\mathbb{P}, \{\mathcal{F}_t\}_{t\in\mathcal{I}})$*-*martingale*, if the following conditions are fulfilled:*

(i) X *is adapted with respect to the filtration* $\{\mathcal{F}_t\}_{t\in\mathcal{I}}$*.*

(ii) For each $t \in \mathcal{I}$*,* X_t *is* $\mathbb{P}$*-integrable (i.e.,* $\mathbb{E}(|X_t|) < \infty$*).*

(iii) For all $s, t \in \mathcal{I}$ *with* $s \leq t$*, we have* $\mathbb{P}$*-almost surely:* $\mathbb{E}(X_t \,|\, \mathcal{F}_s) = X_s$*.*

If in condition (iii) the equality sign "=" is replaced by $\geq$ *or* $\leq$*, then* X *is called a* submartingale *or* supermartingale*, respectively.*

The following lemma can be immediately deduced from this definition.

Lemma A.18 (Counting Processes as Non-Negative Submartingales)**.**
Let $N = (N_t)_{t\geq t_0}$ *be an adapted counting process defined on the filtered probability space* $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq t_0}, \mathbb{P}\right)$*. If* N_t *is* $\mathbb{P}$*-integrable for each* $t \geq t_0$*, then* N *is a non-negative* $\left(\mathbb{P}, \{\mathcal{F}_t\}_{t\geq t_0}\right)$*-submartingale.*

Proof. By Equation (A.2) from Definition A.5, every counting process is non-negative by default. Assuming that condition (ii) of Definition A.17 holds, it suffices to show that for each $t \geq s \geq t_0$ we have $\mathbb{E}(N_t \,|\, \mathcal{F}_s) \geq N_s$ $\mathbb{P}$-almost surely. Recall that by construction the sample paths of N are increasing, and the monotonicity of the conditional expectation yields:

$$\mathbb{E}(N_t \,|\, \mathcal{F}_s) \geq \mathbb{E}(N_s \,|\, \mathcal{F}_s) = N_s \,,$$

where the last equation holds since N is adapted with respect to the filtration $\{\mathcal{F}_t\}_{t\geq t_0}$. □

The following example can be stated in a more general form, but serves in foreshadowing the upcoming concepts from intensity theory.

Example A.19 (Compensated Poisson Process; Karatzas and Shreve 1988, p. 12).
Let $N = (N_t)_{t\geq 0}$ be a homogeneous Poisson process with intensity λ, see Example A.7. The *compensated Poisson process* is defined as

$$M_t := N_t - \lambda t \,, \quad t \geq 0 \,. \tag{A.11}$$

Then, $(M_t)_{t\geq 0}$ is a $\left(\mathbb{P}, \{\mathcal{F}_t^N\}_{t\geq 0}\right)$-martingale. A proof of this minor statement can be found in Appendix A.4.

The essential contribution of the Doob-Meyer decomposition is that *any* adapted counting process can be uniquely compensated in a fashion similar to Equation (A.11). Before we can formulate this result, we must first familiarize ourselves with the concept of predictability.

[33]Sometimes the term *filtered process* - denoted with $(X_t, \mathcal{F}_t)_{t\in\mathcal{I}}$ - is used here, if only the considered filtration is to be indicated and the underlying probability space is negligible, cf. Karatzas and Shreve 1988, p. 11.

Definition A.20 (Predictable σ-Algebra and Processes; Daley and Vere-Jones 2003, p. 425)**.**
Let $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq t_0}, \mathbb{P}\right)$ *be a filtered probability space. The* $\{\mathcal{F}_t\}_{t\geq t_0}$-predictable σ-algebra *over* $[t_0, \infty) \times \Omega$ *is defined as*

$$\mathcal{P}\left(\{\mathcal{F}_t\}_{t\geq t_0}\right) = \sigma\left(\{(s,t] \times A : t_0 \leq s \leq t, A \in \mathcal{F}_s\}\right),$$

that is, the sub-σ-algebra of $\mathcal{B}([t_0,\infty)) \otimes \mathcal{F}$ generated by the rectangles of the form[34]

$$(s,t] \times A, \quad t_0 \leq s \leq t, A \in \mathcal{F}_s. \tag{A.12}$$

A continuous-parameter process $X = (X_t)_{t\geq t_0}$ with state space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is said to be $\{\mathcal{F}_t\}_{t\geq t_0}$-predictable *when it is* $\mathcal{P}\left(\{\mathcal{F}_t\}_{t\geq t_0}\right)$*-$\mathcal{B}(\mathbb{R})$-measurable.*

Remark A.21 (Alternative Characterizations of Predictability)**.**
While Definition A.20 is considered a classical definition of predictability, a variety of equivalent characterizations exists. We present only a short selection that is tailored to our needs, but all these relevant to us can be found in the concise standard reference Dellacherie and Meyer 1978, pp. 121–126.

(i) The most frequently used characterization besides the one given in Definition A.20 is based on left-continuous, adapted processes (occasionally also requiring the existence of right limits). More precisely, we have:

$$\begin{aligned}\mathcal{P}\left(\{\mathcal{F}_t\}_{t\geq t_0}\right) &= \sigma\left(\{X : X \text{ is adapted and left-continuous.}\}\right)\\ &= \sigma\left(\{X : X \text{ is adapted and left-continuous with right limits.}\}\right)\\ &= \sigma\left(\{X : X \text{ is adapted and continuous.}\}\right).\end{aligned}$$

As a consequence, any adapted (left-)continuous process is predictable.

(ii) Another concept closely interwoven with predictability is that of "naturalness", which provides a martingale characterization of predictability. An integrable, increasing, right-continuous process X that is adapted with respect to $\{\mathcal{F}_t\}_{t\geq t_0}$ is called *natural* if for every bounded càdlàg $\left(\mathbb{P}, \{\mathcal{F}_t\}_{t\geq t_0}\right)$-martingale $(M_t)_{t\geq t_0}$ we have

$$\mathbb{E}\left[\int_0^t M_s \,\mathrm{d}X_s\right] = \mathbb{E}\left[\int_0^t M_{s-} \,\mathrm{d}X_s\right],$$

where here the Lebesgue-Stieltjes integral is considered. However, for such adapted, integrable, increasing, right-continuous processes, the notions of predictability and naturalness coincide (Karr 1991, p. 416). The whole Section III.8 of Protter 2005 is devoted to proving this surprising equivalence that is due to Doléans-Dade 1968.

We already noticed that given the usual conditions, any martingale permits a càdlàg modification, whereas the archetypal predictable process is càglàd[35] by virtue of Remark A.21 (i). This reinforces that the terminology of "predictability" is well chosen, since

[34]Instead of the traditional set of generators given in Equation (A.12), one could use the simpler set $\{(s,\infty) \times A : s \geq t_0, A \in \mathcal{F}_s\}$, see Brémaud 1981, p. 9.
[35]Abbreviation of the French term *continue à gauche, limite à droite*, i.e. left-continuous with right limits.

by left-continuity the value of the process at time t is fixed just before the time itself. Moreover, there is an orthogonality between martingales and predictable processes: a (finite variation) process that is both a (local) martingale and predictable is indistinguishable from a constant process (Andersen et al. 1993, p. 66, also see Kruglov 2016 for a collection of proofs on this topic). Further justification of the term "predictability" is given by the following lemma, where predictability is considered in lieu of left-continuity.

Lemma A.22 (Theorem 3.10 of Liptser and Shiryaev 2001, p. 74)**.**
Let $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq t_0}, \mathbb{P}\right)$ be a filtered probability space satisfying the usual conditions of Definition A.14 and define[36]

$$\mathcal{F}_{t_0-} := \mathcal{F}_{t_0}\,, \quad \mathcal{F}_{t-} := \sigma\left(\bigcup_{t_0\leq s<t} \mathcal{F}_s\right)\,, \quad \textit{for all } t > t_0\,.$$

Then, any $\{\mathcal{F}_t\}_{t\geq t_0}$-predictable process is $\{\mathcal{F}_{t-}\}_{t\geq t_0}$-adapted.

Proof. A detailed proof is given in Liptser and Shiryaev 2001, pp. 74-75. For a less profound sketch of an alternative proof, see Daley and Vere-Jones 2003, p. 425. □

In heuristic terms, Lemma A.22 indicates that all information about a predictable process at time t is already determined at previous points in time, reflecting the prior statement about left-continuity. For this reason, predictable processes are particularly suitable for modeling the systematic part of a random system - the so-called compensator.

A.2.3. The Doob-Meyer Decomposition

The concept of compensators can be motivated in several ways, such as in Example A.19 above. However, the prevalent approach is to incorporate the definition of the compensator into the assertion of the Doob-Meyer decomposition. For instance, Daley and Vere-Jones 2008 and Jacobsen 2006 each follow this line, whereas Dellacherie and Meyer 1978 and Karr 1991 adhere to a more general conception. The version of the Doob-Meyer decomposition presented here is based on the classical formulation found in Ethier and Kurtz 1986, pp. 74–77 or Karatzas and Shreve 1988, pp. 24–27, but has been adapted to fit our framework of counting processes and avoid unnecessary complexity. A similar modern representation is given in Pang, Talreja and Whitt 2007, p. 208.

Theorem A.23 (Doob-Meyer Decomposition; based on Ethier and Kurtz 1986, pp. 74–75)**.**
Let $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq t_0}, \mathbb{P}\right)$ be a filtered probability space satisfying the usual conditions of Definition A.14 and let $X = (X_t)_{t\geq t_0}$ be a right-continuous, non-negative $\left(\mathbb{P}, \{\mathcal{F}_t\}_{t\geq t_0}\right)$-submartingale. Then there exists an $\{\mathcal{F}_t\}_{t\geq t_0}$-predictable, integrable, increasing, right-continuous process $\Lambda = (\Lambda_t)_{t\geq t_0}$ with $\Lambda_{t_0} = 0$ such that the process $M = (M_t)_{t\geq t_0}$ given by

$$M_t := X_t - \Lambda_t\,, \quad t \geq t_0\,, \tag{A.13}$$

is a right-continuous $\left(\mathbb{P}, \{\mathcal{F}_t\}_{t\geq t_0}\right)$-martingale. The process Λ is unique up to indistinguishability[37] *and is called the* compensator *of X.*

[36] The filtration $\{\mathcal{F}_{t-}\}_{t\geq t_0}$ is the left-continuous analogue of the right-continuous filtration $\{\mathcal{F}_{t+}\}_{t\geq t_0}$ introduced in Definition A.14.

[37] This means that the sample paths of any two versions must be equal with probability 1, see Protter 2005, p. 4.

Proof. The Doob-Meyer decomposition owes its name to the fact that it is an extension of the Doob decomposition for discrete-parameter processes to continuous-parameter processes, a result due to Meyer 1962 and Meyer 1963 (existence and uniqueness, respectively). At that, the above decomposition can be stated more generally for processes of class (DL) which includes the right-continuous, non-negative submartingales (see Problem 4.9 of Karatzas and Shreve 1988, p. 24). Avoiding any details, the abbreviation (DL) stands for the local class (D), whose designation in turn goes back to Doob 1956, p. 60 and which is extensively discussed in Dellacherie and Meyer 1978, pp. 82,84. Several different methods in proving this theorem can be found throughout the literature (cf. Beiglböck, Schachermayer and Veliyev 2012 for an overview and alternative proof), but the more recent of them rely on the techniques of the proof given in Karatzas and Shreve 1988, pp. 24–27. The procedure originates from Rao 1969 and can be briefly outlined as follows: First, discrete approximations of Equation (A.13) are obtained by applying the Doob decomposition on dyadic rationals. Then, the Dunford-Pettis compactness criterion (see Dunford and Schwartz 1957, p. 294) yields weak convergence of these approximations in $\mathrm{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ towards the desired decomposition. Finally, the process Λ is shown to be natural, and predictability follows from this by virtue of Remark A.21 (ii) (Beiglböck, Schachermayer and Veliyev 2012, p. 1). Further details are omitted and can be found in the bibliographical references. □

Remark A.24 (On the Preconditions of the Doob-Meyer Decomposition)**.**
On initial study of any proof for Theorem A.23, it may not be immediately clear why (i) the usual conditions and (ii) the predictability of the compensator are required.

(i) The usual conditions ensure that the compensator admits a right-continuous modification, see Ethier and Kurtz 1986, p. 77. Consequently, showing the validity of the usual conditions is a non-issue if the (right-)continuity of the compensator is assumed for all our modeling purposes.

(ii) The predictability of the compensator provides that the decomposition in Equation (A.13) is unique (cf. Liptser and Shiryaev 2001, pp. 67–70). In fact, uniqueness was not part of the original solution to Doob's decomposition problem given by Meyer 1962.

Let us consider the compensator Λ of X and derive from Equation (A.13) and the martingale property of $M = X - \Lambda$ that for any $t_0 \leq s < t$ we have:

$$\begin{aligned}\mathbb{E}\left(X_t - X_s \,|\, \mathcal{F}_s\right) &= \mathbb{E}\left((M_t + \Lambda_t) - (M_s + \Lambda_s) \,|\, \mathcal{F}_s\right) \\ &= \underbrace{\mathbb{E}\left(M_t - M_s \,|\, \mathcal{F}_s\right)}_{=0} + \mathbb{E}\left(\Lambda_t - \Lambda_s \,|\, \mathcal{F}_s\right) = \mathbb{E}\left(\Lambda_t - \Lambda_s \,|\, \mathcal{F}_s\right) . \end{aligned} \tag{A.14}$$

Consequently, the expected increase of the process X on the interval $(s, t]$ given the information at time s is equal to the expected increase of the predictable compensator. Furthermore, from Equation (A.14) it follows by the tower property that

$$\mathbb{E}\left(X_t - X_s\right) = \mathbb{E}\left(\mathbb{E}\left(X_t - X_s \,|\, \mathcal{F}_s\right)\right) = \mathbb{E}\left(\mathbb{E}\left(\Lambda_t - \Lambda_s \,|\, \mathcal{F}_s\right)\right) = \mathbb{E}\left(\Lambda_t - \Lambda_s\right) ,$$

and for $s = t_0$ we get

$$\mathbb{E}\left(X_t - X_{t_0}\right) = \mathbb{E}\left(\Lambda_t\right) ,$$

since $\Lambda_{t_0} = 0$ holds deterministically. Finally, in the case that $X_{t_0} = 0$ - as satisfied by any counting process - we obtain:

$$\mathbb{E}(X_t) = \mathbb{E}(\Lambda_t) , \quad \text{for all } t \geq t_0 . \tag{A.15}$$

It is therefore natural to view the compensator as a predictor of the process X and hence to model its qualitative behavior based on the compensator. The questions arise how such compensators can be determined and, conversely, which predictable, integrable, increasing, right-continuous processes appear as compensators of counting processes in the first place. The intensity theory will lend itself to answering both questions, at least in the special case of so-called intrinsic histories. Before that, we extend the ongoing example of the Poisson process (see Examples A.7 and A.19) by Watanabe's characterization.

Example A.25 (Watanabe's Characterization of the Poisson Process; Watanabe 1964, pp. 58–59).
Let $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P}\right)$ be a filtered probability space and $N = (N_t)_{t\geq 0}$ be an adapted counting process. If there exists a constant $\lambda \in (0,\infty)$ such that $N_t - \lambda t$ is a $\left(\mathbb{P}, \{\mathcal{F}_t\}_{t\geq 0}\right)$-martingale, then N is a homogeneous Poisson process with intensity λ.

Recall that in Example A.19 we established that the Poisson process compensated by $\Lambda(t) := \lambda t$ is a martingale with respect to its natural filtration. Watanabe's characterization demonstrates that the converse is also true, in that a counting process compensable by the above Λ is necessarily a Poisson process. Accordingly, this serves as a first example that a counting process can be completely determined by the associated compensator. A framework that allows this type of characterization for counting processes other than the Poisson process is provided by the intensity theory.

A.2.4. Intensity Theory

The abstract definition of the compensator by virtue of the Doob-Meyer decomposition hides its dependence on both the probability measure $\mathbb{P}$ and the underlying filtration $\{\mathcal{F}_t\}_{t\geq t_0}$ to some extent; in particular, finding an explicit representation of the compensator (and thus an intelligible interpretation of the above dependence) seems hardly feasible. However, if the sample paths of the compensator Λ are absolutely continuous with respect to the Lebesgue measure[38], then it admits a density $\lambda = (\lambda_t)_{t\geq t_0}$,

$$\Lambda_t = \int_{t_0}^{t} \lambda_u \, \mathrm{d}u \, .$$

If we assume λ to be bounded by an integrable random variable, we can compute for $t \geq t_0$ using the averaging and the dominated convergence theorem (cf. Aalen 1978, p. 705):

$$\begin{aligned}\lim_{h\downarrow 0} \frac{1}{h}\mathbb{E}\left(N_{t+h} - N_t \,|\, \mathcal{F}_t\right) &= \lim_{h\downarrow 0} \frac{1}{h}\mathbb{E}\left(\Lambda_{t+h} - \Lambda_t \,|\, \mathcal{F}_t\right) \\ &= \lim_{h\downarrow 0} \mathbb{E}\left(\frac{1}{h}\int_t^{t+h} \lambda_u \, \mathrm{d}u \,\middle|\, \mathcal{F}_t\right)\end{aligned}$$

[38] i.e., for each $\omega \in \Omega$ the measure induced by the increasing function $t \mapsto \Lambda(t,\omega)$ is absolutely continuous with respect to the Lebesgue measure. The induced measure on $\mathcal{B}([t_0,\infty))$ is given by $\mu_\Lambda([t_0,t]) := \Lambda(t,\omega)$ (cf. Ethier and Kurtz 1986, p. 74), where the dependence on ω is suppressed.

$$
\begin{aligned}
&= \mathbb{E}\left(\lim_{h\downarrow 0} \frac{1}{h}\int_t^{t+h} \lambda_u \,\mathrm{d}u \,\middle|\, \mathcal{F}_t\right) \\
&= \mathbb{E}\left(\lambda_{t+} \middle| \mathcal{F}_t\right) = \lambda_{t+}\,, \qquad \text{(A.16)}
\end{aligned}
$$

where $\lambda_{t+} = \lim_{u\downarrow t} \lambda_u$, and the last equation holds since λ_{t+} is $\mathcal{F}_t$-measurable (a proof of this statement can be found in Appendix A.4). This identity motivates the suggestive relation[39]

$$
\lambda(t)\,\mathrm{d}t \approx \mathbb{E}\left(N\left(\mathrm{d}t\right) \mid \mathcal{F}_{t-}\right)\,, \qquad \text{(A.17)}
$$

commonly deployed in the literature (e.g., Daley and Vere-Jones 2003, p. 232, Karr 1991, p. 69 or Andersen et al. 1993, p. 51), which in a heuristical vein is obtainable by transition to $t-$. According to Equation (A.17), the continuous-parameter process λ turns out to be the conditional instantaneous average rate for the occurrence of a jump of the associated counting process given the strict past. Moreover, this relation implicitly proposes that λ_t is $\mathcal{F}_{t-}$-measurable, something that can be achieved by retaining the predictability requirement. We formalize these considerations by incorporating them into the following definition.

Definition A.26 (Stochastic Intensity; Karr 1991, p. 69)**.**
If in the setting of the Doob-Meyer decomposition (Theorem A.23) a counting process N is considered and there exists a non-negative, $\{\mathcal{F}_t\}_{t\geq t_0}$-predictable process $\lambda = (\lambda_t)_{t\geq t_0}$ satisfying

$$
\Lambda_t = \int_{t_0}^{t} \lambda_u \,\mathrm{d}u\,, \quad \textit{for all } t \geq t_0\,, \qquad \text{(A.18)}
$$

then we say that N admits the $\left(\mathbb{P}, \{\mathcal{F}_t\}_{t\geq t_0}\right)$-intensity λ.

Remark A.27 (On the Existence and Uniqueness of Stochastic Intensities)**.**

(i) Although much emphasis is placed in the literature on counting processes that admit stochastic intensities, the question of their existence is rarely addressed. For the compensator to have absolutely continuous sample paths with respect to the Lebesgue measure, it must necessarily be continuous. The class of counting processes whose compensator is continuous are precisely the *regular* counting processes, and while we do not want to delve further into this topic, the notion accompanies us in the profound study of the underlying theory (e.g., Daley and Vere-Jones 2003 and Karatzas and Shreve 1988). Sufficient conditions that ensure the existence of stochastic intensities are given in Dolivo 1974, pp. 100–106 and Boel, Varaiya and Wong 1975, p. 1007.

(ii) For the mere definition of a stochastic intensity by virtue of Equation (A.18), only the progressive measurability[40] of the process λ must be required instead of predictability (cf. Andersen et al. 1993, p. 75 and Brémaud 1981, p. 27). In this case, however, one can always find a predictable version of the intensity. Furthermore, if the stochastic intensity is constrained to be predictable, it is essentially unique, in that for any two predictable stochastic intensities λ, $\tilde{\lambda}$, we have $\mathbb{P}$-almost surely $\lambda_t = \tilde{\lambda}_t$ μ_Λ-almost

[39] In fact, a semi-constructive proof of the Doob-Meyer decomposition can be based on this identity, see Equations (5.6) and (5.7) of Ethier and Kurtz 1986, p. 76.

[40] We have already encountered this notion in the Proof of Proposition A.9 and are content to think of it as a stronger form of measurability within the framework of martingale theory.

everywhere, where μ_Λ once again denotes the measure on $\mathcal{B}([t_0, \infty))$ induced by the increasing function $t \mapsto \Lambda(t, \omega)$ (Brémaud 1981, pp. 30–31).

Example A.28 (Intensity of the Homogeneous Poisson Process).
Remember that according to Watanabe's characterization (Example A.25) a counting process N with compensator $\Lambda(t) = \lambda t$ is necessarily a Poisson process with intensity λ. This notion is well chosen and conforms to the definition of stochastic intensities, since

$$\Lambda(t) = \int_0^t \lambda \, \mathrm{d}u \, .$$

In order to see that not every counting process admits a stochastic intensity, it is sufficient to enlarge the respective filtration. If in particular we have an intrinsic filtration (i.e., $\mathcal{F}_t = \mathcal{F}_t^N \vee \mathcal{G}_0$ for each $t \geq t_0$) satisfying

$$\mathcal{F}_\infty^N := \sigma\left(\bigcup_{t \geq t_0} \mathcal{F}_t^N\right) \subset \mathcal{G}_0 \, ,$$

all information regarding N is available at time t_0 and the counting process is forced to be its own compensator. To see this, let Λ denote the compensator as given by the Doob-Meyer decomposition and compute:

$$N_t - \Lambda_t = \mathbb{E}\left(N_t - \Lambda_t \,|\, \mathcal{F}_{t_0}\right) = N_{t_0} - \Lambda_{t_0} = 0 \, ,$$

where we exploit that $\mathcal{F}_t \equiv \mathcal{G}_0$ and hence $N_t - \Lambda_t$ is $\mathcal{F}_s$-measurable even for $t_0 \leq s < t$. We conclude that the smaller the filtration, the more likely a stochastic intensity is to exist (Karr 1991, p. 69, where also another somewhat less degenerative example is discussed). A formal representation of this intuitive interrelation is the subject of the following lemma.

Lemma A.29 (Change of Filtration for Intensities; Brémaud 1981, p. 32)**.**
Let N be a counting process defined on the filtered probability space $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq t_0}, \mathbb{P}\right)$, and let $\{\mathcal{G}_t\}_{t \geq t_0}$ be another filtration satisfying

$$\mathcal{F}_t^N \subset \mathcal{G}_t \subset \mathcal{F}_t \, , \quad \textit{for all } t \geq t_0 \, .$$

If N admits the $\left(\mathbb{P}, \{\mathcal{F}_t\}_{t \geq t_0}\right)$-intensity λ, then $\tilde{\lambda} = \left(\tilde{\lambda}_t\right)_{t \geq t_0}$ given by

$$\tilde{\lambda}_t := \mathbb{E}\left(\lambda_t \,|\, \mathcal{G}_t\right) \, , \quad t \geq t_0 \, , \tag{A.19}$$

is the $\left(\mathbb{P}, \{\mathcal{G}_t\}_{t \geq t_0}\right)$-intensity of N.

Proof. A simplified proof, sufficient for all applications where actual computations are performed, can be found in Brémaud 1981, pp. 32–33. For its validity, it is required that the process $\tilde{\lambda}$ defined via Equation (A.19) is $\{\mathcal{G}_t\}_{t \geq t_0}$-predictable, which in practice can be achieved by considering a left-continuous version of $\tilde{\lambda}$. The proof then turns out to be a straightforward application of Fubini's theorem. $\square$

The implications of Lemma A.29 in terms of the aforementioned intrinsic filtrations are twofold: Firstly, considering intrinsic filtrations in lieu of internal filtrations proves to be a useful extension, since in the absence of preliminary information (i.e., $\mathcal{G}_0 = \{\emptyset, \Omega\}$) the updated intensity can be calculated from the former one. Secondly and more importantly, external effects may cause an increase in information that cannot be reflected by intrinsic filtrations, in which case the intensity $\tilde{\lambda}$ still emerges as the best mean square estimate of λ within the restricted information (cf. Dolivo 1974, p. 106). This henceforth serves as a justification for focusing the primary attention of this thesis on intrinsic filtrations. Furthermore, the simple structure of these filtrations carries an additional benefit: It allows us to deduce an explicit representation of the compensator and thus of the corresponding intensity function, which dates back to Jacod 1973 and is occasionally referred to as "Jacod's formula for the intensity process" (cf. Andersen et al. 1993, pp. 95–96). This result given in Theorem A.33 concludes a series of lemmas dealing with the successive derivation of said formula relying on the concepts presented so far. While they are largely based on the elaboration found in Daley and Vere-Jones 2008, pp. 356–365, we supplement essential arguments and correct the erroneous proof of Lemma 14.1.III. The proofs are predominantly of a technical nature and can be reviewed in Appendix A.4. For readers more familiar with intensity theory, skipping the following results altogether and consulting only the condensed Summary 1 at the end of the subsection is advised. We start with the simplest example, a one-point process consisting of only a single point whose location is defined by some random variable X.

Lemma A.30 (One-Point Process: Compensator w.r.t. the Internal Filtration; Lemma 14.1.II. of Daley and Vere-Jones 2008, p. 359)**.**
Let X be a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in $([t_0, \infty], \mathcal{B}([t_0, \infty]))$. Let F denote the cumulative distribution function of X and define the one-point process N by

$$N : [t_0, \infty) \times \Omega \longrightarrow \mathbb{R} : (t, \omega) \longmapsto N(t, \omega) = \mathbb{1}_{(-\infty, t]}(X(\omega)) = \begin{cases} 1, & t \geq X(\omega), \\ 0, & t < X(\omega). \end{cases}$$

The one-point process N has the $\left(\mathbb{P}, \left\{\mathcal{F}_t^N\right\}_{t \geq t_0}\right)$-compensator Λ given by

$$\Lambda : [t_0, \infty) \times \Omega \longrightarrow \mathbb{R} : (t, \omega) \longmapsto \Lambda(t, \omega) = H(t \wedge X(\omega)) = \begin{cases} H(X(\omega)), & t \geq X(\omega), \\ H(t), & t < X(\omega), \end{cases}$$

where $\left\{\mathcal{F}_t^N\right\}_{t \geq t_0}$ denotes the internal filtration of N and H is defined via

$$H(t) := \int_{t_0}^{t} \frac{\mathrm{d}F(x)}{1 - F(x-)}. \tag{A.20}$$

Remark A.31 (Hazard Measure and Hazard Function; Jacobsen 2006, p. 34 and Daley and Vere-Jones 2003, p. 109)**.**
If we consider the case of an absolutely continuous cumulative distribution function F in the situation of Lemma A.30, F admits a density f and the function H from Equation (A.20) satisfies

$$H(t) = \int_{t_0}^{t} \frac{f(x)}{1 - F(x)} \, \mathrm{d}x .$$

Thus, H can equivalently be stated in terms of the hazard function

$$h(x) := \frac{f(x)}{1 - F(x)} = \frac{f(x)}{S(x)}\,, \tag{A.21}$$

where $S(x) = 1 - F(x)$ denotes the survival function of X. For this reason, the function H is called the *integrated hazard function* (IHF), even in the case where F is not absolutely continuous. The associated *hazard measure* Q defined on $\mathcal{B}\left([t_0, \infty)\right)$ can be obtained through its Radon-Nikodym derivative with respect to the measure induced by F: Corresponding to the integrated form of Equation (A.20), we have

$$Q\left([t_0, t]\right) = \int_{t_0}^{t} \frac{\mathrm{d}F(x)}{1 - F(x-)}\,, \quad t \geq t_0\,,$$

where this equation can alternatively be written in the differentiated form

$$\mathrm{d}Q\left(x\right) = \frac{\mathrm{d}F\left(x\right)}{1 - F(x-)}\,. \tag{A.22}$$

An informal proof of Lemma A.30 can be given in the situation of Remark A.31 where F admits a continuous density f, illustrating the relevance of the heuristic presented in Equation (A.16). However, the identity below may be considered only on $\{X > t\}$ (note that $\mathcal{F}_t^N = \{\emptyset, \{X > t\}, \{X \leq t\}, \Omega\}$), which turned out to be the more interesting case in the scope of Lemma A.30:

$$\begin{aligned}
\lim_{h \downarrow 0} \frac{1}{h} \mathbb{E}\left(N_{t+h} - N_t \,\middle|\, \mathcal{F}_t^N\right) &= \lim_{h \downarrow 0} \frac{1}{h} \mathbb{E}\left(N_{t+h} - N_t \,|\, X > t\right) \\
&= \lim_{h \downarrow 0} \frac{1}{h} \mathbb{P}\left(N_{t+h} - N_t = 1 \,|\, X > t\right) \\
&= \lim_{h \downarrow 0} \frac{1}{h} \mathbb{P}\left(t < X \leq t + h \,|\, X > t\right) \\
&= \lim_{h \downarrow 0} \frac{1}{h} \frac{\mathbb{P}\left(t < X \leq t + h\right)}{\mathbb{P}\left(X > t\right)} \\
&= \lim_{h \downarrow 0} \frac{1}{h} \frac{F(t+h) - F(t)}{S(t)} \\
&= \frac{f(t)}{S(t)} = h(t)\,,
\end{aligned}$$

so that on $\{X > t\}$ the intensity function is equal to the hazard function. On $\{X \leq t\}$, the difference $N_{t+h} - N_t$ vanishes and therefore the same is true for the intensity. Thus,

$$\lambda(t, \omega) = \begin{cases} h(t)\,, & t < X(\omega)\,, \\ 0\,, & t \geq X(\omega)\,, \end{cases}$$

which by integration yields exactly the compensator from Lemma A.30. The advantage of a formal proof, given in Appendix A.4, lies in its capability to be easily generalized to the cases of intrinsic filtrations and counting processes. The following result extends Lemma A.30 to intrinsic filtrations.

Lemma A.32 (One-Point Process: Compensator w.r.t. an Intrinsic Filtration; Lemma 14.1.III. of Daley and Vere-Jones 2008, p. 361)**.**
In the situation of Lemma A.30, let $\mathcal{G}_0$ denote the prior σ-algebra of an intrinsic filtration $\mathcal{F}_t = \mathcal{F}_t^N \vee \mathcal{G}_0$. If a regular conditional distribution function $F\left(\cdot \,|\, \mathcal{G}_0\right)$ for X exists, the one-point process N has the $\left(\mathbb{P}, \{\mathcal{F}_t\}_{t \geq t_0}\right)$-compensator Λ given by

$$\begin{aligned} \Lambda \,:\, & [t_0, \infty) \times \Omega \longrightarrow \mathbb{R} \\ & (t, \omega) \longmapsto \Lambda(t, \omega) = H(t \wedge X(\omega) \,|\, \mathcal{G}_0) = \begin{cases} H(X(\omega) \,|\, \mathcal{G}_0)\,, & t \geq X(\omega)\,, \\ H(t \,|\, \mathcal{G}_0)\,, & t < X(\omega)\,, \end{cases} \end{aligned}$$

where $H\left(\cdot \,|\, \mathcal{G}_0\right)$ is the conditional integrated hazard function associated with $F\left(\cdot \,|\, \mathcal{G}_0\right)$,

$$H\left(t \,|\, \mathcal{G}_0\right) := \int_{t_0}^{t} \frac{\mathrm{d}F\left(x \,|\, \mathcal{G}_0\right)}{1 - F\left(x- \,|\, \mathcal{G}_0\right)}\,.$$

Jacod's formula for the intensity process (respectively the compensator in the absence of its absolute continuity) concludes this successive sequence of lemmata. Thereby, the essential step is the conception of a counting process as a superposition of one-point processes, writing as in Equation (A.2):

$$N(t, \omega) = \sum_{i=1}^{\infty} \underbrace{\mathbb{1}_{(-\infty, t]}\left(T_i(\omega)\right)}_{=: N_i(t, \omega)}\,, \quad t \geq t_0\,.$$

By additivity, it then suffices to derive the compensator individually for the one-point processes N_i. However, this simplification comes at a price: Since we are now considering an infinite sequence of points (i.e., the point process $(T_i)_{i \in \mathbb{N}}$) rather than a single random variable X, the information at a given time t can no longer be easily decomposed into a large atom and the remainder. Instead, we need to contemplate the information of the associated counting process N up to the random times T_i, that is, the evolution of N until the arrival of the ith event. This is reflected in the definition of the stopped σ-algebra $\mathcal{F}_{T_i}$, where for $i \in \mathbb{N}$ we have:

$$\mathcal{F}_{T_i} := \sigma\left(\{N_{t \wedge T_i} \,:\, t \geq t_0\}\right) \vee \mathcal{G}_0\,. \tag{A.23}$$

Note that stopped σ-algebras are a common concept encountered in further study of stopping times, see Dellacherie 1972, p. 117 for details, whereas the representation given in Equation (A.23) is specific to our situation, as outlined in Karr 1991, pp. 54–56. Since $(T_i)_{i \in \mathbb{N}}$ is an almost surely strictly increasing sequence and $\mathcal{F}_{T_i}$ contains knowledge about N up to its ith jump, conditioning on $\mathcal{F}_{T_i}$ means we are given $\mathcal{G}_0$ as well as $T_1, \ldots, T_i$. Setting $T_0 \equiv t_0$ as in Example A.7 finally yields $\mathcal{F}_{T_0} = \mathcal{G}_0$, a direct consequence of Equation (A.23) (compare Equation (A.9) on this regard).
After these preliminary considerations, we are now able to state Jacod's formula, which is given under this name in Andersen et al. 1993, p. 96. In order to preserve the readability and the previous notation, we cite Daley and Vere-Jones 2008 again.

Theorem A.33 (Jacod's Formula for the Intensity Process; Theorem 14.1.IV. of Daley and Vere-Jones 2008, pp. 363–364)**.**
Let $N = (N_t)_{t\geq t_0}$ be a counting process and $T = (T_i)_{i\in\mathbb{N}}$ the associated simple point process. Let $\mathcal{F}_t = \mathcal{F}_t^N \vee \mathcal{G}_0$ denote an intrinsic filtration with prior σ-algebra $\mathcal{G}_0$. Suppose there exist regular versions $F_i\left(\cdot \,|\, \mathcal{F}_{T_{i-1}}\right)$ of the conditional distribution functions of the interarrival times $W_i = T_i - T_{i-1}$, given $\mathcal{F}_{T_{i-1}}$ as in Equation (A.23)*, such that $1 - F_i\left(x- \,|\, \mathcal{F}_{T_{i-1}}\right) > 0$ for $x > 0$. Let N_i denote the one-point process given by*

$$N_i \,:\, [t_0, \infty) \times \Omega \longrightarrow \mathbb{R} \,:\, (t, \omega) \longmapsto N_i(t, \omega) = \mathbb{1}_{(-\infty,t]}\left(T_i(\omega)\right) \,,$$

so that $N = \sum_{i=1}^{\infty} N_i$. Then the $\left(\mathbb{P}, \{\mathcal{F}_t\}_{t\geq t_0}\right)$-compensator Λ_i for N_i has the form

$$\Lambda_i(t, \omega) = \begin{cases} 0\,, & t < T_{i-1}(\omega)\,, \\ H_i\left(t - T_{i-1} \,|\, \mathcal{F}_{T_{i-1}}\right)\,, & T_{i-1}(\omega) \leq t < T_i(\omega)\,, \\ H_i\left(T_i - T_{i-1} \,|\, \mathcal{F}_{T_{i-1}}\right)\,, & T_i(\omega) \leq t\,, \end{cases} \tag{A.24}$$

where $H_i\left(\cdot \,|\, \mathcal{F}_{T_{i-1}}\right)$ is the conditional integrated hazard function associated with $F_i\left(\cdot \,|\, \mathcal{F}_{T_{i-1}}\right)$,

$$H_i\left(t \,|\, \mathcal{F}_{T_{i-1}}\right) = \int_0^t \frac{\mathrm{d}F_i\left(x \,|\, \mathcal{F}_{T_{i-1}}\right)}{1 - F_i\left(x- \,|\, \mathcal{F}_{T_{i-1}}\right)}\,. \tag{A.25}$$

Thus, a version of the $\left(\mathbb{P}, \{\mathcal{F}_t\}_{t\geq t_0}\right)$-compensator Λ for N is given by

$$\Lambda(t, \omega) = \sum_{i=1}^{\infty} \Lambda_i(t, \omega)\,. \tag{A.26}$$

Although we treated the general case first, Theorem A.33, as its name suggests, serves primarily for the explicit determination of the intensity function in the absolutely continuous case. The following corollary can be seen as a continuation of Remark A.31, which dealt with absolute continuity in the case of a one-point process.

Corollary A.34 (Jacod's Formula, absolutely continuous case; cf. Brémaud 1981, pp. 61–63 and Daley and Vere-Jones 2008, pp. 364–365)**.**
In the situation of Theorem A.33, the $\left(\mathbb{P}, \{\mathcal{F}_t\}_{t\geq t_0}\right)$-compensator Λ is almost surely absolutely continuous if and only if the conditional distribution functions $F_i\left(\cdot \,|\, \mathcal{F}_{T_{i-1}}\right)$ have absolutely continuous versions with densities $f_i\left(\cdot \,|\, \mathcal{F}_{T_{i-1}}\right)$. In this case, one version of Λ is given by

$$\Lambda(t, \omega) = \int_{t_0}^t \lambda^*(u, \omega)\,\mathrm{d}u\,,$$

where

$$\lambda^*(t, \omega) = \sum_{i=1}^{\infty} \lambda_i^*(t, \omega) \equiv \sum_{i=1}^{\infty} \frac{f_i\left(t - T_{i-1} \,|\, \mathcal{F}_{T_{i-1}}\right)}{1 - F_i\left(t - T_{i-1} \,|\, \mathcal{F}_{T_{i-1}}\right)} \mathbb{1}_{\{T_{i-1} \leq t < T_i\}}\,. \tag{A.27}$$

An $\{\mathcal{F}_t\}_{t\geq t_0}$-predictable version λ of λ^* and hence the $\left(\mathbb{P}, \{\mathcal{F}_t\}_{t\geq t_0}\right)$-intensity of N is *defined by*

$$\lambda(t,\omega) = \sum_{i=1}^{\infty} \lambda_i(t,\omega) \equiv \sum_{i=1}^{\infty} \frac{f_i\left(t - T_{i-1} \,|\, \mathcal{F}_{T_{i-1}}\right)}{1 - F_i\left(t - T_{i-1} \,|\, \mathcal{F}_{T_{i-1}}\right)} \mathbb{1}_{\{T_{i-1} < t \leq T_i\}} \,.$$

One of the key features of Jacod's formula is the capability to completely characterize a process's intensity by the conditional hazard functions $h_i\left(\cdot \,|\, \mathcal{F}_{T_{i-1}}\right)$, which are in turn fully determined by the conditional densities $f_i\left(\cdot \,|\, \mathcal{F}_{T_{i-1}}\right)$. Nevertheless, a reasonable modelling approach tries to avoid having to specify the entire family of conditional distributions. Hence, the reason why Corollary A.34 is vital for intensity-based modelling is that the converse statement also holds: a counting process is uniquely characterized by its intensity.

Proposition A.35 (Characterization of Counting Processes via Stochastic Intensities; Prop. 7.2.IV. of Daley and Vere-Jones 2003, p. 233)**.**
In the situation of Corollary A.34, the $\left(\mathbb{P}, \{\mathcal{F}_t\}_{t \geq t_0}\right)$*-intensity* λ *determines the probability structure of* N *uniquely.*

Proof. The proposition is partly embedded in the uniqueness theorem, stating that at most one probability measure corresponds to a given intensity, see Theorem 2.19 of Karr 1991, p. 63 or Brémaud 1981, pp. 63–64, 77–78, where the latter extensively deals with the special case of internal filtrations. A detailed elaboration for the absolutely continuous case can be found in Daley and Vere-Jones 2003, pp. 229–233. A prominent example to illustrate how the reconstruction of the counting process from its intensity proceeds is Watanabe's characterization of the Poisson process, see Example A.25. □

Remark A.36 (Technical Preconditions in Intensity-Based Modelling)**.**
Incorporating the findings of Proposition A.35, we want to shed new light on Remark A.24. Initially, we were faced with the challenge that an intensity-based modelling approach should satisfy the usual conditions of Definition A.14 on the one hand and ensure predictability of the compensator on the other. However, we recognized in Remark A.24 that the usual conditions primarily serve to find a right-continuous modification of the compensator, whereas left-continuity implies predictability according to Remark A.21. Now, if we provide a model for the stochastic intensity λ, Proposition A.35 states that the cumulative intensity

$$\Lambda(t,\omega) = \int_{t_0}^{t} \lambda(u,\omega)\, \mathrm{d}u$$

is an absolutely continuous compensator of the counting process associated with λ. In particular, the compensator is right-continuous, which eliminates the need to check the usual conditions, and left-continuous, which makes it predictable. Consequently, for many practical applications, we do not need to bother with the technical requirements of intensity theory and circumvent the necessity of proving the compensator's predictability.

We round out this subsection by summarizing the key insights in simplified form sufficient for the most common applications. In doing so, we will also slightly shift perspective by specifying the intensity based on the point process itself rather than on the interarrival times. This can be achieved by considering the shifted conditional distribution functions $\tilde{F}_i\left(\cdot \,|\, \mathcal{F}_{T_{i-1}}\right)$ satisfying

$$\tilde{F}_i\left(t \,|\, \mathcal{F}_{T_{i-1}}\right) := F_i\left(t - T_{i-1} \,|\, \mathcal{F}_{T_{i-1}}\right) ,$$

as well as the shifted conditional integrated hazard function $\tilde{H}_i\left(\cdot \,|\, \mathcal{F}_{T_{i-1}}\right)$ defined by

$$\tilde{H}_i\left(t \,|\, \mathcal{F}_{T_{i-1}}\right) := \int_{T_{i-1}}^{t} \frac{\mathrm{d}\tilde{F}_i\left(x \,|\, \mathcal{F}_{T_{i-1}}\right)}{1 - \tilde{F}_i\left(x- \,|\, \mathcal{F}_{T_{i-1}}\right)} = H_i\left(t - T_{i-1} \,|\, \mathcal{F}_{T_{i-1}}\right) . \tag{A.28}$$

If defined as above, $\tilde{F}_i\left(\cdot \,|\, \mathcal{F}_{T_{i-1}}\right)$ represents the conditional distribution function of T_i given $\mathcal{F}_{T_{i-1}}$ and Λ_i from Equation (A.24) can be stated as

$$\Lambda_i(t,\omega) = \begin{cases} 0 , & t < T_{i-1}(\omega) , \\ \tilde{H}_i\left(t \,|\, \mathcal{F}_{T_{i-1}}\right) , & T_{i-1}(\omega) \leq t < T_i(\omega) , \\ \tilde{H}_i\left(T_i \,|\, \mathcal{F}_{T_{i-1}}\right) , & T_i(\omega) \leq t , \end{cases}$$

which becomes evident by plugging in the identity of Equation (A.28).

Summary 1 (Intensity Theory of Counting Processes)**.**
Let $t_1 < t_2 < \ldots$ with $t_i \in [t_0, \infty]$ be realizations of a point process T. Let $f_i\left(t \,|\, t_1, \ldots, t_{i-1}\right)$ denote the absolutely continuous conditional density function of T_i after the observation of $T_1 = t_1, \ldots, T_{i-1} = t_{i-1}$ and

$$S_i\left(t \,|\, t_1, \ldots, t_{i-1}\right) := 1 - \int_{t_{i-1}}^{t} f_i\left(u \,|\, t_1, \ldots, t_{i-1}\right) \mathrm{d}u \tag{A.29}$$

the associated survival function. The corresponding hazard functions are given by

$$h_i\left(t \,|\, t_1, \ldots, t_{i-1}\right) := \frac{f_i\left(t \,|\, t_1, \ldots, t_{i-1}\right)}{S_i\left(t \,|\, t_1, \ldots, t_{i-1}\right)} . \tag{A.30}$$

By integrating Equation (A.30), one obtains the cumulative conditional hazard functions:

$$H_i\left(t \,|\, t_1, \ldots, t_{i-1}\right) := \int_{t_{i-1}}^{t} h_i\left(u \,|\, t_1, \ldots, t_{i-1}\right) \mathrm{d}u . \tag{A.31}$$

Furthermore, the conditional intensity function can then be piecewise defined as follows:

$$\lambda^*(t) := \begin{cases} h_1(t) , & t_0 \leq t < t_1 , \\ h_i\left(t \,|\, t_1, \ldots, t_{i-1}\right) , & t_{i-1} \leq t < t_i ,\, i \geq 2 . \end{cases} \tag{A.32}$$

The conditional intensity can be interpreted as the conditional risk of the occurance of an event at t, given the realizations of the process over $[0,t)$. Since densities are unique only up to its values on sets with Lebesgue measure zero, the conditional intensity need not necessarily be left-continuous. Uniqueness can then be achieved by taking the left-continuous modification[41] $\lambda(t)$ of $\lambda^*(t)$ (i.e., $\lambda(t) = \lambda^*(t-)$). This modification emerges as the canonical choice because it ensures predictability. The cumulative intensity

[41]Formally this is possible since we presume the existence of continuous conditional density functions for the T_i, see proofs of Lemma A.29 and Corollary A.34 for details.

process Λ is defined as the pointwise integral

$$\Lambda(t) := \int_{t_0}^{t} \lambda(u)\,\mathrm{d}u = \int_{t_0}^{t} \lambda^*(u)\,\mathrm{d}u\,. \tag{A.33}$$

Then, Λ is an $\{\mathcal{F}_t^N\}_{t\geq t_0}$-predictable process and $M(t) = N(t) - \Lambda(t)$ is an $\left(\mathbb{P}, \{\mathcal{F}_t^N\}_{t\geq t_0}\right)$-martingale. The cumulative intensity process Λ can thus be identified with the compensator obtained through the Doob-Meyer decomposition of N. Finally, when substituting the conditional hazard function $h_i\left(\cdot\,|\,t_1,\ldots,t_{i-1}\right)$ in Equation (A.31) with the predictable version λ of the conditional intensity function λ^* from Equation (A.32), one obtains for $t \leq t_i$:

$$H_i\left(t\,|\,t_1,\ldots,t_{i-1}\right) = \int_{t_{i-1}}^{t} \lambda(u)\,\mathrm{d}u = \Lambda(t) - \Lambda(t_{i-1})\,, \tag{A.34}$$

where the definition of the cumulative intensity process Λ from Eq. (A.33) is considered.

In the presence of an absolutely continuous compensator Λ, Corollary A.34 allows us to draw a useful Lemma from Summary 1. It states that with probability one, countably many independent copies of the underlying counting process have no common discontinuities.

Lemma A.37 (Common Discontinuities of Independent Counting Processes)**.**
Let $N^{(1)}, N^{(2)}, \ldots$ be an at most countable collection of independent counting processes with index set $[t_0, \infty)$ defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Suppose that, for each $j \in \mathbb{N}$, the associated compensator $\Lambda^{(j)}$ with respect to the canonical filtration of $N^{(j)}$ is absolutely continuous. Let $T^{(1)}, T^{(2)}, \ldots$ denote the corresponding counting processes. Then, for all $i, j, k, l \in \mathbb{N}$ where $j \neq l$ or $i \neq k$, the following holds:

$$\mathbb{P}\left(T_i^{(j)} = T_k^{(l)}\,, T_i^{(j)} < \infty\right) = 0\,. \tag{A.35}$$

In terms of the processes $N^{(1)}, N^{(2)}, \ldots$, this implies that $\mathbb{P}$-almost surely they exhibit no common discontinuities on $[t_0, \infty)$. In particular, this applies on any subinterval $\mathcal{I} \subset [t_0, \infty)$.

Proof. The lemma is a generalization of the elementary statement that two independent random variables with absolutely continuous probability distributions almost surely take on different values. A proof is given in Appendix A.4. □

A.2.5. The Itô Isometry for Square-Integrable Martingales

We conclude the current Section A.2 on filtrations, martingales and compensators with a brief digression on stochastic integrals whose integrator is given by a square-integrable martingale. Since a comprehensive coverage of the topic of stochastic integration is outside the scope of this thesis and at the same time promises little insight, we refer for an introductory reading to the textbooks Kuo 2006 (providing a good overview without excessive technical depth) and Øksendal 2013 (where the Brownian motion is considered in place of square-integrable martingales) as well as the monographs Protter 2005 (for an in-depth look at the general theory) and Kallianpur 1980 (see in particular pages 52 to 59 on useful properties of stochastic integrals). Moreover, the proofs presented here are each intended to illustrate only the idea of concept, but we consistently refer to specialized literature for the details that are left out. Our emphasis is on the Itô isometry for square-integrable martingales, which are the subject of a first definition.

Definition A.38 (Square-Integrable Martingale; Karatzas and Shreve 1988, p. 30).
Let $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq t_0}, \mathbb{P}\right)$ *be a filtered probability space and let* $M = (M_t)_{t \geq t_0}$ *be a* $\left(\mathbb{P}, \{\mathcal{F}_t\}_{t \geq t_0}\right)$*-martingale. Then, M is called* square-integrable *if for all* $t \geq t_0$ *it holds:*

$$\mathbb{E}\left(M_t^2\right) < \infty \,.$$

For any such square-integrable martingale, we obtain by *Jensen's inequality* (see Protter 2005, p. 12) and the martingale property:

$$\mathbb{E}\left(M_t^2 \,|\, \mathcal{F}_s\right) \geq \left(\mathbb{E}\left(M_t \,|\, \mathcal{F}_s\right)\right)^2 = M_s^2\,, \quad \text{for all } t \geq s \geq t_0\,.$$

Accordingly, $(M_t^2)_{t \geq t_0}$ is a non-negative $\left(\mathbb{P}, \{\mathcal{F}_t\}_{t \geq t_0}\right)$-submartingale. Moreover, if M is right-continuous or has a right-continuous modification (e.g., if the filtered probability space satisfies the usual conditions, see Remark A.24), the same holds for M^2 and hence the Doob-Meyer decomposition from Theorem A.23 can be applied. The resulting compensator is often denoted by $\langle M \rangle$ and referred to as the *(predictable) quadratic variation process* of M (see Karatzas and Shreve 1988, p. 31 and Appendix B of Karr 1991). Remarkably, if M itself emerges as the martingale from the Doob-Meyer decomposition of a counting process N with compensator Λ, the quadratic variation process $\langle M \rangle$ can be explicitly stated in terms of Λ. If Λ is assumed to be continuous, then in fact $\langle M \rangle = \Lambda$ holds. This preliminary consideration forms the framework of the following theorem, which can be viewed as an extension of the aforesaid Doob-Meyer decomposition.

Theorem A.39 (Extension of the Doob-Meyer Decomposition; cf. Pang, Talreja and Whitt 2007, p. 211).
Let $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq t_0}, \mathbb{P}\right)$ *be a filtered probability space satisfying the usual conditions of Definition A.14 and let* $N = (N_t)_{t \geq t_0}$ *be a* $\mathbb{P}$*-integrable adapted counting process. If the compensator* Λ *of* N *provided by the Doob-Meyer decomposition of Theorem A.23 is continuous, then it holds:*

(i) $M = N - \Lambda$ *is a square-integrable* $\left(\mathbb{P}, \{\mathcal{F}_t\}_{t \geq t_0}\right)$*-martingale.*

(ii) The compensator of M^2 *is given by* Λ.

Proof. The complete proof can be found in Pang, Talreja and Whitt 2007, pp. 265–267. We give here only a sketch of the proof, which covers the main arguments. The idea is roughly to explicitly decompose the process M^2 into a martingale and a predictable increasing process, and then infer the statement from the uniqueness property of the Doob-Meyer decomposition. The first step involves an application of the extended integration-by-parts formula[42] found in Appendix A.50 to obtain:

$$\begin{aligned} M_t^2 &= 2\int_{t_0}^t M_{s-}\,\mathrm{d}M_s + \sum_{t_0 \leq s \leq t} (\Delta M_s)^2 && |M_s = N_s - \Lambda_s \\ &= 2\int_{t_0}^t M_{s-}\,\mathrm{d}M_s + \sum_{t_0 \leq s \leq t} \left[(\Delta N_s)^2 - 2\Delta N_s \Delta\Lambda_s + (\Delta\Lambda_s)^2\right] && |\Delta N_s \in \{0,1\} \end{aligned}$$

[42] Applying the ordinary version given in Theorem A.49 proves equally effective, but allows for slightly less understanding, as will be demonstrated here. As a difference of two increasing càdlàg functions, M is itself a càdlàg function of bounded variation over finite intervals.

$$
\begin{aligned}
&= 2\int_{t_0}^{t} M_{s-}\,\mathrm{d}M_s + \sum_{t_0\le s\le t}\left[(1-2\Delta\Lambda_s)\,\Delta N_s + (\Delta\Lambda_s)^2\right] \\
&= 2\int_{t_0}^{t} M_{s-}\,\mathrm{d}M_s + \int_{t_0}^{t}(1-2\Delta\Lambda_s)\,\mathrm{d}N_s + \int_{t_0}^{t}\Delta\Lambda_s\,\mathrm{d}\Lambda_s\,, \qquad\qquad |N_s = M_s + \Lambda_s \\
&= \int_{t_0}^{t}(2M_{s-}+1-2\Delta\Lambda_s)\,\mathrm{d}M_s + \int_{t_0}^{t}(1-\Delta\Lambda_s)\,\mathrm{d}\Lambda_s\,, \qquad\qquad (A.36)
\end{aligned}
$$

where the second to last step bridges the difference between the two versions of the integration-by-parts formula, since for $f = g = \Lambda$ we have:

$$
\begin{aligned}
& \int_{t_0}^{t}\Lambda_s\,\mathrm{d}\Lambda_s + \int_{t_0}^{t}\Lambda_{s-}\,\mathrm{d}\Lambda_s = 2\int_{t_0}^{t}\Lambda_{s-}\,\mathrm{d}\Lambda_s + \sum_{t_0\le s\le t}(\Delta\Lambda_s)^2 \\
\Longleftrightarrow \quad & \int_{t_0}^{t}\Lambda_s\,\mathrm{d}\Lambda_s - \int_{t_0}^{t}\Lambda_{s-}\,\mathrm{d}\Lambda_s = \sum_{t_0\le s\le t}(\Delta\Lambda_s)^2 \\
\Longleftrightarrow \quad & \int_{t_0}^{t}\Delta\Lambda_s\,\mathrm{d}\Lambda_s = \sum_{t_0\le s\le t}(\Delta\Lambda_s)^2\ .
\end{aligned}
$$

One continues with Equation (A.36) by showing that the first summand is a $\left(\mathbb{P}, \{\mathcal{F}_t\}_{t\ge t_0}\right)$-martingale and the second summand is a predictable increasing process (i.e., Equation (A.36) already represents the Doob-Meyer decomposition of M^2). In the case that Λ is continuous[43] and therefore we have $\Delta\Lambda_s \equiv 0$, this simplifies to

$$
M_t^2 = \int_{t_0}^{t}(2M_{s-}+1)\,\mathrm{d}M_s + \Lambda_t\,, \qquad (A.37)
$$

because of $\Lambda_{t_0} = 0$. As the remaining integrand is left-continuous and hence predictable, the martingale property of the first summand then follows from a standard result for integrals with respect to bounded variation martingales, which for convenience is given in the Lemma A.40 below. As the compensator of N, the second summand is predictable and increasing by default, so part (ii) follows. Moreover, Equation (A.37) provides a rationale as to why M is square-integrable, since then

$$
\mathbb{E}\left(M_t^2\right) = \mathbb{E}\left(\Lambda_t\right) < \infty\,, \quad \text{for all } t \ge t_0\,,
$$

by the integrability of Λ. Note, however, that in the process we have implicitly assumed square-integrability in order to ultimately show it (e.g., in utilizing the Doob-Meyer decomposition). For the mathematically rigorous argument, additional regularity conditions are needed and the required boundedness is achieved by localization via stopping times. We once again refer to Pang, Talreja and Whitt 2007, p. 266 for the remainder of part (i). □

In particular, Theorem A.39 implies that a counting process "with a continuous compensator is locally and conditionally a Poisson process in the sense that its mean and variance are equal" (Karr 1991, p. 64). This is indicated by the following heuristic expression in

[43] The reasoning remains valid even if Λ is not continuous, cf. Theorem 2.21 of Karr 1991, p. 64. However, with this approach we immediately recognize why the continuity of Λ is beneficial.

infinitesimal form:

$$\mathbb{E}\left[\ \big(\underbrace{N(\mathrm{d}t) - \mathbb{E}\left(N(\mathrm{d}t)\,|\,\mathcal{F}_{t-}\right)}_{=M(\mathrm{d}t)}\big)^2\,|\,\mathcal{F}_{t-}\ \right] = \Lambda(\mathrm{d}t) = \mathbb{E}\left(N(\mathrm{d}t)\,|\,\mathcal{F}_{t-}\right)\,,$$

which is understood in the same way as the suggestive relation from Equation (A.17), recall Remark A.48 in Appendix A.5. The result itself foreshadows Theorem A.44 of the coming Section A.3, where we will recognize counting processes with absolutely continuous compensators as Poisson processes under a random time transformation.

The next lemma establishes another relation between predictable processes and martingales (of bounded variation). It turns out to be useful not only for the proof of Theorem A.39, but also in the later consideration of a minimum (L^2-)distance estimator.

Lemma A.40 (Integration with Respect to Bounded Variation Martingales; Theorem T6 of Brémaud 1981, p. 10)**.**
Let $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq t_0}, \mathbb{P}\right)$ be a filtered probability space and $M = (M_t)_{t\geq t_0}$ a $\left(\mathbb{P}, \{\mathcal{F}_t\}_{t\geq t_0}\right)$-martingale of bounded variation (i.e., $\mathbb{P}$-almost all sample paths have bounded variation over finite intervals). Let $|M| = (|M|_t)_{t\geq t_0}$ denote the total variation process[44] associated with M. Suppose further that M is of locally integrable variation, that is,

$$\mathbb{E}\left[\int_{t_0}^{t} \mathrm{d}\,|M|_s\right] < \infty\,, \quad \textit{for all } t \geq t_0\,.$$

Then, for each $\{\mathcal{F}_t\}_{t\geq t_0}$-predictable process $f = (f_t)_{t\geq t_0}$ satisfying

$$\mathbb{E}\left[\int_{t_0}^{t} |f_s|\ \mathrm{d}\,|M|_s\right] < \infty\,, \quad \textit{for all } t \geq t_0\,, \tag{A.38}$$

the stochastic process $X = (X_t)_{t\geq t_0}$, which for each $t \geq t_0$ is defined by

$$X_t = \int_{t_0}^{t} f_s\,\mathrm{d}M_s\,, \tag{A.39}$$

is a $\left(\mathbb{P}, \{\mathcal{F}_t\}_{t\geq t_0}\right)$-martingale.

Proof. A brief outline of the proof is given in Karr 1991, p. 59. It relies on first proving the statement for the elementary $\{\mathcal{F}_t\}_{t\geq t_0}$-predictable processes which, according to Equation (A.12), have the form

$$f_t = \mathbb{1}_A \cdot \mathbb{1}_{\{u<t\leq v\}}\,, \quad \text{where } t_0 \leq u \leq v\,, A \in \mathcal{F}_u\,, \tag{A.40}$$

and then applying the monotone class theorem as seen in the proof of Theorem A.33. Since $|f_t| \leq 1$ for such an elementary process, the condition (A.38) is satisfied for any martingale of locally integrable variation. Substituting f into Equation (A.39), we get:

$$X_t = \mathbb{1}_A \cdot \left(M_{\min\{v,\max\{u,t\}\}} - M_u\right)\,. \tag{A.41}$$

It is shown in an addendum in Appendix A.4 that X is indeed a martingale. The profound

[44] Accordingly, $|M|_t < \infty$ is the total variation of the path $s \mapsto M_s$ over $[t_0, t]$ and is not to be confused with $|M_t|$ (Karr 1991, p. 59).

proof employs functional analytic tools such as the Hahn-Banach extension theorem. For details, see Brémaud 1981, pp. 10–11. □

Lemma A.40 further consolidates the role of predictable processes in the context of martingale theory. If M is a martingale of integrable bounded variation, each of the elementary predictable processes $f = (f_t)_{t \geq t_0}$ given in Equation (A.40) satisfies the identity

$$\mathbb{E}\left[\int_{t_0}^{\infty} f_s \, \mathrm{d}M_s\right] = 0\,, \tag{A.42}$$

a direct consequence of X being a martingale with $X_{t_0} = 0$. But the converse is equally true: If M is an adapted process of integrable bounded variation satisfying Equation (A.42) for every such f, then M is already a martingale. This is evident from the following observation, where for arbitrary $t_0 \leq u \leq v$ we have:

$$\begin{aligned}
& \mathbb{E}\left[\int_{t_0}^{\infty} \mathbb{1}_A \mathbb{1}_{\{u<s\leq v\}} \, \mathrm{d}M_s\right] = 0\,, \quad \text{for all } A \in \mathcal{F}_u\,, \\
\iff\quad & \mathbb{E}\left[\mathbb{1}_A \left(M_v - M_u\right)\right] = 0\,, \quad \text{for all } A \in \mathcal{F}_u\,, \\
\iff\quad & \int_A M_v \, \mathrm{d}\mathbb{P} = \int_A M_u \, \mathrm{d}\mathbb{P}\,, \quad \text{for all } A \in \mathcal{F}_u\,, \\
\iff\quad & \mathbb{E}\left(M_v \,|\, \mathcal{F}_u\right) = M_u\,,
\end{aligned}$$

which is the martingale property from Definition A.17(iii). The elementary predictable processes can therefore even be considered as part of the definition of martingales (cf. Brémaud 1981, p. 11). The requirement of predictability in the construction of the integral

$$\int_{t_0}^{t} f_s \, \mathrm{d}M_s$$

therefore comes as no surprise, but its main purpose is to compensate the jumps of the martingale M. If M were not to exhibit any jumps - as it is the case for the Brownian motion - one could settle for the adaptedness of the integrand instead (see for example Øksendal 2013, p. 33). For a more elaborate motivation, see Kuo 2006, pp. 75–80. We have now acquired all the necessary prior knowledge to give the Itô isometry for square-integrable martingales. It provides a natural extension of the eponymous result usually formulated only for the Brownian motion, see Øksendal 2013, p. 29. We state it specifically for the case where M is the innovation martingale from the Doob-Meyer decomposition of a counting process N.

Theorem A.41 (Itô Isometry for Square-Integrable Martingales; cf. Kuo 2006, p. 88)**.** *Given the situation from Theorem A.39, the martingale M meets the conditions from Lemma A.40. For any predictable process f satisfying*

$$\mathbb{E}\left[\int_{t_0}^{t} |f|^2 \, \mathrm{d}\Lambda_s\right] < \infty\,, \quad \textit{for all}\; t \geq t_0\,, \tag{A.43}$$

the following equality holds:

$$\mathbb{E}\left(|X_t|^2\right) = \mathbb{E}\left[\left|\int_{t_0}^{t} f_s \, \mathrm{d}M_s\right|^2\right] = \mathbb{E}\left[\int_{t_0}^{t} |f_s|^2 \, \mathrm{d}\Lambda_s\right]\,. \tag{A.44}$$

Proof. In Chapter 4 of Kuo 2006, pp. 43–48, the original Itô isometry for the Brownian motion is proved (see Theorem 4.3.5). The proof is carried out in two major steps: First, Lemma 4.3.2 verifies the isometry for elementary processes, and this can be easily generalized to square-integrable martingales. Second, Lemma 4.3.3 shows that the isometry (and even the definition of the stochastic integral itself) extends to all adapted square-integrable processes. To generalize the second part, one requires predictability instead of adaptedness. But this is covered in Kallianpur 1980, p. 52, and therefore the result remains valid. □

Remark A.42 (The Isometric Property in Theorem A.41)**.**
Borrowing the notation from Kuo 2006, p. 84, we denote the space of all predictable processes with index set $[t_0, t^0]$ satisfying Equation (A.43) as $\mathrm{L}^2_{\text{pred}}\left([t_0, t^0]_\Lambda \times \Omega\right)$. Similarly, let $\mathrm{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ be the space of square-integrable random variables defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then, the mappings $\mathbb{I}_{t^0}$ defined for each $t^0 > t_0$ by

$$\mathbb{I}_{t^0} \, : \, \mathrm{L}^2_{\text{pred}}\left(\left[t_0, t^0\right]_\Lambda \times \Omega\right) \longrightarrow \mathrm{L}^2(\Omega, \mathcal{F}, \mathbb{P}) \, : \, f \longmapsto \int_{t_0}^{t^0} f_s \, \mathrm{d}M_s$$

are isometric according to Equation (A.44).

A.3. The Hazard Transformation

Numerous methods of statistical inference are based on knowledge of the probability distributions involved: For instance, the well-known Kolmogorov-Smirnov test relies on the Kolmogorov distribution, which emerges as the asymptotic distribution of the test statistic[45]. In practise, the underlying distributions are often too complex to be computed explicitly and, if they are, too unwieldy to work with. However, the transformation of random variables or processes often results in simpler, familiar distributions, with the most prominent example being the (inverse) probability integral transform. The aforementioned Kolmogorov-Smirnov test implements this approach, where the probability integral transform is utilized to achieve an asymptotic distribution that is independent of the (continuous) cumulative distribution function under consideration. Notably, while this transformation preserves as much information as possible, it also requires that the particular cumulative distribution function is known in advance. This considerably limits its applicability as soon as more complex random structures such as continuous-parameter processes are involved: In our situation, the (finite dimensional) distribution of a simple point process is generally unknown or not conveniently expressible. Consequently, we are interested in addressing a transformation that is more tailored to our situation: the hazard transformation. The application of the hazard transformation yields independent and exponentially distributed random variables under relatively liberal assumptions. Nonetheless, this comes at a price: the hazard transform not only exhibits a striking resemblance to the inverse probability integral transform, it also bears similar restrictions. In particular, to utilize the methods presented in this subsection, knowledge of the underlying cumulative intensity process or equivalently the compensator of the associated counting process is required. This appears to be an equally significant constraint at first glance, but proves to be helpful in the intensity-based modelling approach on which this

[45] Consequently, the name originates from Andrei N. Kolmogorov, who derived both the test statistic and its asymptotic distribution named in his honour (Kolmogorov 1933), and Nikolai V. Smirnov, who published a table of said distribution (Smirnov 1939).

thesis rests. Beforehand, we start with a motivational introduction on why the hazard transformation is suited to our needs.

Remark A.43 (The Hazard Transform as a Random Time Change; Daley and Vere-Jones 2003, pp. 257–261)**.**
The following description of what we will refer to as the hazard transformation is due to Papangelou, who formulated and proved an earlier version of the subsequent Theorem A.44, which was limited to stationary point processes (cf. Theorem 5 of Papangelou 1974, p. 132). The original citation is given in Harding and Kendall 1974, while an adapted rendition can also be found in Daley and Vere-Jones 2003, p. 258. Merely adjusting the notation to fit this thesis, it reads:

> "Suppose that starting at 0 say, we trace the positive half-line $[0, \infty)$ in such a way that at the time we are passing position t our speed is $1/\lambda^*(t)$, which can be ∞. (The value of $\lambda^*(t)$ is determined by the observations of the past, i.e. of what happened in $[t_0, t)$.) Then the time instants at which we shall meet all the non-negative points of the process form a homogeneous Poisson process."

In the language of intensity theory, the above statement means that the random time change

$$[t_0, \infty) \longrightarrow [0, \infty) \, : \, t \longmapsto \Lambda(t) \tag{A.45}$$

transforms the point process with conditional intensity function λ^* into a homogeneous Poisson process with intensity 1. The above heuristic is formalized in Theorem A.44.

Theorem A.44 (Random Time Transformation for Adapted Counting Processes; cf. Daley and Vere-Jones 2003, p. 258 and Brémaud 1981, p. 40)**.**
Let N be a counting process defined on the filtered probability space $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq t_0}, \mathbb{P}\right)$. Suppose that N is adapted and let Λ denote its $\left(\mathbb{P}, \{\mathcal{F}_t\}_{t \geq t_0}\right)$-compensator which admits a strictly positive conditional intensity λ^*. If N is non-terminating, that is,

$$\mathbb{P}\left(\lim_{t \to \infty} N_t = \infty\right) = 1 \,,$$

then the transformed process $\tilde{N} = \left(\tilde{N}_t\right)_{t \geq 0}$ defined by

$$\tilde{N}_t := N_{\Lambda^{-1}(t)} \tag{A.46}$$

is a homogeneous Poisson process with intensity 1*.*

Proof. A sketch of proof can be found in Daley and Vere-Jones 2003, pp. 258–259, but technical details are largely ignored. For instance, the filtration $\{\mathcal{F}_t\}_{t \geq t_0}$ is transformed in conjunction with the counting process, but both a rigorous definition and a strict proof are omitted. A fuller discussion is deferred to Chapter 14 of Daley and Vere-Jones 2008, culminating in a reformulation of the above theorem in Proposition 14.6.III. Since the necessary preliminary considerations would exceed the framework of this thesis, no further elaboration of this particular Proposition is given.
The existence of the conditional intensity λ^* implies that Λ is continuous. If, moreover, λ^* is strictly positive, then Λ is strictly increasing on $[t_0, \infty)$ and thus a continuous inverse Λ^{-1} exists. However, an analogous result holds even if only a generalized inverse for Λ

can be found, as an alternative proof in Brémaud 1981, pp. 40–43 demonstrates. Once again, that proof involves stopping times, which we decided not to address in Subsection A.2. We therefore retain the requirement that λ be strictly positive to enable a more accessible representation consistent with the notation of previous subsections. Finally, the requirement that N is non-terminating ensures that there is no final point of the process. As stated and illustrated by an example in Daley and Vere-Jones 2003, p. 260, "the basic result remains valid without it, except insofar as the final interval is then infinite and so cannot belong to a [homogeneous] Poisson-process". □

A counting process N satisfying the conditions of Theorem A.44 is sometimes called a *process of Poisson type*, since "all such processes can be derived from a simple Poisson process by a random time transformation" (Daley and Vere-Jones 2003, p. 259). An important implication is that these processes can be easily simulated by the inversion method (the procedure is described in Algorithm 7.4.III. of Daley and Vere-Jones 2003, p. 260). As the later Theorem A.46 shows, this only requires knowledge of the inverse conditional cumulative hazard function.

Example A.45 (Time Change for the Homogeneous Poisson Process).
If $N = (N_t)_{t\geq 0}$ is a homogeneous Poisson process with intensity λ, we have $\Lambda^{-1}(t) = \frac{t}{\lambda}$ according to Example A.28. Then, the transformed process $\tilde{N}$ with $\tilde{N}_t := N_{\frac{t}{\lambda}}$ is again a homogeneous Poisson process and has intensity 1:

$$\begin{aligned}\mathbb{P}\left(\tilde{N}\left(s,t\right)=n\right) &= \mathbb{P}\left(N\left(\frac{s}{\lambda},\frac{t}{\lambda}\right)=n\right)\\ &= \frac{1}{n!}\left(\lambda\left(\frac{t}{\lambda}-\frac{s}{\lambda}\right)\right)^{n}\mathrm{e}^{-\lambda\left(\frac{t}{\lambda}-\frac{s}{\lambda}\right)}\\ &= \frac{1}{n!}(t-s)^{n}\mathrm{e}^{-(t-s)}\,.\end{aligned}$$

The independence of the increments of $\tilde{N}$ is immediately clear, as it is likewise inherited from N. Keep in mind that the time change here is not random, but deterministic.

There remains one question for us to answer in this section: how can the random time change be related to the (cumulative conditional) hazard function that lends its name to the hazard transformation? For this, we need a change of perspective: Instead of studying the effects of the random time change on the counting process N via Equation (A.46), we return to the mapping from Equation (A.45), which immediately tells us how the simple point process $T = (T_i)_{i\in\mathbb{N}}$ associated with N is transformed. We recall the Equations (A.5) and (A.6) from Example A.7 to remind us that the interarrival times of a homogeneous Poisson process with intensity 1 follow a standard exponential distribution. Informally, we obtain:

$$\Lambda\left(T_i\right)-\Lambda\left(T_{i-1}\right)\overset{\text{i.i.d.}}{\sim}\mathcal{E}(1)\,,\quad i\in\mathbb{N}\,. \tag{A.47}$$

In the setting of Summary 1, we identified the conditional intensity λ^* as a piecewise conglomerate of conditional hazard functions, and this relationship carried over to the cumulative intensity process - or, in other words, the compensator. If we proceed to substitute Equation (A.47) into Equation (A.34), we observe the following connection that justifies the term hazard transformation:

$$H_i\left(T_i \,|\, T_1,\ldots,T_{i-1}\right) \overset{\text{i.i.d}}{\sim} \mathcal{E}(1)\,, \quad i \in \mathbb{N}\,. \tag{A.48}$$

Basically, the statement of Equation (A.48) follows directly from Theorem A.44, but we still owe a proof of it. We provide a basic one that relies only on the assumptions of Summary 1 and therefore can be understood without consulting the previous subsections. Beforehand, the hazard transform shall finally be established in the following theorem.

Theorem A.46 (Hazard Transformation)**.**
In the situation of Summary 1, let $T = (T_i)_{i\in\mathbb{N}}$ *be a simple point process and assume* $f_i\left(t \,|\, t_1,\ldots,t_{i-1}\right)$ *to be continuous with* $f_i\left(t \,|\, t_1,\ldots,t_{i-1}\right) > 0$ *almost everywhere for each*[46] $t_1 < \ldots < t_{i-1} \in [0,\infty)$*. Then,* $H_i\left(T_i \,|\, T_1,\ldots,T_{i-1}\right)$ *is exponentially distributed with parameter* 1*. The transformed process* $R = (R_i)_{i\in\mathbb{N}}$ *with*

$$R_i := H_i\left(T_i \,|\, T_1,\ldots,T_{i-1}\right) \tag{A.49}$$

is called the hazard transformation of T*. Accordingly, it holds:*

$$R_1, R_2, \ldots \overset{\text{i.i.d.}}{\sim} \mathcal{E}\left(1\right)\,. \tag{A.50}$$

Proof. Throughout the proof, we will use the abbreviation from Remark 2.1 wherever suitable, but switch to the conventional notation situationally. If we assume the conditional density function of T_i given $T_{1:(i-1)} = t_{1:(i-1)}$ (i.e., $T_1 = t_1, \ldots, T_{i-1} = t_{i-1}$) to be continuous for each $i \in \mathbb{N}$, the associated survival function $S_i\left(t \,\middle|\, t_{1:(i-1)}\right)$ is a differentiable function of t. Thus, it holds by virtue of the fundamental theorem of calculus and Equation (A.29) for all $t > t_{i-1}$:

$$\begin{aligned}
H_i\left(t \,\middle|\, t_{1:(i-1)}\right) &\overset{\text{(A.31)}}{=} \int_{t_{i-1}}^{t} h_i\left(u \,\middle|\, t_{1:(i-1)}\right)\,\mathrm{d}u \overset{\text{(A.30)}}{=} \int_{t_{i-1}}^{t} -\frac{\frac{\partial}{\partial u} S_i\left(u \,\middle|\, t_{1:(i-1)}\right)}{S_i\left(u \,\middle|\, t_{1:(i-1)}\right)}\,\mathrm{d}u \\
&= \int_{t_{i-1}}^{t} -\frac{\partial}{\partial u}\ln\left(S_i\left(u \,\middle|\, t_{1:(i-1)}\right)\right)\mathrm{d}u \\
&= \ln\left(S_i\left(t_{i-1} \,\middle|\, t_{1:(i-1)}\right)\right) - \ln\left(S_i\left(t \,\middle|\, t_{1:(i-1)}\right)\right) \\
&= -\ln\left(S_i\left(t \,\middle|\, t_{1:(i-1)}\right)\right)\,.
\end{aligned} \tag{A.51}$$

Let $F_i\left(t \,\middle|\, t_{1:(i-1)}\right)$ denote the conditional cumulative distribution function (*conditional CDF*) of T_i, so that $F_i\left(t \,\middle|\, t_{1:(i-1)}\right) = 1 - S_i\left(t \,\middle|\, t_{1:(i-1)}\right)$ according to Equation (A.29). As $f_i\left(t \,\middle|\, t_{1:(i-1)}\right) > 0$ a.e., the conditional CDF of T_i is strictly increasing and therefore has an inverse denoted with $F_i^{-1}\left(\cdot \,\middle|\, t_{1:(i-1)}\right)$. It holds for arbitrary $u \in (0,1)$:

$$\begin{aligned}
&\mathbb{P}\left(F_i\left(T_i \,\middle|\, T_{1:(i-1)}\right) \leq u \,\middle|\, T_{1:(i-1)} = t_{1:(i-1)}\right) \\
&\quad = \mathbb{P}\left(F_i\left(T_i \,\middle|\, t_{1:(i-1)}\right) \leq u \,\middle|\, T_{1:(i-1)} = t_{1:(i-1)}\right) \\
&\quad = \mathbb{P}\left(T_i \leq F_i^{-1}\left(u \,\middle|\, t_{1:(i-1)}\right) \,\middle|\, T_{1:(i-1)} = t_{1:(i-1)}\right)
\end{aligned}$$

[46] Because of Equations (A.30) and (A.32), we could equivalently require that $\lambda(t) > 0$ on $\{t \leq T_i\}$.

$$= F_i\left(F_i^{-1}\left(u \,|\, t_{1:(i-1)}\right) \,|\, t_{1:(i-1)}\right) = u\,, \tag{A.52}$$

so that the conditional distribution of $F_i\left(T_i \,|\, T_{1:(i-1)}\right)$ given $T_{1:(i-1)} = t_{1:(i-1)}$ is a uniform distribution on $[0,1]$. Hence, it is easy to see that the conditional distribution of $S_i\left(T_i \,|\, T_{1:(i-1)}\right)$ given $T_{1:(i-1)} = t_{1:(i-1)}$ is also a uniform distribution:

$$\begin{aligned}
&\mathbb{P}\left(S_i\left(T_i \,|\, T_{1:(i-1)}\right) \le u \,|\, T_{1:(i-1)} = t_{1:(i-1)}\right) \\
&\quad= 1 - \mathbb{P}\left(S_i\left(T_i \,|\, T_{1:(i-1)}\right) \ge u \,|\, T_{1:(i-1)} = t_{1:(i-1)}\right) \\
&\quad\overset{(A.29)}{=} 1 - \mathbb{P}\left(F_i\left(T_i \,|\, T_{1:(i-1)}\right) \le 1-u \,|\, T_{1:(i-1)} = t_{1:(i-1)}\right) = 1-(1-u) = u\,,
\end{aligned}$$

where again $u \in (0,1)$. This eventually implies that for all $t \in (0,\infty)$ we have:

$$\begin{aligned}
&\mathbb{P}\left(H_i\left(T_i \,|\, T_{1:(i-1)}\right) \le t \,|\, T_{1:(i-1)} = t_{1:(i-1)}\right) \\
&\quad\overset{(A.51)}{=} \mathbb{P}\left(-\ln\left(S_i\left(T_i \,|\, T_{1:(i-1)}\right)\right) \le t \,|\, T_{1:(i-1)} = t_{1:(i-1)}\right) \\
&\quad= \mathbb{P}\left(S_i\left(T_i \,|\, T_{1:(i-1)}\right) \ge \exp(-t) \,|\, T_{1:(i-1)} = t_{1:(i-1)}\right) = 1-\exp(-t)\,,
\end{aligned} \tag{A.53}$$

which means that the conditional distribution of $R_i = H_i\left(T_i \,|\, T_{1:(i-1)}\right)$ given $T_{1:(i-1)} = t_{1:(i-1)}$ is an exponential distribution with parameter 1. Note that due to the factorization lemma (see Lemma 11.7, Bauer 2001, p. 62), for each $t \in (0,\infty)$ there exists a measurable function $g_{i,t}$ such that

$$\mathbb{P}\left(R_i \le t \,|\, T_1,\ldots,T_{i-1}\right) = \mathbb{E}\left(\mathbb{1}_{\{R_i \le t\}} \,|\, T_1,\ldots,T_{i-1}\right) = g_{i,t} \circ \left(T_1,\ldots,T_{i-1}\right)^\top .$$

We thus conclude by the law of total expectation[47] and Equation (A.53):

$$\begin{aligned}
\mathbb{P}\left(R_i \le t\right) &= \mathbb{E}\left[\mathbb{E}\left(\mathbb{1}_{\{R_i \le t\}} \,|\, T_1,\ldots,T_{i-1}\right)\right] = \mathbb{E}\left[g_{i,t} \circ \left(T_1,\ldots,T_{i-1}\right)^\top\right] \\
&= \int \underbrace{g_{i,t}\left(t_1,\ldots,t_{i-1}\right)}_{\overset{(A.53)}{=} 1-\exp(-t)} f_{(T_1,\ldots,T_{i-1})}\left(t_1,\ldots,t_{i-1}\right) \mathrm{d}\lambda^{i-1}\left(t_1,\ldots,t_{i-1}\right) \\
&= (1-\exp(-t)) \underbrace{\int f_{(T_1,\ldots,T_{i-1})}\left(t_1,\ldots,t_{i-1}\right) \mathrm{d}\lambda^{i-1}\left(t_1,\ldots,t_{i-1}\right)}_{=1} = 1-\exp(-t)\,,
\end{aligned}$$

where again $t \in (0,\infty)$ and λ^d denotes the d-dimensional Lebesgue measure that should not be confused with the previously introduced conditional intensity function. This shows that the random variable R_i follows an exponential distribution with parameter 1, for each $i \in \mathbb{N}$. To complete the proof, we need to show the independence of $\left(R_i\right)_{i\in\mathbb{N}}$.
Once again, this result emerges as an immediate consequence of Equation (A.53). Note that for $i < j \in \mathbb{N}$, we have $\sigma\left(T_1,\ldots,T_i\right) \subset \sigma\left(T_1,\ldots,T_{j-1}\right)$. Therefore, $R_i = H_i\left(T_i \,|\, T_1,\ldots,T_{i-1}\right)$ is measurable with respect to the σ-algebra $\sigma\left(T_1,\ldots,T_{j-1}\right)$ as a function of $T_1,\ldots,T_i$. For arbitrary $t^{(1)}, t^{(2)} \in (0,\infty)$ we thus obtain by virtue of the

[47] Equation (A.53) readily implies $g_{i,t} \circ \left(T_1,\ldots,T_{i-1}\right)^\top = 1-\exp(-t)$ almost surely, so that one can easily deduce $\mathbb{E}\left[g_{i,t} \circ \left(T_1,\ldots,T_{i-1}\right)^\top\right] = 1-\exp(-t)$.

tower property:

$$\begin{aligned}
\mathbb{P}\left(R_i \leq t^{(1)}, R_j \leq t^{(2)}\right) &= \mathbb{E}\left[\mathbb{E}\left(\mathbb{E}\left(\mathbb{1}_{\left\{R_i \leq t^{(1)}, R_j \leq t^{(2)}\right\}} \,\middle|\, T_{1:(j-1)}\right) \,\middle|\, T_{1:(i-1)}\right)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left(\mathbb{E}\left(\mathbb{1}_{\left\{R_i \leq t^{(1)}\right\}} \cdot \mathbb{1}_{\left\{R_j \leq t^{(2)}\right\}} \,\middle|\, T_{1:(j-1)}\right) \,\middle|\, T_{1:(i-1)}\right)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left(\mathbb{1}_{\left\{R_i \leq t^{(1)}\right\}} \cdot \mathbb{E}\left(\mathbb{1}_{\left\{R_j \leq t^{(2)}\right\}} \,\middle|\, T_{1:(j-1)}\right) \,\middle|\, T_{1:(i-1)}\right)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left(\mathbb{1}_{\left\{R_i \leq t^{(1)}\right\}} \cdot \underbrace{g_{j,t^{(2)}} \circ (T_1, \ldots, T_{j-1})^\top}_{\equiv 1-\exp\left(-t^{(2)}\right) \text{ almost surely.}} \,\middle|\, T_{1:(i-1)}\right)\right] \\
&= \left(1 - \exp\left(-t^{(2)}\right)\right) \cdot \mathbb{E}\left[\mathbb{E}\left(\mathbb{1}_{\left\{R_i \leq t^{(1)}\right\}} \,\middle|\, T_{1:(i-1)}\right)\right] \\
&= \left(1 - \exp\left(-t^{(2)}\right)\right) \cdot \mathbb{E}\Big[\underbrace{g_{j,t^{(1)}} \circ (T_1, \ldots, T_{i-1})^\top}_{\equiv 1-\exp\left(-t^{(1)}\right) \text{ almost surely.}}\Big] \\
&= \left(1 - \exp\left(-t^{(2)}\right)\right) \cdot \left(1 - \exp\left(-t^{(1)}\right)\right) \\
&= \mathbb{P}\left(R_i \leq t^{(1)}\right) \cdot \mathbb{P}\left(R_j \leq t^{(2)}\right), \qquad \text{(A.54)}
\end{aligned}$$

which implies the independence of R_i and R_j. The independence of $(R_i)_{i\in\mathbb{N}}$ now easily follows by induction:
Let $\mathbb{I}_{n+1}$ be an arbitrary subset of $\mathbb{N}$ with cardinality $|\mathbb{I}_{n+1}| = n+1$, $\mathbb{I}_{n+1} = \{i_1, \ldots, i_{n+1}\}$ and assume that the independence of the collection $\{R_i : i \in \mathbb{I}_n\}$ is shown for each $\mathbb{I}_n \subset \mathbb{N}$ with $|\mathbb{I}_n| = n$. By Equation (A.54), this is true for $n = 2$.
The independence of $\{R_i : i \in \mathbb{I}_{n+1}\}$ can then be shown analogously to Equation (A.54), as for all $t^{(1)}, \ldots, t^{(n+1)} \in (0, \infty)$ it holds:

$$\begin{aligned}
&\mathbb{P}\left(R_{i_1} \leq t^{(1)}, \ldots, R_{i_n} \leq t^{(n)}, R_{i_{n+1}} \leq t^{(n+1)}\right) \\
&\quad = \mathbb{E}\left[\mathbb{E}\left(\mathbb{1}_{\left\{R_{i_1} \leq t^{(1)}, \ldots, R_{i_n} \leq t^{(n)}, R_{i_{n+1}} \leq t^{(n+1)}\right\}} \,\middle|\, T_1, \ldots, T_{i_{n+1}-1}\right)\right] \\
&\quad = \mathbb{E}\left[\mathbb{1}_{\left\{R_{i_1} \leq t^{(1)}, \ldots, R_{i_n} \leq t^{(n)}\right\}} \cdot g_{i_{n+1}, t^{(n+1)}} \circ \left(T_1, \ldots, T_{i_{n+1}-1}\right)^\top\right] \\
&\quad = \left(1 - \exp\left(-t^{(n+1)}\right)\right) \cdot \mathbb{E}\left[\mathbb{1}_{\left\{R_{i_1} \leq t^{(1)}, \ldots, R_{i_n} \leq t^{(n)}\right\}}\right] \\
&\quad = \mathbb{P}\left(R_{i_{n+1}} \leq t^{(n+1)}\right) \cdot \mathbb{P}\left(R_{i_1} \leq t^{(1)}, \ldots, R_{i_n} \leq t^{(n)}\right) \\
&\quad = \mathbb{P}\left(R_{i_{n+1}} \leq t^{(n+1)}\right) \cdot \mathbb{P}\left(R_{i_1} \leq t^{(1)}\right) \cdot \ldots \cdot \mathbb{P}\left(R_{i_n} \leq t^{(n)}\right),
\end{aligned}$$

by assumption as $|\{i_1, \ldots, i_n\}| = n$, thus completing the proof of Theorem A.46. In the course of this, we observed that conditioning on $T_1, \ldots, T_{i_n-1}$ is not required as the probability $\mathbb{P}\left(R_{i_1} \leq t^{(1)}, \ldots, R_{i_n} \leq t^{(n)}\right)$ breaks down into the known factors $\mathbb{P}\left(R_{i_k} \leq t^{(k)}\right)$ for $1 \leq k \leq n$ considered earlier. In hindsight, the same holds true for Equation (A.54), so that the outer condition could theoretically be neglected. Nevertheless the author decided to maintain the above form in order to allow a better understanding of the calculations shown. □

We conclude this section with a useful corollary of Theorem A.46. It shows that if the conditional hazard function $h_i(\cdot \,|\, T_{1:(i-1)})$ for some $i \in \mathbb{N}$ is $\mathbb{P}$-almost surely constant, then the interarrival time $T_i - T_{i-1}$ also follows an exponential distribution.

Corollary A.47 (Interarrival Times at a Constant Conditional Hazard Function)**.** *In the situation of Theorem A.46, let $i \in \mathbb{N}$ and suppose that $h_i(\cdot \,|\, T_1, \dots, T_{i-1})$ is $\mathbb{P}$-almost surely constant. Then,*

$$h_i(T_{i-1} \,|\, T_1, \dots, T_{i-1})\,(T_i - T_{i-1}) \sim \mathcal{E}(1)\,.$$

We write informally

$$T_i - T_{i-1} \sim \mathcal{E}\big(h_i(T_{i-1} \,|\, T_1, \dots, T_{i-1})\big)\,,$$

which means that the conditional distribution of the interarrival time $T_i - T_{i-1}$ given $T_{1:(i-1)} = t_{1:(i-1)}$ is exponential with rate $h_i(t_{i-1} \,|\, t_1, \dots, t_{i-1})$.

Proof. Let us define the random variable

$$h_i(T_{i-1} \,|\, T_1, \dots, T_{i-1})(\omega) =: \lambda_i(\omega)\,.$$

If $h_i(\cdot \,|\, T_1, \dots, T_{i-1})$ is $\mathbb{P}$-almost surely constant, it follows that, with probability one,

$$h_i(\cdot \,|\, T_1, \dots, T_{i-1}) \equiv \lambda_i\,.$$

According to Theorem A.46,

$$H_i\left(T_i \,|\, T_1, \dots, T_{i-1}\right) \sim \mathcal{E}(1)\,. \tag{A.55}$$

By the definition of the cumulative conditional hazard function in Equation (A.31), we thus obtain:

$$\begin{aligned} H_i\left(T_i \,|\, T_1, \dots, T_{i-1}\right) &= \int_{T_{i-1}}^{T_i} h_i(u \,|\, T_1, \dots, T_{i-1})\,\mathrm{d}u \\ &= \int_{T_{i-1}}^{T_i} \lambda_i\,\mathrm{d}u = \lambda_i\left(T_i - T_{i-1}\right)\,. \end{aligned} \tag{A.56}$$

Substituting Equation (A.56) into Equation (A.55) directly yields

$$\lambda_i\left(T_i - T_{i-1}\right) \sim \mathcal{E}(1)\,.$$

The rest of the assertion then follows from the scaling property of the exponential distribution after division by λ_i, and since λ_i is by construction a function of $T_{1:(i-1)}$. □

A.4. Complementary Proofs

To conclude Appendix A, we provide the proofs that were omitted from Sections A.1 through A.3. For ease of reference, we restate the corresponding result in each case.

Proof of Remark A.6, part (ii)

Remark A.6. We can easily recognize how the properties that define a simple point process are transferred to the associated counting process. Let again $T = (T_i)_{i \in \mathbb{N}}$ denote a simple point process and $N = (N_t)_{t \in [t_0, \infty)}$ its associated counting process.

(ii) The second condition ensures that almost surely the sample paths of the process increase only in jumps of size 1.
Accordingly, for $\mathbb{P}$-almost all $\omega \in \Omega$ it holds that $\Delta N_t(\omega) \in \{0, 1\}$ for all t, where for any *càdlàg*[48] function f the function Δf is defined via (cf. Jacobsen 2006, p. 12)

$$t \longmapsto \Delta f(t) := f(t) - f(t-) := f(t) - \lim_{s \uparrow t} f(s) \, .$$

Proof. We need to prove that

$$\mathbb{P}\big(\{\omega \in \Omega \, : \, \underbrace{\Delta N_t(\omega)}_{\in \overline{\mathbb{N}}_0} \in \{0,1\} \text{ for all } t \in [t_0, \infty) \, \}\big) = 1 \, ,$$

or equivalently by transition to the complementary event:

$$\mathbb{P}\left(\{\omega \in \Omega \, : \, \exists t \in [t_0, \infty) \text{ such that } \Delta N_t(\omega) \geq 2\}\right) = 0 \, .$$

Note that the following holds:

$$\begin{aligned} &\{\omega \in \Omega \, : \, \exists t \in [t_0, \infty) \text{ such that } \Delta N_t(\omega) \geq 2\} \\ &\quad \subset \{\omega \in \Omega \, : \, \exists t \in [t_0, \infty) \; \exists i < j \text{ such that } T_i(\omega) = T_j(\omega) = t\} \\ &\quad \subset \bigcup_{i \in \mathbb{N}} \bigcup_{j > i} \left(\{T_i = T_j\} \cap \{T_j < \infty\}\right) \, . \end{aligned} \tag{A.57}$$

By condition (i) of Definition A.3 we obtain for any $i \in \mathbb{N}$:

$$\begin{aligned} &\mathbb{P}\Big(\bigcup_{j>i} \left(\{T_i = T_j\} \cap \{T_j < \infty\}\right)\Big) \\ &\quad = \mathbb{P}\Big(\bigcup_{j>i} \left(\{T_i = T_j\} \cap \{T_j < \infty\}\right) \cap \{t_0 < T_1 \leq T_2 \leq \ldots\}\Big) \\ &\quad = \mathbb{P}\Big(\bigcup_{j>i} (\underbrace{\{T_i = T_{i+1} = \ldots = T_j\}}_{\subset \{T_i = T_{i+1}\}} \cap \underbrace{\{T_j < \infty\}}_{\subset \{T_i < \infty\}}) \cap \{t_0 < T_1 \leq T_2 \leq \ldots\}\Big) \\ &\quad \leq \mathbb{P}\left(\{T_i = T_{i+1}\} \cap \{T_i < \infty\} \cap \{t_0 < T_1 \leq T_2 \leq \ldots\}\right) \\ &\quad \leq \mathbb{P}\left(\{T_i = T_{i+1}\} \cap \{T_i < \infty\}\right) \, . \end{aligned} \tag{A.58}$$

Combining Equations (A.57) and (A.58) then yields:

$$\begin{aligned} &\mathbb{P}\left(\{\omega \in \Omega \, : \, \exists t \in [t_0, \infty) \text{ such that } \Delta N_t(\omega) \geq 2\}\right) \\ &\quad \leq \mathbb{P}\Big(\bigcup_{i \in \mathbb{N}} \bigcup_{j > i} \{T_i = T_j\} \cap \{T_j < \infty\}\Big) \\ &\quad \leq \sum_{i \in \mathbb{N}} \mathbb{P}\left(\{T_i = T_{i+1}\} \cap \{T_i < \infty\}\right) \, . \end{aligned}$$

We complete the proof by showing that $\mathbb{P}\left(\{T_i = T_{i+1}\} \cap \{T_i < \infty\}\right) = 0$ for each $i \in \mathbb{N}$.

[48] Abbreviation of the French term *continue à droite, limite à gauche*, i.e. right-continuous with left limits.

Note that according to condition (ii) of Definition A.3 we have:

$$
\begin{aligned}
&\mathbb{P}\left(\{T_i < T_{i+1}\} \cap \{T_i < \infty\}\right) \\
&\quad = \mathbb{P}\left(\{T_i < \infty\}\right) \\
&\quad = \mathbb{P}\Big(\left(\{T_i < T_{i+1}\} \cap \{T_i < \infty\}\right) \cup \left(\{T_i \geq T_{i+1}\} \cap \{T_i < \infty\}\right)\Big) \\
&\quad = \mathbb{P}\left(\{T_i < T_{i+1}\} \cap \{T_i < \infty\}\right) + \mathbb{P}\left(\{T_i \geq T_{i+1}\} \cap \{T_i < \infty\}\right), \qquad \text{(A.59)}
\end{aligned}
$$

so by subtracting $\mathbb{P}\left(\{T_i < T_{i+1}\} \cap \{T_i < \infty\}\right)$ on both sides of Equation (A.59) we obtain:

$$\mathbb{P}\left(\{T_i = T_{i+1}\} \cap \{T_i < \infty\}\right) \leq \mathbb{P}\left(\{T_i \geq T_{i+1}\} \cap \{T_i < \infty\}\right) = 0\,,$$

thereby finishing the proof of Remark A.6, part (ii). □

Proof of Proposition A.9

Proposition A.9 (Measurability of Right-Continuous Stochastic Processes)**.**
Let $X = (X_t)_{t \in \mathcal{I}}$ be a right-continuous stochastic process, where $\mathcal{I} \subset \mathbb{R}$ is an interval and the state space is $\left(\mathbb{R}^d, \mathcal{B}\left(\mathbb{R}^d\right)\right)$ for some integer d. Then, X is measurable.

Proof. Without loss of generality we assume $\mathcal{I} = [t_0, \infty)$ with $t_0 = 0$. For all $n \in \mathbb{N}_0$ and $t \in \mathcal{I}$, let

$$
\begin{aligned}
X^{(n)} \;&: \mathcal{I} \times \Omega \longrightarrow (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \\
&(t, \omega) \longmapsto X_t^{(n)}(\omega) := \sum_{k=0}^{\infty} X_{(k+1)2^{-n}}(\omega) \cdot \mathbb{1}_{[k2^{-n},(k+1)2^{-n})}(t)\,, \qquad \text{(A.60)}
\end{aligned}
$$

so that for each n the interval $\mathcal{I}$ is divided into infinitely many subintervals of length 2^{-n}, where on each subinterval $[k2^{-n}, (k+1)2^{-n})$ the sample path $t \mapsto X_t(\omega)$ is uniformly approximated by its value on the right edge of the interval (i.e., $X_{(k+1)2^{-n}}(\omega)$).
The proof is carried out in two steps by showing:

1. $X^{(n)}$ is $(\mathcal{B}(\mathcal{I}) \otimes \mathcal{F})$-$\mathcal{B}(\mathbb{R}^d)$-measurable.

2. $X^{(n)} \to X$ pointwise on $I \times \Omega$ as $n \to \infty$.

For the first step, we obtain for $B \in \mathcal{B}(\mathbb{R}^d)$:

$$
\begin{aligned}
\left(X^{(n)}\right)^{-1}(B) &= \left\{(t,\omega) \in \mathcal{I} \times \Omega \,:\, X_t^{(n)}(\omega) \in B\right\} \\
&= \left\{(t,\omega) \,:\, \exists k \in \mathbb{N}_0 \text{ with } t \in \left[k2^{-n}, (k+1)2^{-n}\right) \text{ and } X_{(k+1)2^{-n}}(\omega) \in B\right\} \\
&= \bigcup_{k=0}^{\infty} \left\{(t,\omega) \,:\, t \in \left[k2^{-n}, (k+1)2^{-n}\right) \text{ and } X_{(k+1)2^{-n}}(\omega) \in B\right\} \\
&= \bigcup_{k=0}^{\infty} \underbrace{\underbrace{\left[k2^{-n}, (k+1)2^{-n}\right)}_{\in \mathcal{B}(\mathcal{I})} \times \underbrace{\left\{X_{(k+1)2^{-n}}(\omega) \in B\right\}}_{\in \mathcal{F} \text{ by Definition A.1.}}}_{\in \mathcal{B}(\mathcal{I}) \otimes \mathcal{F}} \in \mathcal{B}(\mathcal{I}) \otimes \mathcal{F}\,,
\end{aligned}
$$

and thus $X^{(n)}$ is $(\mathcal{B}(\mathcal{I}) \otimes \mathcal{F})$-$\mathcal{B}(\mathbb{R}^d)$-measurable.
For the second step, consider any $(t, \omega) \in \mathcal{I} \times \Omega$ and note that for each n there exists a uniquely determined $k^{(n)} = k_t^{(n)}$ satisfying $t \in \left[k^{(n)}2^{-n}, (k^{(n)}+1)2^{-n}\right)$. Additionally, $\left|t - (k^{(n)}+1)2^{-n}\right| < 2^{-n}$ and $t < (k^{(n)}+1)2^{-n}$ jointly imply $(k^{(n)}+1)2^{-n} \downarrow t$ while $n \to \infty$. As furthermore $X_t^{(n)}(\omega) = X_{\left(k^{(n)}+1\right)2^{-n}}(\omega)$ applies by Equation (A.60), the pointwise convergence follows from the right-continuity of the sample path $t \mapsto X_t(\omega)$:

$$\lim_{n\to\infty} X_t^{(n)}(\omega) = \lim_{n\to\infty} X_{\left(k^{(n)}+1\right)2^{-n}}(\omega) = X_t(\omega)\,.$$

This completes the proof, since pointwise limits of sequences of Borel functions (i.e., measurable functions with state space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ for some integer d) are itself measurable.
□

Proof of Lemma A.15

Lemma A.15 (Completion of a Probability Space; Jacobsen 2006, p. 301)**.**
Let $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space. Then there exists a complete probability space* $\left(\Omega, \overline{\mathcal{F}}, \overline{\mathbb{P}}\right)$ *with* $\mathcal{F} \subset \overline{\mathcal{F}}$ *and*

$$\overline{\mathbb{P}}(A) = \mathbb{P}(A)\,, \quad \textit{for all } A \in \mathcal{F}. \tag{A.10}$$

$\left(\Omega, \overline{\mathcal{F}}, \overline{\mathbb{P}}\right)$ *is called the* completion *of* $(\Omega, \mathcal{F}, \mathbb{P})$.

Proof. Consider the collection of sets $\mathcal{N}$ defined by[49]

$$\mathcal{N} := \{A \subset \Omega \,:\, \exists\, A_0 \in \mathcal{F} \text{ with } \mathbb{P}(A_0) = 0 \text{ and } A \subset A_0\}\,,$$

and let $\overline{\mathcal{F}} = \sigma(\mathcal{F} \cup \mathcal{N})$ be the smallest σ-algebra containing both $\mathcal{F}$ and $\mathcal{N}$. We show:

$$\overline{\mathcal{F}} = \{F \cup N \,:\, F \in \mathcal{F}, N \in \mathcal{N}\}\,. \tag{A.61}$$

Since the inclusion $\supset$ is trivial due to the definition of $\overline{\mathcal{F}}$, it suffices to show that the right-hand side of Equation (A.61) - hereinafter referenced as (RHS) - is itself a σ-algebra: By $\emptyset \in \mathcal{N}$, we have $\emptyset \in$ (RHS) and the fact that the countable union of $\mathbb{P}$-null sets is again a $\mathbb{P}$-null set by virtue of the measure's σ-additivity yields that (RHS) is closed under countable unions. This renders the stability under complementation the remaining property to show. For this, consider any $F \in \mathcal{F}, N \in \mathcal{N}$. By definition there exists $N_0 \in \mathcal{F}$ satisfying $N \subset N_0$ and $\mathbb{P}(N_0) = 0$. We define $\tilde{N} := N_0 \setminus N \in \mathcal{N}$ and compute by consecutive application of *de Morgan's laws*:

$$\begin{aligned}(F \cup N)^{\complement} &= \left(F \cup \left(N_0 \setminus \tilde{N}\right)\right)^{\complement} = F^{\complement} \cap \left(N_0 \setminus \tilde{N}\right)^{\complement} \\ &= F^{\complement} \cap \left(N_0^{\complement} \cup \tilde{N}\right) = \left(F^{\complement} \cap N_0^{\complement}\right) \cup \left(F^{\complement} \cap \tilde{N}\right)\end{aligned}$$

[49] $\mathcal{N}$ is not to be mistaken with the eponymous collection of $\mathbb{P}$-null sets from Definition A.14, hence why this notation only appears in the appendix. Instead, note the resemblance to condition (i) of the very same definition.

$$= \underbrace{(F \cup N_0)^{\complement}}_{\in \mathcal{F}} \cup \underbrace{\left(F^{\complement} \cap \tilde{N}\right)}_{\in \mathcal{N}} .$$

Therefore, Equation (A.61) holds and we can utilize the representation of $\mathcal{F}$ via (RHS) to uniquely extend $\mathbb{P}$ to a probability measure[50] $\overline{\mathbb{P}}$ on $\overline{\mathcal{F}}$:

$$\overline{\mathbb{P}}(F \cup N) := \mathbb{P}(F) , \quad \text{for } F \in \mathcal{F} , N \in \mathcal{N} .$$

From here, Equation (A.10) follows by choosing $N = \emptyset$. Furthermore, $\left(\Omega, \overline{\mathcal{F}}, \overline{\mathbb{P}}\right)$ is complete by construction, thus finishing the proof of Lemma A.15. □

Proof of Lemma A.16

Lemma A.16 (Right-Continuity of the Internal History of a Counting Process; Protter 2005, p. 16)**.**
Let $N = (N_t)_{t \in [t_0, \infty)}$ be a counting process. Then the internal history $\left\{\mathcal{F}_t^N\right\}_{t \geq t_0}$ of N is a right-continuous filtration.

Proof. Recall the state space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ of N and let Γ denote the measurable space of all mappings from $[t_0, \infty)$ to $\mathbb{R}$ endowed with the product σ-algebra[51],

$$\Gamma = \left(\prod_{t \geq t_0} \mathbb{R} , \bigotimes_{t \geq t_0} \mathcal{B}(\mathbb{R}) \right) .$$

If we consider the mappings

$$P_t : \Omega \longrightarrow \Gamma : \omega \longmapsto \left[s \mapsto N_{\min\{s,t\}}(\omega) \right] ,$$

then $P_t(\omega)$ is the sample path of N at ω constantly continued from t onwards. Since the product σ-algebra is the smallest σ-algebra such that all the coordinate projections

$$\pi_s : \Gamma \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})) , : \gamma \longmapsto \gamma(s) , \quad s \in [t_0, \infty) ,$$

are measurable, we have

$$\bigotimes_{t \geq t_0} \mathcal{B}(\mathbb{R}) = \sigma\left(\{\pi_s : s \geq t_0\}\right) ,$$

and hence for P_t to be measurable it is both necessary and sufficient that $\pi_s \circ P_t$ is measurable for each $s \geq t_0$, see Bauer 2001, p. 35, Theorem 7.4. But obviously we have

$$\pi_s \circ P_t = \begin{cases} N_s , & t_0 \leq s \leq t , \\ N_t , & s > t , \end{cases}$$

and from this it immediately follows that

$$\mathcal{F}_t^N = \sigma(\{N_s : t_0 \leq s \leq t\}) = \sigma(\{P_t\}) . \tag{A.62}$$

[50] This is not shown here, but proves to be an easy exercise.
[51] For a detailed discussion of this σ-algebra commonly used in measure theory, see Bauer 1996, pp. 55–64.

Instead of dealing with an infinite collection of random variables, Equation (A.62) allows us to consider only the single function space-valued random variable P_t when generating the σ-algebra $\mathcal{F}_t^N$. Let

$$A \in \mathcal{F}_{t+}^N \overset{(*)}{=} \bigcap_{n \geq 1} \mathcal{F}_{t+\frac{1}{n}}^N = \bigcap_{n \geq 1} \sigma\left(\left\{P_{t+\frac{1}{n}}\right\}\right) ,$$

where $(*)$ holds because a filtration is an increasing family of σ-algebras. For any $n \in \mathbb{N}$ there exists

$$B_n \in \bigotimes_{t \geq t_0} \mathcal{B}(\mathbb{R})$$

such that $A = P_{t+\frac{1}{n}}^{-1}(B_n) = \left\{P_{t+\frac{1}{n}} \in B_n\right\}$. Furthermore, we set

$$W_n := \left\{P_t = P_{t+\frac{1}{n}}\right\} , \quad n \in \mathbb{N} ,$$

thereby defining an increasing sequence of events. By the right-continuity of N, we have for all $\omega \in \Omega$ that

$$\lim_{n \to \infty} N_{t+\frac{1}{n}}(\omega) = N_t(\omega) ,$$

and since N_t increases only in jumps of integer size, there exists an $n \in \mathbb{N}$ such that $s \mapsto N_s(\omega)$ is constant on $\left[t, t+\frac{1}{n}\right]$. Accordingly, $P_t(\omega) = P_{t+\frac{1}{n}}(\omega)$ and thus $\omega \in W_n$, implying that

$$\Omega = \bigcup_{n \in \mathbb{N}} W_n .$$

The monotonicity of the sequence $(W_n)_{n \in \mathbb{N}}$ ensures $\lim\limits_{n \to \infty} W_n = \Omega$ and we conclude:

$$\begin{aligned} A &= \lim_{n \to \infty} (W_n \cap A) = \lim_{n \to \infty} \left(W_n \cap \left\{P_{t+\frac{1}{n}} \in B_n\right\}\right) \\ &= \lim_{n \to \infty} (W_n \cap \{P_t \in B_n\}) = \lim_{n \to \infty} \{P_t \in B_n\} \in \sigma(\{P_t\}) = \mathcal{F}_t^N , \end{aligned}$$

so that $\mathcal{F}_{t+}^N \subset \mathcal{F}_t^N$. For the reverse inclusion, we once again utilize the properties of a filtration to observe:

$$\mathcal{F}_t^N \subset \mathcal{F}_{t+\frac{1}{n}}^N , \text{ for each } n \in \mathbb{N} . \quad \Longrightarrow \quad \mathcal{F}_t^N \subset \bigcap_{n \geq 1} \mathcal{F}_{t+\frac{1}{n}}^N = \mathcal{F}_{t+}^N .$$

□

Proof of Example A.19

Example A.19 (Compensated Poisson Process; Karatzas and Shreve 1988, p. 12).
Let $N = (N_t)_{t \geq 0}$ be a homogeneous Poisson process with intensity λ, see Example A.7. The *compensated Poisson process* is defined as

$$M_t := N_t - \lambda t , \quad t \geq 0 . \tag{A.11}$$

Then $(M_t)_{t \geq 0}$ is a $\left(\mathbb{P}, \{\mathcal{F}_t^N\}_{t \geq 0}\right)$-martingale.

Proof. We proof the assertion in three parts according to Definition A.17.

(i) In order to see that $(M_t)_{t\geq 0}$ is adapted with respect to the filtration $\left\{\mathcal{F}_t^N\right\}_{t\geq 0}$, we show that for each $t \geq 0$, the generated σ-algebras $\mathcal{F}_t^N$ and $\mathcal{F}_t^M$ coincide. We have

$$\mathcal{F}_t^N = \sigma\left(\{N_s \,:\, 0 \leq s \leq t\}\right) = \sigma\left(\bigcup_{0\leq s\leq t} N_s^{-1}\left(\mathcal{B}(\mathbb{R})\right)\right), \tag{A.63}$$

and therefore only need to verify the following equation:

$$N_s^{-1}\left(\mathcal{B}(\mathbb{R})\right) = M_s^{-1}\left(\mathcal{B}(\mathbb{R})\right), \quad 0 \leq s \leq t\,. \tag{A.64}$$

For any $s \geq 0$ and $A \in N_s^{-1}\left(\mathcal{B}(\mathbb{R})\right)$, there exists $B \in \mathcal{B}(\mathbb{R})$ such that $A = N_s^{-1}(B)$. The shifted set

$$B - \lambda s := \{b - \lambda s \,:\, b \in B\}$$

is again a Borel set satisfying $M_s^{-1}\left(B - \lambda s\right) = A = N_s^{-1}\left(B\right)$, and thus the inclusion $\subset$ holds in Equation (A.64). The inclusion $\supset$ is shown analogously, so that substituting Equation (A.64) into Equation (A.63) yields the assertion.

(ii) For the $\mathbb{P}$-integrability of M_t, application of the triangle inequality and utilizing the non-negativity of N_t leads to the desired result:

$$\mathbb{E}\left(|M_t|\right) \leq \mathbb{E}\left(|N_t| + |\lambda t|\right) = \mathbb{E}\left(N_t\right) + \lambda t = \mathbb{E}\left(N(0,t)\right) + \lambda t = 2\lambda t < \infty\,,$$

as $N(0,t) = N_t - N_0$ follows a Poisson distribution with parameter λt.

(iii) For each $t \geq s \geq 0$, note that the independence of increments ensures that $N_t - N_s$ is independent of $\mathcal{F}_s^N$. Therefore, we observe:

$$\mathbb{E}\left(N_t - N_s \,|\, \mathcal{F}_s^N\right) = \mathbb{E}\left(N_t - N_s\right) = \lambda(t-s)\,. \tag{A.65}$$

Furthermore, N_s is $\mathcal{F}_s^N$-measurable and we obtain:

$$\mathbb{E}\left(N_t - N_s \,|\, \mathcal{F}_s^N\right) = \mathbb{E}\left(N_t \,|\, \mathcal{F}_s^N\right) - \mathbb{E}\left(N_s \,|\, \mathcal{F}_s^N\right) = \mathbb{E}\left(N_t \,|\, \mathcal{F}_s^N\right) - N_s \tag{A.66}$$

By combining Equations (A.65) and (A.66) we conclude:

$$\begin{aligned}\mathbb{E}\left(M_t \,|\, \mathcal{F}_s^N\right) &= \mathbb{E}\left(N_t \,|\, \mathcal{F}_s^N\right) - \lambda t = \underbrace{\mathbb{E}\left(N_t \,|\, \mathcal{F}_s^N\right) - N_s}_{=\lambda(t-s)} + N_s - \lambda t \\ &= N_s + \lambda(t-s) - \lambda t = N_s - \lambda s = M_s\,.\end{aligned}$$

By properties (i)-(iii), the compensated Poisson process is a $\left(\mathbb{P}, \{\mathcal{F}_t^N\}_{t\geq 0}\right)$-martingale. $\square$

Proof of Equation (A.16)

If we assume λ to be bounded by an integrable random variable, we can compute for $t \geq t_0$ using the averaging and the dominated convergence theorem (cf. Aalen 1978, p. 705):

$$\lim_{h\downarrow 0} \frac{1}{h}\mathbb{E}\left(N_{t+h} - N_t \,|\, \mathcal{F}_t\right) = \lim_{h\downarrow 0} \frac{1}{h}\mathbb{E}\left(\Lambda_{t+h} - \Lambda_t \,|\, \mathcal{F}_t\right)$$

$$
\begin{aligned}
&= \lim_{h \downarrow 0} \mathbb{E}\left(\frac{1}{h}\int_t^{t+h} \lambda_u \,\mathrm{d}u \,\middle|\, \mathcal{F}_t\right) \\
&= \mathbb{E}\left(\lim_{h \downarrow 0} \frac{1}{h}\int_t^{t+h} \lambda_u \,\mathrm{d}u \,\middle|\, \mathcal{F}_t\right) \\
&= \mathbb{E}\left(\lambda_{t+} \middle| \mathcal{F}_t\right) = \lambda_{t+}\,, \qquad \text{(A.16)}
\end{aligned}
$$

where $\lambda_{t+} = \lim_{u \downarrow t} \lambda_u$, and the last equation holds since λ_{t+} is $\mathcal{F}_t$-measurable.

Proof. Only the last step of Equation (A.16) demands further explanation, so it suffices to show that λ_{t+} is $\mathcal{F}_t$-measurable. As before, the Lebesgue averaging theorem yields:

$$\lambda_{t+} = \lim_{h \to 0} \frac{1}{h}\int_t^{t+h} \lambda_u \,\mathrm{d}u = \lim_{h \to 0} \frac{\Lambda_{t+h} - \Lambda_t}{h}\,. \tag{A.67}$$

Since Λ is $\{\mathcal{F}_t\}_{t \geq t_0}$-predictable and hence adapted by virtue of Lemma A.22, Λ_t is $\mathcal{F}_u$-measurable for each $u \geq t$. In particular, for any $0 < h \leq h_0$, the difference quotients

$$\frac{\Lambda_{t+h} - \Lambda_t}{h}$$

from Equation (A.67) are $\mathcal{F}_{t+h_0}$-measurable and thus the same holds for the limit λ_{t+}. Given that h_0 can be chosen arbitrarily small, it follows that

$$\sigma(\,\{\lambda_{t+}\}\,) \subset \bigcap_{h_0 > 0} \mathcal{F}_{t+h_0} = \bigcap_{s > t} \mathcal{F}_s = \mathcal{F}_{t+} = \mathcal{F}_t$$

holds by the right-continuity of the filtration $\{\mathcal{F}_t\}_{t \geq t_0}$, which completes the proof. $\square$

Proof of Lemma A.30

Lemma A.30 (One-Point Process: Compensator w.r.t. the Internal Filtration; Lemma 14.1.II. of Daley and Vere-Jones 2008, p. 359)**.**
Let X be a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in $([t_0, \infty], \mathcal{B}([t_0, \infty]))$. Let F denote the cumulative distribution function of X and define the one-point process N by

$$N \,:\, [t_0, \infty) \times \Omega \longrightarrow \mathbb{R} \,:\, (t, \omega) \longmapsto N(t, \omega) = \mathbb{1}_{(-\infty, t]}\left(X(\omega)\right) = \begin{cases} 1\,, & t \geq X(\omega)\,, \\ 0\,, & t < X(\omega)\,. \end{cases}$$

The one-point process N has the $\left(\mathbb{P}, \left\{\mathcal{F}_t^N\right\}_{t \geq t_0}\right)$-compensator Λ given by

$$\Lambda \,:\, [t_0, \infty) \times \Omega \longrightarrow \mathbb{R} \,:\, (t, \omega) \longmapsto \Lambda(t, \omega) = H(t \wedge X(\omega)) = \begin{cases} H(X(\omega))\,, & t \geq X(\omega)\,, \\ H(t)\,, & t < X(\omega)\,, \end{cases}$$

where $\left\{\mathcal{F}_t^N\right\}_{t \geq t_0}$ denotes the internal filtration of N and H is defined via

$$H(t) := \int_{t_0}^{t} \frac{\mathrm{d}F(x)}{1 - F(x-)}\,. \tag{A.20}$$

Proof. The proof proceeds along the lines of Daley and Vere-Jones 2008, pp. 359–360, but is adapted to our situation. We first find that N and Λ are integrable. This is obviously true for the former, while for the latter we have:

$$\mathbb{E}\left(|\Lambda_t|\right) = \mathbb{E}\left(H(t \wedge X)\right) \leq H(t) < \infty\,, \quad \text{for all } t \geq t_0\,.$$

Since H and thus Λ is right-continuous and increasing in t[52] with $\Lambda(t_0, \omega) = 0$, it suffices to show that (a) Λ is $\{\mathcal{F}_t\}_{t \geq t_0}$-predictable and (b) $N - \Lambda$ is a $\left(\mathbb{P}, \{\mathcal{F}_t\}_{t \geq t_0}\right)$-martingale.

(a) To verify that Λ is predictable, we first check that $\xi(t, \omega) = t \wedge X(\omega)$ is predictable. We study the sets $\{(t, \omega) \,:\, \xi(t, \omega) > s\}$ generating the σ-algebra $\sigma(\xi)$ and obtain:

$$\begin{aligned} \{(t, \omega) \,:\, \xi(t, \omega) > s\} &= \{(t, \omega) \,:\, t > s \,\wedge\, X(\omega) > s\} \\ &= \{t \,:\, t > s\} \times \underbrace{\{\omega \,:\, X(\omega) > s\}}_{\in \mathcal{F}_s^N} \\ &= (s, \infty) \times A\,, \end{aligned} \tag{A.68}$$

for some $A \in \mathcal{F}_s^N$. This can be seen as follows:

$$\begin{aligned} \{\omega \,:\, X(\omega) > s\}^{\complement} = \{\omega \,:\, X(\omega) \leq s\} &= \{\omega \,:\, N_s(\omega) \geq 1\} \\ &= N_s^{-1}\left([1, \infty)\right) \in \mathcal{F}_s^N\,, \end{aligned}$$

whereby the identity from Equation (A.4) is mirrored. If we examine the predictable σ-algebra $\mathcal{P}\left(\{\mathcal{F}_t\}_{t \geq t_0}\right)$ from Definition A.20, we recognize that the set in Equation (A.68) has the form of a generating set for this σ-algebra. Therefore, $\sigma(\xi) \subset \mathcal{P}\left(\{\mathcal{F}_t\}_{t \geq t_0}\right)$ and thus ξ is predictable. Since H is increasing and right-continuous, we have with an argument similar to the proof of the simulation lemma[53]:

$$\{(t, \omega) \,:\, H\left(\xi(t, \omega)\right) \geq s\} = \{(t, \omega) \,:\, \xi(t, \omega) \geq H^{\leftarrow}(s)\}\,, \tag{A.69}$$

where $H^{\leftarrow}(s) := \inf\{t \,:\, H(t) \geq s\}$ denotes the generalized inverse of H. Recall that, in general, H is not left-continuous and thus has no inverse, since the distribution of X may contain atoms causing jumps of the associated cumulative distribution function F. The predictability of ξ now ensures that the sets in Equation (A.69) are again included in the predictable σ-algebra, and hence the predictability of $\Lambda = H \circ \xi$ follows with the same arguments as above.

(b) In proving that $N - \Lambda$ is a $\left(\mathbb{P}, \{\mathcal{F}_t\}_{t \geq t_0}\right)$-martingale, only condition (iii) of Definition A.17 remains to be shown. We first note that $\{\omega \,:\, X(\omega) > t\}$ constitutes an atom of $\mathcal{F}_t^N$. In order to see this, we bear in mind that

$$\sigma\left(N_t\right) = \{\emptyset, \underbrace{\{N_t = 0\}}_{=\{X > t\}}, \underbrace{\{N_t = 1\}}_{=\{X \leq t\}}, \Omega\}\,,$$

[52] The right-continuity is inherited from the cumulative distribution function F, whereas the monotonicity is trivial. For further details on this regard, see Appendix A4 of Brémaud 1981, pp. 334–339 and Proposition 4.6.V. of Daley and Vere-Jones 2003, p. 109.

[53] This result is also widely known as inverse transform sampling, the basics of which can be found among others in Graham and Talay 2013, p. 22.

and thereby $\mathcal{F}_t^N$ is generated by sets of the form $\{\omega : X(\omega) > s\}$ for $t_0 \leq s \leq t$. Accordingly, $\{\omega : X(\omega) > t\}$ can not be further decomposed within $\mathcal{F}_t^N$. We take advantage of this to explicitly calculate the conditional expectations involved in verifying the martingale property. For any non-negative measurable function g, we obtain[54]:

$$\mathbb{E}\left(g(X) \,|\, \mathcal{F}_t^N\right) = \mathbb{E}(g(X) \,|\, X > t) \quad \text{on } \{\omega : X(\omega) > t\} \,,$$

and consequently in this case the conditional expectation is given by (see Bauer 1996, p. 110 for reference)

$$\mathbb{E}\left(g(X) \,|\, \mathcal{F}_t^N\right) = \frac{\int_{\{X>t\}} g(X)\,\mathrm{d}\mathbb{P}}{\mathbb{P}(X > t)} = \frac{\mathbb{E}\left(g(X)\mathbb{1}_{\{X>t\}}\right)}{\mathbb{E}\left(\mathbb{1}_{\{X>t\}}\right)} \quad \text{on } \{\omega : X(\omega) > t\} \,,$$

which in terms of the cumulative distribution function F of X can be written as

$$\mathbb{E}\left(g(X) \,|\, \mathcal{F}_t^N\right) = \frac{1}{1 - F(t)} \int_t^\infty g(x)\,\mathrm{d}F(x) \quad \text{on } \{\omega : X(\omega) > t\} \,. \tag{A.70}$$

Choosing $g(x) = \mathbb{1}_{(-\infty,t]}(x)$ in Equation (A.70) then yields for $t_0 \leq s \leq t$:

$$\begin{aligned}
\mathbb{E}\left(N_t \,|\, \mathcal{F}_s^N\right) = \mathbb{E}\left(g(X) \,|\, \mathcal{F}_s^N\right) &= \frac{1}{1 - F(s)} \int_s^\infty g(x)\,\mathrm{d}F(x) \\
&= \frac{1}{1 - F(s)} \int_s^\infty \mathbb{1}_{(-\infty,t]}(x)\,\mathrm{d}F(x) \\
&= \frac{1}{1 - F(s)} \int_s^t \mathrm{d}F(x) \\
&= \frac{F(t) - F(s)}{1 - F(s)} \quad \text{on } \{\omega : X(\omega) > s\} \,.
\end{aligned} \tag{A.71}$$

If instead we choose $g(x) = H(t \wedge x)$, once again applying Equation (A.70) provides

$$\begin{aligned}
\mathbb{E}\left(\Lambda_t \,|\, \mathcal{F}_s^N\right) &= \frac{1}{1 - F(s)} \int_s^\infty H(t \wedge x)\,\mathrm{d}F(x) \\
&= \frac{1}{1 - F(s)} \left(\int_s^t H(x)\,\mathrm{d}F(x) + \int_t^\infty H(t)\,\mathrm{d}F(x)\right) \\
&= \frac{1}{1 - F(s)} \left(\int_s^t H(x)\,\mathrm{d}F(x) + H(t)\,(1 - F(t))\right) \text{ on } \{\omega : X(\omega) > s\} \,,
\end{aligned}$$

and hence we can compute on $\{\omega : X(\omega) > s\}$:

$$\begin{aligned}
&[1 - F(s)] \left[\mathbb{E}\left(\Lambda_t \,|\, \mathcal{F}_s^N\right) - H(s)\right] \\
&\quad = \int_s^t H(x)\,\mathrm{d}F(x) + H(t)\,(1 - F(t)) - (1 - F(s))H(s) \\
&\quad = \int_s^t (H(x) - H(s))\,\mathrm{d}F(x) + H(t)\,(1 - F(t)) - H(s) + H(s)F(t) \,,
\end{aligned} \tag{A.72}$$

[54] Calculating the conditional expectation on selected subsets of $\mathcal{F}_t^N$ may seem perplexing at first, but keep in mind that the conditional expectation is itself a random variable defined on that same σ-algebra.

where we used that

$$\int_s^t -H(s)\,\mathrm{d}F(x) = -H(s)\,(F(t)-F(s)) = H(s)F(s) - H(s)F(t)\,.$$

Proceeding from Equation (A.72), we can apply the Product Formula of the Stieltjes-Lebesgue Calculus (Theorem A.49 of Appendix A.5) in $(*)$ to receive:

$$\begin{aligned}
&\int_s^t (H(x)-H(s))\,\mathrm{d}F(x) + H(t)\,(1-F(t)) - H(s) + H(s)F(t) \\
&\quad= \int_s^t (\underbrace{H(x)-H(s)}_{\hat{=} f(x) \text{ in Thm. A.49}})\,\mathrm{d}F(x) + (H(t)-H(s))\,(1-F(t)) \qquad (*) \\
&\quad= (H(t)-H(s))\,F(t) - \int_s^t F(x-)\,\mathrm{d}(H(x)-\cancel{H(s)}) + (H(t)-H(s))\,(1-F(t)) \\
&\quad= \underbrace{H(t)-H(s)}_{=\int_s^t 1\,\mathrm{d}H(x)} - \int_s^t F(x-)\,\mathrm{d}H(x) = \int_s^t (1-F(x-))\,\mathrm{d}H(x)\,, \qquad \text{(A.73)}
\end{aligned}$$

since the addition of constant terms to the integrator does not change the Lebesgue-Stieltjes integral. From here, we observe that the integration in Equation (A.73) is in terms of the hazard measure, so by substituting Equation (A.22) we have:

$$\begin{aligned}
\int_s^t (1-F(x-))\,\mathrm{d}H(x) &= \int_s^t (1-F(x-))\,\frac{\mathrm{d}F(x)}{1-F(x-)} \\
&= \int_s^t \mathrm{d}F(x) = F(t)-F(s)\,. \qquad \text{(A.74)}
\end{aligned}$$

Thus, merging Equations (A.72), (A.73) and (A.74) yields:

$$\mathbb{E}\left(\Lambda_t \,\middle|\, \mathcal{F}_s^N\right) - H(s) = \frac{F(t)-F(s)}{1-F(s)} = \mathbb{E}\left(N_t \,\middle|\, \mathcal{F}_s^N\right) \quad \text{on } \{\omega : X(\omega) > s\}\,. \qquad \text{(A.75)}$$

Considering that $N_s = 0$ and $\Lambda_s = H(s \wedge X) = H(s)$ on $\{\omega : X(\omega) > s\}$, Equation (A.75) can be reformulated as

$$\mathbb{E}\left(N_t - \Lambda_t \,\middle|\, \mathcal{F}_s^N\right) = -H(s) = N_s - \Lambda_s \quad \text{on } \{\omega : X(\omega) > s\}\,,$$

which is the desired martingale property. On the other hand, on $\{\omega : X(\omega) > s\}^{\complement} = \{\omega : X(\omega) \leq s\}$ we have $N_t = N_s = 1$ and $\Lambda_t = \Lambda_s = H(X)$, where X and thus $H(X)$ is $\mathcal{F}_s^N$-measurable due to $\{\omega : X(\omega) \leq s\} = \{\omega : N(s,\omega) = 1\} \in \mathcal{F}_s^N$. Therefore,

$$\mathbb{E}\left(N_t - \Lambda_t \,\middle|\, \mathcal{F}_s^N\right) = 1 - H(X) = N_s - \Lambda_s \quad \text{on } \{\omega : X(\omega) > s\}^{\complement}\,,$$

and hence the martingale property applies on $\Omega = \{\omega : X(\omega) > s\} \cup \{\omega : X(\omega) > s\}^{\complement}$.

□

Proof of Lemma A.32

Lemma A.32 (One-Point Process: Compensator w.r.t. an Intrinsic Filtration; Lemma 14.1.III. of Daley and Vere-Jones 2008, p. 361)**.**
In the situation of Lemma A.30, let $\mathcal{G}_0$ denote the prior σ-algebra of an intrinsic filtration $\mathcal{F}_t = \mathcal{F}_t^N \vee \mathcal{G}_0$. If a regular conditional distribution function $F\left(\cdot \,|\, \mathcal{G}_0\right)$ for X exists, the one-point process N has the $\left(\mathbb{P}, \{\mathcal{F}_t\}_{t \geq t_0}\right)$-compensator Λ given by

$$
\begin{aligned}
\Lambda \,:\, & [t_0, \infty) \times \Omega \longrightarrow \mathbb{R} \\
& (t, \omega) \longmapsto \Lambda(t, \omega) = H(t \wedge X(\omega) \,|\, \mathcal{G}_0) = \begin{cases} H(X(\omega) \,|\, \mathcal{G}_0)\,, & t \geq X(\omega)\,, \\ H(t \,|\, \mathcal{G}_0)\,, & t < X(\omega)\,, \end{cases}
\end{aligned}
$$

where $H\left(\cdot \,|\, \mathcal{G}_0\right)$ is the conditional integrated hazard function associated with $F\left(\cdot \,|\, \mathcal{G}_0\right)$,

$$
H\left(t \,|\, \mathcal{G}_0\right) := \int_{t_0}^{t} \frac{\mathrm{d}F\left(x \,|\, \mathcal{G}_0\right)}{1 - F\left(x- \,|\, \mathcal{G}_0\right)}\,.
$$

Proof. The proof found here is a corrected version of the erroneous one in Daley and Vere-Jones 2008, pp. 361-362, which nevertheless provides the key arguments needed to extend the result from the previous lemma. It is hence performed in analogy to that of Lemma A.30; in particular, the predictability of Λ can be shown with the exact same arguments as before. For the required martingale property, recall that we had for any non-negative measurable function g:

$$
\mathbb{E}\left(g(X) \,|\, \mathcal{F}_t^N\right) = \begin{cases} g(X)\,, & \text{on } \{\omega \,:\, X(\omega) \leq t\}\,, \\ \frac{\mathbb{E}\left(g(X)\mathbb{1}_{\{X>t\}}\right)}{\mathbb{E}\left(\mathbb{1}_{\{X>t\}}\right)}\,, & \text{on } \{\omega \,:\, X(\omega) > t\}\,. \end{cases}
$$

We claim that upon considering $\mathcal{F}_t = \mathcal{F}_t^N \vee \mathcal{G}_0$, the following holds:

$$
\mathbb{E}\left(g(X) \,|\, \mathcal{F}_t\right) = \begin{cases} g(X)\,, & \text{on } \{\omega \,:\, X(\omega) \leq t\}\,, \\ \frac{\mathbb{E}\left(g(X)\mathbb{1}_{\{X>t\}} \,|\, \mathcal{G}_0\right)}{\mathbb{E}\left(\mathbb{1}_{\{X>t\}} \,|\, \mathcal{G}_0\right)}\,, & \text{on } \{\omega \,:\, X(\omega) > t\}\,. \end{cases} \tag{A.76}
$$

Since $\{\omega \,:\, X(\omega) \leq t\} = \{\omega \,:\, N(t, \omega) = 1\} \in \mathcal{F}_t^N \subset \mathcal{F}_t$, the first part of Equation (A.76) follows again by measurability. For the second part, we utilize once more that $\{\omega \,:\, X(\omega) > t\}$ can not be further decomposed within $\mathcal{F}_t^N$ and hence the restriction[55]

$$
\mathcal{F}_t\big|_{\{X>t\}} = \{A \cap \{X > t\} \,:\, A \in \mathcal{F}_t\} \subset \mathcal{F}_t
$$

of $\mathcal{F}_t$ to $\{\omega \,:\, X(\omega) > t\}$ consists entirely of sets of the form $U \cap \{\omega \,:\, X(\omega) > t\}$ for some $U \in \mathcal{G}_0$. For each such U, we can write by the definition of the conditional expectation:

$$
\int_{U \cap \{X>t\}} g(X)\,\mathrm{d}\mathbb{P} = \int_U g(X)\mathbb{1}_{\{X>t\}}\,\mathrm{d}\mathbb{P} = \int_U \mathbb{E}\left(g(X)\mathbb{1}_{\{X>t\}} \,|\, \mathcal{G}_0\right)\,\mathrm{d}\mathbb{P}\,. \tag{A.77}
$$

Furthermore, if we exploit the $\mathcal{G}_0$-measurability of the conditional expectation with respect

[55] In the proof of Lemma A.30, this restriction would contain only the atom $\{\omega \,:\, X(\omega) > t\}$.

to $\mathcal{G}_0$, we can find that

$$\begin{aligned}
&\mathbb{E}\left(\frac{\mathbb{1}_{\{X>t\}}}{\mathbb{E}\left(\mathbb{1}_{\{X>t\}}\,\middle|\,\mathcal{G}_0\right)}\cdot\mathbb{E}\left(g(X)\mathbb{1}_{\{X>t\}}\,\middle|\,\mathcal{G}_0\right)\,\middle|\,\mathcal{G}_0\right)\\
&\quad=\frac{\mathbb{E}\left(g(X)\mathbb{1}_{\{X>t\}}\,\middle|\,\mathcal{G}_0\right)}{\mathbb{E}\left(\mathbb{1}_{\{X>t\}}\,\middle|\,\mathcal{G}_0\right)}\cdot\mathbb{E}\left(\mathbb{1}_{\{X>t\}}\,\middle|\,\mathcal{G}_0\right)\\
&\quad=\mathbb{E}\left(g(X)\mathbb{1}_{\{X>t\}}\,\middle|\,\mathcal{G}_0\right)\,, \qquad\qquad (\text{A.78})
\end{aligned}$$

so since $U\in\mathcal{G}_0$, substituting Equation (A.78) into Equation (A.77) yields:

$$\begin{aligned}
\int_U \mathbb{E}\left(g(X)\mathbb{1}_{\{X>t\}}\,\middle|\,\mathcal{G}_0\right)\,\mathrm{d}\mathbb{P}&=\int_U\frac{\mathbb{1}_{\{X>t\}}}{\mathbb{E}\left(\mathbb{1}_{\{X>t\}}\,\middle|\,\mathcal{G}_0\right)}\cdot\mathbb{E}\left(g(X)\mathbb{1}_{\{X>t\}}\,\middle|\,\mathcal{G}_0\right)\,\mathrm{d}\mathbb{P}\\
&=\int_{U\cap\{X>t\}}\frac{\mathbb{E}\left(g(X)\mathbb{1}_{\{X>t\}}\,\middle|\,\mathcal{G}_0\right)}{\mathbb{E}\left(\mathbb{1}_{\{X>t\}}\,\middle|\,\mathcal{G}_0\right)}\,\mathrm{d}\mathbb{P}\,. \qquad (\text{A.79})
\end{aligned}$$

As we have $\mathcal{G}_0\subset\mathcal{F}_t$, the integrand of Equation (A.79) is $\mathcal{F}_t$-measurable and the comparison of Equations (A.77) and (A.79) shows that

$$\mathbb{E}\left(g(X)\,|\,\mathcal{F}_t\right)=\frac{\mathbb{E}\left(g(X)\mathbb{1}_{\{X>t\}}\,\middle|\,\mathcal{G}_0\right)}{\mathbb{E}\left(\mathbb{1}_{\{X>t\}}\,\middle|\,\mathcal{G}_0\right)}\,,\quad \text{on }\{\omega\,:\,X(\omega)>t\}\,,$$

and thereby the second part of Equation (A.76) as asserted. Assuming that X has a regular conditional distribution function $F\left(\cdot\,|\,\mathcal{G}_0\right)$, this expression can be further reduced to obtain:

$$\mathbb{E}\left(g(X)\,|\,\mathcal{F}_t\right)=\frac{1}{1-F\left(t\,|\,\mathcal{G}_0\right)}\int_t^\infty g(x)\,\mathrm{d}F\left(x\,|\,\mathcal{G}_0\right)\,, \qquad (\text{A.80})$$

representing a conditional version of Equation (A.70). From here, the martingale property can be established as in the proof of Lemma A.30, using $F\left(\cdot\,|\,\mathcal{G}_0\right)$ in place of the unconditional cumulative distribution function F. $\square$

Proof of Theorem A.33

Theorem A.33 (Jacod's Formula for the Intensity Process; Theorem 14.1.IV. of Daley and Vere-Jones 2008, pp. 363–364)**.**
Let $N=(N_t)_{t\geq t_0}$ be a counting process and $T=(T_i)_{i\in\mathbb{N}}$ the associated simple point process. Let $\mathcal{F}_t=\mathcal{F}_t^N\vee\mathcal{G}_0$ denote an intrinsic filtration with prior σ-algebra $\mathcal{G}_0$. Suppose there exist regular versions $F_i\left(\cdot\,|\,\mathcal{F}_{T_{i-1}}\right)$ of the conditional distribution functions of the interarrival times $W_i=T_i-T_{i-1}$, given $\mathcal{F}_{T_{i-1}}$ as in Equation (A.23)*, such that $1-F_i\left(x-\,|\,\mathcal{F}_{T_{i-1}}\right)>0$ for $x>0$. Let N_i denote the one-point process given by*

$$N_i\,:\,[t_0,\infty)\times\Omega\longrightarrow\mathbb{R}\,:\,(t,\omega)\longmapsto N_i(t,\omega)=\mathbb{1}_{(-\infty,t]}\left(T_i(\omega)\right)\,,$$

so that $N = \sum_{i=1}^{\infty} N_i$. *Then the* $\left(\mathbb{P}, \{\mathcal{F}_t\}_{t \geq t_0}\right)$*-compensator* Λ_i *for* N_i *has the form*

$$\Lambda_i(t,\omega) = \begin{cases} 0\,, & t < T_{i-1}(\omega)\,, \\ H_i\left(t - T_{i-1} \,|\, \mathcal{F}_{T_{i-1}}\right)\,, & T_{i-1}(\omega) \leq t < T_i(\omega)\,, \\ H_i\left(T_i - T_{i-1} \,|\, \mathcal{F}_{T_{i-1}}\right)\,, & T_i(\omega) \leq t\,, \end{cases} \tag{A.24}$$

where $H_i\left(\cdot \,|\, \mathcal{F}_{T_{i-1}}\right)$ *is the conditional integrated hazard function associated with* $F_i\left(\cdot \,|\, \mathcal{F}_{T_{i-1}}\right)$*,*

$$H_i\left(t \,|\, \mathcal{F}_{T_{i-1}}\right) = \int_0^t \frac{\mathrm{d}F_i\left(x \,|\, \mathcal{F}_{T_{i-1}}\right)}{1 - F_i\left(x- \,|\, \mathcal{F}_{T_{i-1}}\right)}\,. \tag{A.25}$$

Thus, a version of the $\left(\mathbb{P}, \{\mathcal{F}_t\}_{t \geq t_0}\right)$*-compensator* Λ *for* N *is given by*

$$\Lambda(t,\omega) = \sum_{i=1}^{\infty} \Lambda_i(t,\omega)\,. \tag{A.26}$$

Proof. A sketch of the proof is given in Daley and Vere-Jones 2008, pp. 362–364. Since some of the arguments used in it fall somewhat short, we give a more detailed proof based in part on the results presented in Karr 1991[56]:
To establish the form of the compensator for N_i given in Equation (A.24), note that for the predictability, the previous methods can be applied in conjunction with Proposition 2.6. of Karr 1991, p. 57 (see Remark A.51 of Appendix A.5). While that result itself is fundamental, it is in turn based on the theory of stopping times, and we are therefore content to refer to the intuitive proof of Lemma A.30 rather than go into further detail. Specifically, if the conditional distribution functions each have absolutely continuous versions, the predictability of the intensity process (and thus the associated compensator) follows directly from Theorem A2, T24 of Brémaud 1981, p. 304, which provides a generalization of the above proposition.
The assumption that $1 - F_i\left(x- \,|\, \mathcal{F}_{T_{i-1}}\right) > 0$ holds for $x > 0$ ensures the integrability of the conditional integrated hazard function $H_i\left(\cdot \,|\, \mathcal{F}_{T_{i-1}}\right)$ and hence of Λ_i. As before, it is sufficient to prove the requisite equality

$$\mathbb{E}\left(N_i(t) - \Lambda_i(t) \,|\, \mathcal{F}_s\right) = N_i(s) - \Lambda_i(s) \tag{A.81}$$

for $t_0 < s \leq t$, where the dependence on ω is neglected in favor of a shorter notation. Note that both the series $\sum_{i=1}^{\infty} N_i(t)$ and $\sum_{i=1}^{\infty} \Lambda_i(t)$ are almost surely absolutely convergent:

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^{\infty} |N_i(t)| = \infty\right) &= \mathbb{P}\left(\sum_{i=1}^{\infty} \mathbb{1}_{(-\infty,t]}\left(T_i\right) = \infty\right) \\ &= \mathbb{P}(\#\{i \in \mathbb{N} : T_i \leq t\} = \infty) \\ &= \mathbb{P}\left(\liminf_{i \to \infty} T_i \leq t\right) = 0\,, \end{aligned}$$

by the Bolzano-Weierstraß theorem and since the explosion of counting processes is prohibited by condition (iii) of Definition A.3, see Remark A.6(iii). Since $\Lambda_i(t,\omega) = 0$ for

[56] In fact, a version of Jacod's formula can be found in Theorem 2.18 of Karr 1991, p. 62, where the compensator is derived using dual predictable projections.

$t < T_{i-1}(\omega)$, $\sum_{i=1}^{\infty} \Lambda_i(t,\omega) = \infty$ implies that $t \geq T_i(\omega)$ for infinitely many $i \in \mathbb{N}$, and we obtain in a similar vein:

$$\mathbb{P}\left(\sum_{i=1}^{\infty} |\Lambda_i(t)| = \infty\right) \leq \mathbb{P}(\#\{i \in \mathbb{N} \,:\, T_i \leq t\} = \infty) = 0\,,$$

and hence the desired result. This serves as a justification that if Equation (A.81) applies for each $i \in \mathbb{N}$, we have by additivity and dominated convergence of the conditional expectation:

$$\begin{aligned}
\mathbb{E}\left(N(t) - \Lambda(t) \,|\, \mathcal{F}_s\right) &= \mathbb{E}\left(\sum_{i=1}^{\infty} N_i(t) - \sum_{i=1}^{\infty} \Lambda_i(t) \,\middle|\, \mathcal{F}_s\right) \\
&= \mathbb{E}\left(\sum_{i=1}^{\infty} (N_i(t) - \Lambda_i(t)) \,\middle|\, \mathcal{F}_s\right) \\
&= \sum_{i=1}^{\infty} \mathbb{E}\left(N_i(t) - \Lambda_i(t) \,|\, \mathcal{F}_s\right) \\
&= \sum_{i=1}^{\infty} (N_i(s) - \Lambda_i(s)) = N(s) - \Lambda(s)\,,
\end{aligned}$$

so that Equation (A.26) holds. In order to prove Equation (A.81), we establish the equality separately on the sets $\{\omega : T_{i-1}(\omega) \leq s\}$ and $\{\omega : T_{i-1}(\omega) \leq s\}^{\complement} = \{\omega : T_{i-1}(\omega) > s\}$, which are further decomposed according to Figure 20. We will first deal with the decomposition (2a) illustrated in this figure:

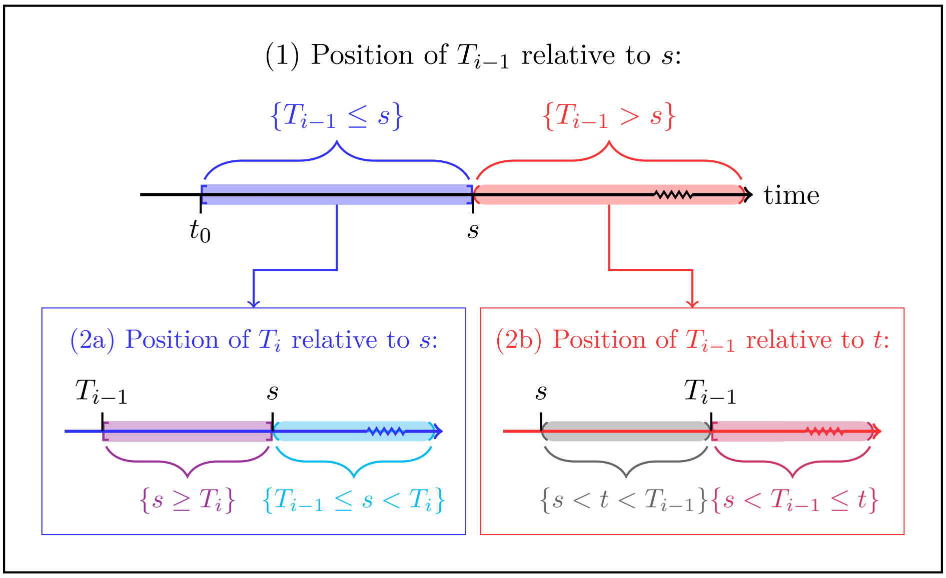


Figure 20: Decomposition of the sample space Ω required for the proof of Jacod's formula. The decomposition is performed in two steps, which are illustrated here. In total, Ω needs to be divided into four disjoint subsets.

We observe that the set $\{\omega : T_{i-1}(\omega) \leq s\}$ is contained in both $\mathcal{F}_{T_{i-1}}$ (by construction) and $\mathcal{F}_s$ (since $\{\omega : T_{i-1}(\omega) \leq s\} = \{\omega : N(s,\omega) \geq i-1\}$ by Equation (A.4)). The relationship between these two σ-algebras is further elucidated by the identity

$$\mathcal{F}_s \cap \{\omega \,:\, T_{i-1}(\omega) \leq s < T_i(\omega)\} = \mathcal{F}_{T_{i-1}} \cap \{\omega \,:\, T_{i-1}(\omega) \leq s < T_i(\omega)\} \;, \tag{A.82}$$

which means that, given any $U \in \mathcal{F}_s$, there exists $U' \in \mathcal{F}_{T_{i-1}}$ such that

$$U \cap \{\omega \,:\, T_{i-1}(\omega) \leq s < T_i(\omega)\} = U' \cap \{\omega \,:\, T_{i-1}(\omega) \leq s < T_i(\omega)\} \;,$$

and conversely. For the validity of Equation (A.82), one considers the basic sets generating $\mathcal{F}_t^N$ (i.e., $\{\omega \,:\, N(s,\omega) = j\}$ for $t_0 < s \leq t$), and computes (cf. Karr 1991, p. 57):

$$\begin{aligned}
&\{\omega \,:\, N(s,\omega) = j\} \cap \{\omega \,:\, T_{i-1}(\omega) \leq t < T_i(\omega)\} \\
&\quad = \underbrace{\{\omega \,:\, N\left(s \wedge T_{i-1}(\omega), \omega\right) = j\}}_{\in \mathcal{F}_{T_{i-1}}} \cap \{\omega \,:\, T_{i-1}(\omega) \leq t < T_i(\omega)\} \;,
\end{aligned}$$

since even for $s > T_{i-1}(\omega)$ no further jump of N can occur due to $s \leq t < T_i(\omega)$. Equation (A.82) can then be obtained by application of the monotone class theorem, see Brémaud 2020, pp. 58–59 for reference[57]. The relevance of this identity is that on the set above, the waiting time $W_i = T_i - T_{i-1}$ plays the same role for N_i as X does for the one-point process from Lemma A.32, with $\mathcal{F}_{T_{i-1}}$ here playing the role of $\mathcal{G}_0$ there. Accordingly, a result similar to Equation (A.76) holds by retracing Equations (A.77), (A.78) and (A.79), where for $U \in \mathcal{F}_s$ and any non-negative measurable function g we have:

$$\begin{aligned}
\int_{U \cap \{T_{i-1} \leq s < T_i\}} g\left(W_i\right) \,\mathrm{d}\mathbb{P} &= \int_{U' \cap \{T_{i-1} \leq s < T_i\}} g\left(W_i\right) \,\mathrm{d}\mathbb{P} \\
&= \int_{U'} g\left(W_i\right) \mathbb{1}_{\{T_{i-1} \leq s < T_i\}} \,\mathrm{d}\mathbb{P} \\
&= \int_{U'} \mathbb{E}\left(g\left(W_i\right) \mathbb{1}_{\{T_{i-1} \leq s < T_i\}} \,\middle|\, \mathcal{F}_{T_{i-1}}\right) \mathrm{d}\mathbb{P} \\
&= \int_{U'} \frac{\mathbb{1}_{\{T_{i-1} \leq s < T_i\}}}{\mathbb{E}\left(\mathbb{1}_{\{T_{i-1} \leq s < T_i\}} \,\middle|\, \mathcal{F}_{T_{i-1}}\right)} \mathbb{E}\left(g\left(W_i\right) \mathbb{1}_{\{T_{i-1} \leq s < T_i\}} \,\middle|\, \mathcal{F}_{T_{i-1}}\right) \mathrm{d}\mathbb{P} \\
&= \int_{U' \cap \{T_{i-1} \leq s < T_i\}} \frac{\mathbb{E}\left(g\left(W_i\right) \mathbb{1}_{\{T_{i-1} \leq s < T_i\}} \,\middle|\, \mathcal{F}_{T_{i-1}}\right)}{\mathbb{E}\left(\mathbb{1}_{\{T_{i-1} \leq s < T_i\}} \,\middle|\, \mathcal{F}_{T_{i-1}}\right)} \,\mathrm{d}\mathbb{P} \\
&= \int_{U \cap \{T_{i-1} \leq s < T_i\}} \frac{\mathbb{E}\left(g\left(W_i\right) \mathbb{1}_{\{T_{i-1} \leq s < T_i\}} \,\middle|\, \mathcal{F}_{T_{i-1}}\right)}{\mathbb{E}\left(\mathbb{1}_{\{T_{i-1} \leq s < T_i\}} \,\middle|\, \mathcal{F}_{T_{i-1}}\right)} \,\mathrm{d}\mathbb{P} \,,
\end{aligned}$$

for some appropriate $U' \in \mathcal{F}_{T_{i-1}}$ according to Equation (A.82). As $\{T_{i-1} \leq s < T_i\} = \{T_{i-1} \leq s\} \cap \{T_i \leq s\}^{\complement} \in \mathcal{F}_s$ by the usual argument, we therefore obtain:

$$\mathbb{E}\left(g\left(W_i\right) \,|\, \mathcal{F}_s\right) = \frac{\mathbb{E}\left(g\left(W_i\right) \mathbb{1}_{\{T_{i-1} \leq s < T_i\}} \,\middle|\, \mathcal{F}_{T_{i-1}}\right)}{\mathbb{E}\left(\mathbb{1}_{\{T_{i-1} \leq s < T_i\}} \,\middle|\, \mathcal{F}_{T_{i-1}}\right)} \,, \quad \text{on } \{T_{i-1} \leq s < T_i\} \;. \tag{A.83}$$

Furthermore, we have $\{T_{i-1} \leq s < T_i\} = \{W_i > s - T_{i-1} \geq 0\}$, so that Equation (A.83)

[57] Given that the prior σ-algebra $\mathcal{G}_0$ is included in both $\mathcal{F}_{T_{i-1}}$ and $\mathcal{F}_t$, it does not affect the application of the monotone class theorem and can thus be neglected in a proof of Equation (A.82).

can be restated as:

$$\mathbb{E}\left(g\left(W_i\right) \mid \mathcal{F}_s\right) = \frac{\mathbb{E}\left(g\left(W_i\right) \mathbb{1}_{\{W_i > s - T_{i-1}\}} \mid \mathcal{F}_{T_{i-1}}\right)}{\mathbb{E}\left(\mathbb{1}_{\{W_i > s - T_{i-1}\}} \mid \mathcal{F}_{T_{i-1}}\right)}$$
$$= \frac{1}{1 - F_i\left(s - T_{i-1} \mid \mathcal{F}_{T_{i-1}}\right)} \int_{s-T_{i-1}}^{\infty} g(x) \, \mathrm{d}F_i\left(x \mid \mathcal{F}_{T_{i-1}}\right) , \tag{A.84}$$

and we find ourselves in the situation of Equation (A.80). Thus, on $\{T_{i-1} \leq s < T_i\}$, the proof of the martingale equality can be carried out as in Lemma A.30. Additionally, $\{T_i \leq s\}$ resembles the case $\{X \leq s\}$ of Lemma A.30, and since we have $\Lambda_i(t, \omega) = H_i\left(T_i - T_{i-1} \mid \mathcal{F}_{T_{i-1}}\right) = \Lambda_i(s, \omega)$ as well as $N_i(t, \omega) = 1 = N_i(s, \omega)$, the martingale equality is trivially fulfilled. We now turn to decomposition (2b) from Figure 20:
In contrast to decomposition (2a), where the sets $\{T_i \leq s\}$ and $\{T_{i-1} \leq s < T_i\}$ could be identified with the sets $\{X \leq s\}$ and $\{X > s\}$, respectively, the decomposition (2b) has no such counterpart within the proof of Lemma A.30 and appears only as the special (but trivial) case $s = 0$. As before, the case $\{s < t < T_{i-1}\}$ is trivial (all terms involved are zero). For the remaining case $\{s < T_{i-1} \leq t\}$, scrutiny of the restricted σ-algebras yields as per construction:

$$\mathcal{F}_{T_{i-1}}\big|_{\{T_{i-1} > s\}} \supset \mathcal{F}_s\big|_{\{T_{i-1} > s\}} .$$

On $\{s < T_{i-1} \leq t\} \subset \{s < T_{i-1}\}$, this allows us to exploit the tower property for nested σ-algebras to obtain:

$$\mathbb{E}\left(N_i(t) - \Lambda_i(t) \mid \mathcal{F}_s\right) = \mathbb{E}\Big(\mathbb{E}\left(N_i(t) - \Lambda_i(t) \mid \mathcal{F}_{T_{i-1}}\right) \mid \mathcal{F}_s\Big) , \tag{A.85}$$

which significantly facilitates the calculation of the conditional expectation, since we have:

$$\mathbb{E}\left(g\left(W_i\right) \mid \mathcal{F}_{T_{i-1}}\right) = \int_0^{\infty} g(x) \, \mathrm{d}F_i\left(x \mid \mathcal{F}_{T_{i-1}}\right) , \tag{A.86}$$

for each non-negative measurable function g. This mirrors our initial comment in that this identity - albeit not mathematically rigorous! - appears as the special case $s = T_{i-1}$ (which corresponds to $s = 0$ in Lemma A.30) of Equation (A.84). Indeed, from Equation (A.86) we derive analogously to Equations (A.71) and (A.74) that

$$\mathbb{E}\left(N_i(t) \mid \mathcal{F}_{T_{i-1}}\right) = F_i\left(t - T_{i-1} \mid \mathcal{F}_{T_{i-1}}\right) = \mathbb{E}\left(\Lambda_i(t) \mid \mathcal{F}_{T_{i-1}}\right) , \quad \text{on } \{s < T_{i-1} \leq t\} ,$$

which in conjunction with Equation (A.85) implies:

$$\mathbb{E}\left(N_i(t) - \Lambda_i(t) \mid \mathcal{F}_s\right) = 0 , \quad \text{on } \{s < T_{i-1} \leq t\} .$$

The martingale equality follows from this, as clearly $N_i(s) = 0 = \Lambda_i(s)$ holds on $\{s < T_{i-1}\}$, thereby completing the proof. □

Proof of Corollary A.34

Corollary A.34 (Jacod's Formula, absolutely continuous case; cf. Brémaud 1981, pp. 61–63 and Daley and Vere-Jones 2008, pp. 364–365).
In the situation of Theorem A.33, the $\left(\mathbb{P}, \{\mathcal{F}_t\}_{t \geq t_0}\right)$*-compensator* Λ *is almost surely absolutely continuous if and only if the conditional distribution functions* $F_i\left(\cdot \mid \mathcal{F}_{T_{i-1}}\right)$

have absolutely continuous versions with densities $f_i\left(\cdot \mid \mathcal{F}_{T_{i-1}}\right)$. In this case, one version of Λ is given by

$$\Lambda(t,\omega) = \int_{t_0}^{t} \lambda^*(u,\omega)\,\mathrm{d}u\,,$$

where

$$\lambda^*(t,\omega) = \sum_{i=1}^{\infty} \lambda_i^*(t,\omega) \equiv \sum_{i=1}^{\infty} \frac{f_i\left(t - T_{i-1} \mid \mathcal{F}_{T_{i-1}}\right)}{1 - F_i\left(t - T_{i-1} \mid \mathcal{F}_{T_{i-1}}\right)} \mathbb{1}_{\{T_{i-1} \le t < T_i\}}\,. \tag{A.27}$$

An $\{\mathcal{F}_t\}_{t\ge t_0}$-predictable version λ of λ^* and hence the $\left(\mathbb{P}, \{\mathcal{F}_t\}_{t\ge t_0}\right)$-intensity of N is defined by

$$\lambda(t,\omega) = \sum_{i=1}^{\infty} \lambda_i(t,\omega) \equiv \sum_{i=1}^{\infty} \frac{f_i\left(t - T_{i-1} \mid \mathcal{F}_{T_{i-1}}\right)}{1 - F_i\left(t - T_{i-1} \mid \mathcal{F}_{T_{i-1}}\right)} \mathbb{1}_{\{T_{i-1} < t \le T_i\}}\,.$$

Proof. A direct proof of Corollary A.34 can be found in Brémaud 1981, pp. 61–63, but we can infer the results immediately from Theorem A.33. In the absolutely continuous case, Equation (A.25) can be stated as[58]

$$H_i\left(t \mid \mathcal{F}_{T_{i-1}}\right) = \int_0^t \frac{f_i\left(x \mid \mathcal{F}_{T_{i-1}}\right)}{1 - F_i\left(x \mid \mathcal{F}_{T_{i-1}}\right)}\,\mathrm{d}x$$

with Lebesgue density

$$h_i\left(x \mid \mathcal{F}_{T_{i-1}}\right) := \frac{f_i\left(x \mid \mathcal{F}_{T_{i-1}}\right)}{1 - F_i\left(x \mid \mathcal{F}_{T_{i-1}}\right)}\,,$$

the conditional equivalent of the hazard function introduced in Remark A.31. By Equation (A.24), the compensator Λ_i of N_i is constant on $\{t < T_{i-1}\}$ as well as $\{T_i \le t\}$, which implies that the intensity vanishes almost everywhere on $\{T_{i-1} \le t < T_i\}^{\complement}$, while on $\{T_{i-1} \le t < T_i\}$ we have:

$$\begin{aligned}
\Lambda_i(t,\cdot) &= H_i\left(t - T_{i-1} \mid \mathcal{F}_{T_{i-1}}\right) \\
&= \int_0^{t-T_{i-1}} \frac{f_i\left(x \mid \mathcal{F}_{T_{i-1}}\right)}{1 - F_i\left(x \mid \mathcal{F}_{T_{i-1}}\right)}\,\mathrm{d}x \\
&= \int_{T_{i-1}}^{t} \frac{f_i\left(x - T_{i-1} \mid \mathcal{F}_{T_{i-1}}\right)}{1 - F_i\left(x - T_{i-1} \mid \mathcal{F}_{T_{i-1}}\right)}\,\mathrm{d}x \\
&= \int_0^{t} \underbrace{\frac{f_i\left(x - T_{i-1} \mid \mathcal{F}_{T_{i-1}}\right)}{1 - F_i\left(x - T_{i-1} \mid \mathcal{F}_{T_{i-1}}\right)} \mathbb{1}_{\{T_{i-1} \le t < T_i\}}}_{=:\,\lambda_i^*(t,\cdot)}\,\mathrm{d}x\,.
\end{aligned}$$

By additivity, we thus obtain the density specified in Equation (A.27). Since the difference between λ^* and λ occurs only at the T_i's, λ is a version of λ^*, as we have for each $t \ge t_0$:

$$\mathbb{P}\left(\{\omega\,:\,\lambda^*(t,\omega) \ne \lambda(t,\omega)\}\right) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} \{\omega\,:\,T_i(\omega) = t\}\right) = 0\,,$$

[58]Note that an absolutely continuous version of $F_i\left(\cdot \mid \mathcal{F}_{T_{i-1}}\right)$ is necessarily continuous, which renders the consideration of the left limit obsolete.

because the countable union of $\mathbb{P}$-null sets is itself a null set and

$$\begin{aligned}\mathbb{P}\left(T_i = t\right) &= \mathbb{E}\left(\mathbb{1}_{\{T_i=t\}}\right) = \mathbb{E}\left(\mathbb{E}\left(\mathbb{1}_{\{T_i=t\}} \,|\, \mathcal{F}_{T_{i-1}}\right)\right) \\ &= \mathbb{E}\left(\mathbb{E}\left(\mathbb{1}_{\{W_i=t-T_{i-1}\}} \,|\, \mathcal{F}_{T_{i-1}}\right)\right) = \mathbb{E}\Big(\mathbb{P}\left(W_i = t - T_{i-1} \,|\, \mathcal{F}_{T_{i-1}}\right)\Big) = 0\,,\end{aligned}$$

by the absolute continuity of the conditional distribution functions $F_i\left(\cdot \,|\, \mathcal{F}_{T_{i-1}}\right)$. Furthermore, the shape of this density entails its predictability by virtue of Theorem A2, T24 of Brémaud 1981, p. 304, a corollary of the aforementioned Proposition 2.6 of Karr 1991, p. 57. Thus, according to Remark A.27, λ is the "unique" $\left(\mathbb{P}, \{\mathcal{F}_t\}_{t\geq t_0}\right)$-intensity of N. $\square$

Proof of Lemma A.37

Lemma A.37 (Common Discontinuities of Independent Counting Processes)**.**
Let $N^{(1)}, N^{(2)}, \ldots$ be an at most countable collection of independent counting processes with index set $[t_0, \infty)$ defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Suppose that, for each $j \in \mathbb{N}$, the associated compensator $\Lambda^{(j)}$ with respect to the canonical filtration of $N^{(j)}$ is absolutely continuous. Let $T^{(1)}, T^{(2)}, \ldots$ denote the corresponding counting processes. Then, for all $i, j, k, l \in \mathbb{N}$ where $j \neq l$ or $i \neq k$, the following holds:

$$\mathbb{P}\left(T_i^{(j)} = T_k^{(l)}\,, T_i^{(j)} < \infty\right) = 0\,. \tag{A.35}$$

In terms of the processes $N^{(1)}, N^{(2)}, \ldots$, this implies that $\mathbb{P}$-almost surely they exhibit no common discontinuities on $[t_0, \infty)$. In particular, this applies on any subinterval $\mathcal{I} \subset [t_0, \infty)$.

Proof. We first prove Equation (A.35) for any i, j, k, l with $j \neq l$ or $i \neq k$. We only need to consider the case $j \neq l$, since otherwise conditions (i) and (ii) of Definition A.3 imply that

$$\begin{aligned}\mathbb{P}\left(T_i^{(j)} = T_k^{(j)}\,, T_i^{(j)} < \infty\right) &= \mathbb{P}\left(T_i^{(j)} = T_{i+1}^{(j)} = \ldots = T_k^{(j)}\,, T_i^{(j)} < \infty\right) \\ &\leq \mathbb{P}\left(T_i^{(j)} = T_{i+1}^{(j)}\,, T_i^{(j)} < \infty\right) = 0\,,\end{aligned}$$

where without loss of generality we assumed $i < k$ and used the argument following Equation (A.59), see the proof of Remark A.6(ii) for details.
We observe that the absolute continuity of the compensator $\Lambda^{(j)}$ entails the absolute continuity of the conditional distribution functions $F_i^{(j)}(\,\cdot\,|\, T_1^{(j)}, \ldots, T_{i-1}^{(j)})$ by virtue of Corollary A.34 and denote by $f_i^{(j)}(\,\cdot\,|\, T_1^{(j)}, \ldots, T_{i-1}^{(j)})$ the corresponding Lebesgue densities. We can then step by step reconstruct the Lebesgue density of the joint distribution of $(T_1^{(j)}, \ldots, T_i^{(j)})$, that is:

$$\begin{aligned}f_{T_1^{(j)}}\left(t_1\right) &= f_1^{(j)}\left(t_1\right)\,, \\ f_{\left(T_1^{(j)}, T_2^{(j)}\right)}\left(t_1, t_2\right) &= f_{T_1^{(j)}}\left(t_1\right) \cdot f_2^{(j)}\left(t_2 \,|\, t_1\right)\,, \\ &\;\vdots \\ f_{\left(T_1^{(j)}, \ldots, T_i^{(j)}\right)}\left(t_1, \ldots, t_i\right) &= f_{\left(T_1^{(j)}, \ldots, T_{i-1}^{(j)}\right)}\left(t_1, \ldots, t_{i-1}\right) \cdot f_i^{(j)}\left(t_i \,|\, t_1, \ldots, t_{i-1}\right)\,,\end{aligned}$$

and integrating out accordingly yields the density of the marginal distribution of $T_i^{(j)}$. In particular, $T_i^{(j)}$ has an absolutely continuous probability distribution, and the same procedure can be repeated for $T_k^{(l)}$. Since $j \neq l$ implies the independence of these random variables, a Lebesgue density of the joint distribution of $T_i^{(j)}$ and $T_k^{(l)}$ can be obtained by the product of the respective marginal distributions. However, this means that the joint distribution of $T_i^{(j)}$ and $T_k^{(l)}$ is again absolutely continuous with respect to the Lebesgue measure. But then follows

$$\mathbb{P}\left(T_i^{(j)} = T_k^{(l)}, T_i^{(j)} < \infty\right) = \mathbb{P}^{\left(T_i^{(j)}, T_k^{(l)}\right)}(\underbrace{\{(x,x) \, : \, x \in [t_0, \infty)\}}_{=:W}) = 0\,,$$

because the (one-dimensional) half-line W is a Lebesgue null set due to the dimension deficiency. Turning to the associated counting processes, the discontinuities of the jth process $N^{(j)}$ are located at

$$\mathcal{J}_{N^{(j)}}(\omega) = \{T_i^{(j)}(\omega) \, : \, i \in \mathbb{N}\} \cap [t_0, \infty)\ .$$

Thus, two counting process $N^{(j)}$ and $N^{(l)}$ have a common discontinuity if and only if

$$\mathcal{J}_{N^{(j)}}(\omega) \cap \mathcal{J}_{N^{(l)}}(\omega) \neq \emptyset\,. \tag{A.87}$$

If $\omega \in \Omega$ is fixed, then in order for Equation (A.87) to hold, $i, k \in \mathbb{N}$ must exist such that $T_i^{(j)}(\omega) = T_k^{(l)}(\omega)$ and $T_i^{(j)}(\omega) < \infty$ are both satisfied. Hence,

$$\{\omega \in \Omega \, : \, \mathcal{J}_{N^{(j)}}(\omega) \cap \mathcal{J}_{N^{(l)}}(\omega) \neq \emptyset\} = \bigcup_{i,k \in \mathbb{N}} \left\{\omega \in \Omega \, : \, T_i^{(j)}(\omega) = T_k^{(l)}(\omega)\,, T_i^{(j)}(\omega) < \infty\right\}\ .$$

However, due to Equation (A.35), this event has probability zero as a countable union of $\mathbb{P}$-null sets. In the same way, the probability that countable many counting processes share a common point of discontinuity is again a $\mathbb{P}$-null set. $\square$

Addendum to the Proof of Lemma A.40

Lemma A.40 (Integration with Respect to Bounded Variation Martingales; Theorem T6 of Brémaud 1981, p. 10)**.**
Let $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq t_0}, \mathbb{P}\right)$ *be a filtered probability space and* $M = (M_t)_{t \geq t_0}$ *a* $\left(\mathbb{P}, \{\mathcal{F}_t\}_{t \geq t_0}\right)$*-martingale of bounded variation (i.e.,* $\mathbb{P}$*-almost all sample paths have bounded variation over finite intervals). Let* $|M| = (|M|_t)_{t \geq t_0}$ *denote the total variation process*[59] *associated with* M*. Suppose further that* M *is of locally integrable variation, that is,*

$$\mathbb{E}\left[\int_{t_0}^{t} \mathrm{d}\,|M|_s\right] < \infty\,, \quad \textit{for all}\ t \geq t_0\,.$$

Then, for each $\{\mathcal{F}_t\}_{t \geq t_0}$*-predictable process* $f = (f_t)_{t \geq t_0}$ *satisfying*

$$\mathbb{E}\left[\int_{t_0}^{t} |f_s|\ \mathrm{d}\,|M|_s\right] < \infty\,, \quad \textit{for all}\ t \geq t_0\,, \tag{A.38}$$

[59] Accordingly, $|M|_t < \infty$ is the total variation of the path $s \mapsto M_s$ over $[t_0, t]$ and is not to be confused with $|M_t|$ (Karr 1991, p. 59).

the stochastic process $X = (X_t)_{t \geq t_0}$*, which for each* $t \geq t_0$ *is defined by*

$$X_t = \int_{t_0}^{t} f_s \, \mathrm{d}M_s \,, \tag{A.39}$$

is a $\left(\mathbb{P}, \{\mathcal{F}_t\}_{t \geq t_0}\right)$*-martingale.*

Proof. We only show here that the process X from Equation (A.41) is indeed a martingale. Recall that X is given by

$$X_t = \mathbb{1}_A \cdot \left(M_{\min\{v,\max\{u,t\}\}} - M_u\right) \,. \tag{A.41}$$

For X to be a martingale, we need to verify that for all $t_0 \leq s \leq t$ holds:

$$\mathbb{E}\left(X_t - X_s \,|\, \mathcal{F}_s\right) = 0 \,. \tag{A.88}$$

This requires a case differentiation where we distinguish the relative positions of s and t with respect to u and v. We consider a total of six cases:

$t \leq u$: Here, $X_t = 0 = X_s$ and thus $X_t - X_s = 0$ holds unconditionally.

$s < u < t \leq v$: Again, $X_s = 0$. The tower property yields (recall that $A \in \mathcal{F}_u$):

$$\begin{aligned}\mathbb{E}\left(X_t \,|\mathcal{F}_s\right) &= \mathbb{E}\left(\mathbb{1}_A \cdot (M_t - M_u) \,|\mathcal{F}_s\right) = \mathbb{E}\left[\mathbb{E}\left(\mathbb{1}_A \cdot (M_t - M_u) \,|\mathcal{F}_u\right) \,|\, \mathcal{F}_s\right] \\ &= \mathbb{E}\left[\mathbb{1}_A \cdot \mathbb{E}\left(M_t - M_u \,|\mathcal{F}_u\right) \,|\, \mathcal{F}_s\right] = 0 \,,\end{aligned}$$

by the martingale property of M.

$s < u < v < t$: We proceed analogously to the previous case, replacing only M_t by M_v.

$u \leq s < t \leq v$: For this and the following case, note that $A \in \mathcal{F}_u \subset \mathcal{F}_s$. Hence, we have:

$$\begin{aligned}\mathbb{E}\left(X_t - X_s \,|\mathcal{F}_s\right) &= \mathbb{E}\left(\mathbb{1}_A \cdot (M_t - M_u - (M_s - M_u)) \,|\mathcal{F}_s\right) \\ &= \mathbb{1}_A \cdot \mathbb{E}\left(M_t - M_s \,|\mathcal{F}_s\right) = 0 \,.\end{aligned}$$

$u \leq s < v < t$: Once more, only M_t needs to be replaced by M_v.

$s \geq v$: Similar to the first case, $X_t = X_s$ implies $X_t - X_s = 0$ unconditionally.

Therefore, Equation (A.88) holds for all $t_0 \leq s \leq t$. Since the adaptedness and integrability of X are inherited from M, it follows that X is a martingale. □

A.5. Further Explanations and Remarks

This part of Appendix A contains further notes that aid in the understanding of this thesis, but would impair the flow of reading if incorporated into the main body. Unlike its technical counterpart in Section A.4, the current section is characterized by mathematically less rigorous remarks that do not claim the status of a proof. Consequently, the following comments merely serve to round off the overall picture, but by no means as a basis for a mathematically precise discussion of the subject.

Remark A.48 (Heuristical Explanation of the Suggestive Relation (A.17)).
In Equation (A.17), we established the suggestive relation

$$\lambda(t)\,\mathrm{d}t \approx \mathbb{E}\left(N(\mathrm{d}t)\,|\,\mathcal{F}_{t-}\right)\,,$$

but refrained from giving a comprehensible justification. In order to provide one, we follow Karr 1991, pp. 61 & 69 and first recall that for any $t_0 \leq s < t$ we have by virtue of the martingale property:

$$\mathbb{E}\left(N_t - N_s\,|\,\mathcal{F}_s\right) = \mathbb{E}\left(\Lambda_t - \Lambda_s\,|\,\mathcal{F}_s\right)\,. \tag{A.89}$$

Note that since $\{\mathcal{F}_t\}_{t\geq t_0}$ is a filtration and hence an increasing sequence of σ-algebras, we have

$$\sigma\left(\lim_{s\uparrow t}\mathcal{F}_s\right) = \sigma\left(\bigcup_{s<t}\mathcal{F}_s\right) = \mathcal{F}_{t-}\,.$$

In infinitesimal form (i.e., for an infinitesimally small difference $t-s$ denoted with $\mathrm{d}t$), Equation (A.89) therefore becomes the heuristic expression

$$\mathbb{E}\left(N(\mathrm{d}t)\,|\,\mathcal{F}_{t-}\right) = \mathbb{E}\left(\Lambda(\mathrm{d}t)\,|\,\mathcal{F}_{t-}\right) = \Lambda(\mathrm{d}t)\,, \tag{A.90}$$

where the last equation "holds" due to Lemma A.22, because Λ is $\{\mathcal{F}_t\}_{t\geq t_0}$-predictable. If for $\mathbb{P}$-almost all $\omega \in \Omega$, the function $t \mapsto \Lambda(t,\omega)$ is absolutely continuous with respect to the Lebesgue measure, it admits a Radon-Nikodym derivative λ and the extension of the fundamental theorem of calculus to Lebesgue integrals (Theorem 7.11 of Rudin 1987, p. 141) yields

$$\lim_{s\to t}\frac{\Lambda_t - \Lambda_s}{t-s} = \lambda_t \tag{A.91}$$

almost everywhere. Transferring Equation (A.91) once again into infinitesimal form gives

$$\frac{\Lambda(\mathrm{d}t)}{\mathrm{d}t} = \lambda_t \qquad \rightsquigarrow \qquad \Lambda(\mathrm{d}t) = \lambda_t\,\mathrm{d}t\,, \tag{A.92}$$

and substituting Equation (A.92) into Equation (A.90) results in the desired relation.

The following theorem and subsequent remark revolve around càdlàg functions of bounded variation (over finite intervals), see Appendix A4 of Brémaud 1981, pp. 334–339 for an overview.

Theorem A.49 (Product Formula of the Stieltjes-Lebesgue Calculus; cf. Daley and Vere-Jones 2003, p. 107 and Brémaud 1981, p. 336).
Let f and g be two càdlàg functions of bounded variation over finite intervals. Then the following holds:

$$f(t)g(t) = f(s)g(s) + \int_s^t f(x)\,\mathrm{d}g(x) + \int_s^t g(x-)\,\mathrm{d}f(x)\,.$$

This result is also known as the Integration-by-Parts Formula.

Proof. The proof amounts to an application of Fubini's theorem and can be found in Brémaud 1981, pp. 336-337. □

A further extension of Theorem A.49 is found in Proposition 2.8 of Karr 1991.

Remark A.50 (Extension of the Integration-by-Parts Formula; Karr 1991, p. 58).
Recall the notation of Remark A.6. In the situation of Theorem A.49, it further holds:

$$f(t)g(t) = f(s)g(s) + \int_s^t f(x-)\,\mathrm{d}g(x) + \int_s^t g(x-)\,\mathrm{d}f(x) + \sum_{s \le x \le t} \Delta f(x)\Delta g(x)\,.$$

Remark A.51 (Predictability in Jacod's Formula for the Intensity Process).
A naive approach to the proof of predictability in Theorem A.33 can be given based on the technique presented in the proof of Lemma A.30. We note that the asserted form of the compensator from Equation (A.24) can equivalently be stated as

$$\Lambda_i(t,\omega) = H_i\left(t \wedge T_i(\omega) - t \wedge T_{i-1}(\omega) \,|\, \mathcal{F}_{T_{i-1}}\right)\,. \tag{A.93}$$

Scrutiny of Equation (A.68) shows that the processes

$$(t,\omega) \longmapsto t \wedge T_i(\omega)\,, \quad i \in \mathbb{N}\,,$$

are predictable (note that this is true regardless of the concrete value of i), and since the integrated conditional hazard function is again increasing and right-continuous, one is tempted to argue by means of Equation (A.69) that Λ_i must also be predictable. While certainly valid from a heuristic standpoint, the above proof is not mathematically rigorous: The transition from the unconditional to the conditional integrated hazard function entails a dependence on ω itself, which has been suppressed here (e.g., in Equation (A.93)) in favor of easier comprehensibility. However, if we apply Proposition 2.6 of Karr 1991, p. 57 instead of Equation (A.69), the desired result can be achieved nonetheless.

Remark A.52 (Effects of Type I Censoring on the Hazard Transformation).
Suppose that in the situation of Theorem A.46, the points of the simple point process $T = (T_i)_{i\in\mathbb{N}}$ are right-censored at a preset time τ. We deal in this remark only with the case where τ is deterministic, although in the context of intensity-based load sharing models, random type I censoring may occur. The hazard transformed process $R = (R_i)_{i\in\mathbb{N}}$ then no longer consists of independent exponentially distributed random variables. This can be seen by (semi-)explicitly computing the conditional distribution of R_i for $i \in \mathbb{N}$: Note that due to censoring at τ, we need to consider the process $\left(T_i^{(c)}\right)_{i\in\mathbb{N}}$, where

$$T_i^{(c)} := \min\{T_i, \tau\}\,, \quad i \in \mathbb{N}\,.$$

Let $0 \le t_1 < \ldots < t_{i-1} < \tau$ and let $S_i^{(c)}\left(t \,|\, t_{1:(i-1)}\right)$ denote the survival function of $T_i^{(c)}$ given $T_{1:(i-1)}^{(c)} = t_{1:(i-1)}$. Since $t_{i-1} < \tau$, we can replace $T_{1:(i-1)}^{(c)}$ with $T_{1:(i-1)}$ in the condition to obtain:

$$\begin{aligned}
S_i^{(c)}\left(t \,|\, t_{1:(i-1)}\right) &= \mathbb{P}\left(T_i^{(c)} > t \,|\, T_{1:(i-1)}^{(c)} = t_{1:(i-1)}\right) \\
&= \mathbb{P}\left(\min\{T_i, \tau\} > t \,|\, T_{1:(i-1)} = t_{1:(i-1)}\right) \\
&= \mathbb{P}\left(\{T_i > t\} \cap \{\tau > t\} \,|\, T_{1:(i-1)} = t_{1:(i-1)}\right) \\
&= \mathbb{P}\left(T_i > t \,|\, T_{1:(i-1)} = t_{1:(i-1)}\right) \cdot \mathbb{P}\left(\tau > t\right)
\end{aligned}$$

$$= \begin{cases} S_i\left(t \,\middle|\, t_{1:(i-1)}\right), & t < \tau\,, \\ 0\,, & t \geq \tau\,, \end{cases} \tag{A.94}$$

where in the penultimate step we utilized that τ is independent of $T_1, \ldots, T_i$ as a deterministic constant. With the same argument, the probability of $\{\tau > t\}$ is either 1 (if $t < \tau$) or 0 (if $t \geq \tau$), explaining the last step. If $t_{i-1} \geq \tau$, then $T_i^{(c)} = \min\{T_i, \tau\} = \tau$ and we receive analogously:

$$S_i^{(c)}\left(t \,\middle|\, t_{1:(i-1)}\right) = \mathbb{P}\left(\tau > t \,\middle|\, T_{1:(i-1)}^{(c)} = t_{1:(i-1)}\right) = \mathbb{P}\left(\tau > t\right) = \begin{cases} 1\,, & t < \tau\,, \\ 0\,, & t \geq \tau\,. \end{cases}$$

We now discuss the conditional distribution of the random variable $S_i^{(c)}\left(T_i^{(c)} \,\middle|\, T_{1:(i-1)}^{(c)}\right)$, which is closely tied to the hazard transform R_i of $T_i^{(c)}$ by virtue of Equation (A.51). For $u \in (0,1)$ and $t_{i-1} < \tau$, we calculate by once again replacing $T_{1:(i-1)}^{(c)}$ with $T_{1:(i-1)}$ wherever necessary:

$$\begin{aligned}
&\mathbb{P}\left(S_i^{(c)}\left(T_i^{(c)} \,\middle|\, T_{1:(i-1)}^{(c)}\right) \geq u \,\middle|\, T_{1:(i-1)}^{(c)} = t_{1:(i-1)}\right) \\
&= \mathbb{P}\left(\left\{S_i^{(c)}\left(T_i^{(c)} \,\middle|\, t_{1:(i-1)}\right) \geq u\right\} \cap \{T_i < \tau\} \,\middle|\, T_{1:(i-1)} = t_{1:(i-1)}\right) \\
&\quad + \mathbb{P}\left(\left\{S_i^{(c)}\left(T_i^{(c)} \,\middle|\, t_{1:(i-1)}\right) \geq u\right\} \cap \{T_i \geq \tau\} \,\middle|\, T_{1:(i-1)}^{(c)} = t_{1:(i-1)}\right) \\
&\overset{\text{(A.94)}}{=} \mathbb{P}\left(\left\{S_i\left(T_i \,\middle|\, t_{1:(i-1)}\right) \geq u\right\} \cap \{T_i < \tau\} \,\middle|\, T_{1:(i-1)} = t_{1:(i-1)}\right) \\
&\quad + \underbrace{\mathbb{P}\left(\{0 \geq u\} \cap \{T_i \geq \tau\} \,\middle|\, T_{1:(i-1)}^{(c)} = t_{1:(i-1)}\right)}_{= 0, \text{ since } u > 0.} \\
&= \mathbb{P}\left(\left\{T_i \leq S_i^{-1}\left(u \,\middle|\, t_{1:(i-1)}\right)\right\} \cap \{T_i < \tau\} \,\middle|\, T_{1:(i-1)} = t_{1:(i-1)}\right) \\
&= \mathbb{P}\left(T_i \leq \min\left\{S_i^{-1}\left(u \,\middle|\, t_{1:(i-1)}\right), \tau\right\} \,\middle|\, T_{1:(i-1)} = t_{1:(i-1)}\right) \\
&= F_i\left(\min\left\{S_i^{-1}\left(u \,\middle|\, t_{1:(i-1)}\right), \tau\right\} \,\middle|\, t_{1:(i-1)}\right) \\
&= \begin{cases} F_i\left(S_i^{-1}\left(u \,\middle|\, t_{1:(i-1)}\right) \,\middle|\, t_{1:(i-1)}\right), & \text{if } S_i^{-1}\left(u \,\middle|\, t_{1:(i-1)}\right) < \tau\,, \\ F_i\left(\tau \,\middle|\, t_{1:(i-1)}\right), & \text{if } S_i^{-1}\left(u \,\middle|\, t_{1:(i-1)}\right) \geq \tau\,, \end{cases} \\
&= \begin{cases} 1 - u\,, & \text{if } u > S_i\left(\tau \,\middle|\, t_{1:(i-1)}\right), \\ F_i\left(\tau \,\middle|\, t_{1:(i-1)}\right), & \text{if } u \leq S_i\left(\tau \,\middle|\, t_{1:(i-1)}\right), \end{cases}
\end{aligned}$$

utilizing that $\mathbb{P}\left(T_i = \tau \,\middle|\, T_{1:(i-1)} = t_{1:(i-1)}\right) = 0$ according to the proof of Theorem A.46. For $t < H_i^{(c)}\left(\tau \,\middle|\, t_{1:(i-1)}\right)$, plugging in Equation (A.51) immediately yields:

$$\begin{aligned}
&\mathbb{P}\left(H_i^{(c)}\left(T_i^{(c)} \,\middle|\, T_{1:(i-1)}^{(c)}\right) \leq t \,\middle|\, T_{1:(i-1)}^{(c)} = t_{1:(i-1)}\right) \\
&= \mathbb{P}\left(S_i^{(c)}\left(T_i^{(c)} \,\middle|\, T_{1:(i-1)}^{(c)}\right) \geq \exp(-t) \,\middle|\, T_{1:(i-1)}^{(c)} = t_{1:(i-1)}\right) \\
&= 1 - \exp(-t)\,,
\end{aligned} \tag{A.95}$$

since the remaining case $\exp(-t) \leq S_i\left(\tau \,\middle|\, t_{1:(i-1)}\right)$ translates to $t \geq H_i^{(c)}\left(\tau \,\middle|\, t_{1:(i-1)}\right)$, which is considered separately due to the lack of differentiability of the conditional survival function in τ. Here, Equations (A.30) and (A.31) state that $H_i^{(c)}\left(\cdot \,\middle|\, t_{1:(i-1)}\right)$ becomes constant starting from τ, so that for $t \geq H_i^{(c)}\left(\tau \,\middle|\, t_{1:(i-1)}\right)$ we observe:

$$\begin{aligned}
&\mathbb{P}\left(H_i^{(c)}\left(T_i^{(c)} \,\middle|\, T_{1:(i-1)}^{(c)}\right) \leq t \,\middle|\, T_{1:(i-1)}^{(c)} = t_{1:(i-1)}\right) \\
&\quad \geq \mathbb{P}\left(H_i^{(c)}\left(T_i^{(c)} \,\middle|\, t_{1:(i-1)}\right) \leq H_i^{(c)}\left(\tau \,\middle|\, t_{1:(i-1)}\right) \,\middle|\, T_{1:(i-1)}^{(c)} = t_{1:(i-1)}\right) = 1\,. \qquad \text{(A.96)}
\end{aligned}$$

Combining Equations (A.95) and (A.96) yields:

$$\begin{aligned}
&\mathbb{P}\left(H_i^{(c)}\left(T_i^{(c)} \,\middle|\, T_{1:(i-1)}^{(c)}\right) \leq t \,\middle|\, T_{1:(i-1)}^{(c)} = t_{1:(i-1)}\right) \\
&\quad = \begin{cases} 1 - \exp(-t)\,, & \text{if } t < -\ln\left(S_i\left(\tau \,\middle|\, t_{1:(i-1)}\right)\right)\,, \\ 1\,, & \text{if } t \geq -\ln\left(S_i\left(\tau \,\middle|\, t_{1:(i-1)}\right)\right)\,. \end{cases}
\end{aligned}$$

As $\tau \to \infty$ we have $-\ln\left(S_i\left(\tau \,\middle|\, t_{1:(i-1)}\right)\right) \to \infty$, so this formula conforms with Equation (A.53). However, for $\tau < \infty$ it becomes apparent that the conditional distribution is neither exponential nor independent of the points $t_1, \ldots, t_{i-1}$. This also carries over to the unconditional distribution, although the dichotomy of $T_{i-1} < \tau$ and $T_{i-1} \geq \tau$ complicates its computation. We therefore refrain from any further analysis and close this remark.

B. Technical Proofs and Complementary Notes

B.1. Proofs of Uniform Bounds for the Intensity and its Partial Derivatives

This part of Appendix B supplements Section 2.4 of the thesis. We provide the technical proofs omitted there. For the convenience of the reader, we repeat the corresponding results with the numbering used in the main body.

Lemma 2.14 (Integrals of the Natural Powers of ln)**.**
Let $t > 0$, $q > -1$ and $p \in \mathbb{N}_0$. Then,

$$\int_0^t x^q \cdot (\ln x)^p \,\mathrm{d}x = \frac{t^{1+q}}{1+q} \sum_{k=0}^{p} \left(\frac{-1}{1+q}\right)^{p-k} \frac{p!}{k!} (\ln t)^k \,. \tag{2.33}$$

Proof. The validity of Equation (2.33) is easily verified by induction on p. For the base case $p = 0$, it boils down to

$$\int_0^t x^q \,\mathrm{d}x \overset{!}{=} \frac{t^{1+q}}{1+q} \,,$$

which is obviously true since $q > -1$ implies

$$\int_0^t x^q \,\mathrm{d}x = \left[\frac{x^{1+q}}{1+q}\right]_{x=0}^{t} \,.$$

For the induction step, if we assume Equation (2.33) to hold for some $p \in \mathbb{N}_0$, integration by parts (differentiating $(\ln x)^{p+1}$ and integrating x^q) yields:

$$\begin{aligned}
\int_0^t x^q \cdot (\ln x)^{p+1} \,\mathrm{d}x &= \left[\frac{x^{1+q}}{1+q} \cdot (\ln x)^{p+1}\right]_{x=0}^{t} - \int_0^t \frac{x^{1+q}}{1+q} \cdot \frac{p+1}{x} (\ln x)^p \,\mathrm{d}x \\
&\overset{(*)}{=} \frac{t^{1+q}}{1+q} \cdot (\ln t)^{p+1} - \frac{p+1}{1+q} \int_0^t x^q \cdot (\ln x)^p \,\mathrm{d}x \quad \big| \text{substitute Eq. (2.33)} \\
&= \frac{t^{1+q}}{1+q} \cdot (\ln t)^{p+1} - \frac{p+1}{1+q} \left[\frac{t^{1+q}}{1+q} \sum_{k=0}^{p} \left(\frac{-1}{1+q}\right)^{p-k} \frac{p!}{k!} (\ln t)^k\right] \\
&= \frac{t^{1+q}}{1+q} \cdot \underbrace{\left(\frac{-1}{1+q}\right)^{(p+1)-(p+1)} \frac{(p+1)!}{(p+1)!}}_{=1} (\ln t)^{p+1} \\
&\quad + \frac{t^{1+q}}{1+q} \sum_{k=0}^{p} \left(\frac{-1}{1+q}\right)^{(p+1)-k} \frac{(p+1)!}{k!} (\ln t)^k \\
&= \frac{t^{1+q}}{1+q} \sum_{k=0}^{p+1} \left(\frac{-1}{1+q}\right)^{(p+1)-k} \frac{(p+1)!}{k!} (\ln t)^k \,.
\end{aligned}$$

Formally, the above integral is improper for $q \in (-1, 0]$. Depending on $p \in \mathbb{N}_0$, the integrand $x^q \cdot (\ln x)^{p+1}$ tends to either ∞ or $-\infty$ as $t \to 0$ and therefore cannot be evaluated at 0. Therefore, in $(*)$ we take limits implicitly to allow for easier reading, keeping in mind that by L'Hôpital's rule for each $p \in \mathbb{N}_0$ we have:

$$\lim_{x \downarrow 0} x^{1+q} (\ln x)^p = \lim_{x \downarrow 0} \frac{(\ln x)^p}{x^{-(1+q)}} = \lim_{x \downarrow 0} \frac{p (\ln x)^{p-1} \cdot \frac{1}{x}}{-(1+q) \cdot x^{-(1+q)-1}} = p \left(\frac{-1}{1+q}\right) \lim_{x \downarrow 0} \frac{(\ln x)^{p-1}}{x^{-(1+q)}}$$

$$= \ldots = p! \left(\frac{-1}{1+q}\right)^p \lim_{x\downarrow 0} \frac{(\ln(x))^{p-p}}{x^{-(1+q)}} = p! \left(\frac{-1}{1+q}\right)^p \lim_{x\downarrow 0} x^{1+q} = 0\,.$$

□

Lemma 2.15 (Integrable Bounds for the Intensity Partial Derivatives in the Basquin Load Sharing Model with Multiplicative Damage Accumulation)**.**
Let ${}^{\times D}\lambda_\theta^{(j)}(t)$ be the conditional intensity function of the Basquin load sharing model with multiplicative damage accumulation given in Definition 2.8 of Subsection 2.3.2, that is:

$$^{\times D}\lambda_\theta^{(j)}(t) := \theta_1 \Bigg(\underbrace{s_j \frac{I}{I - N_{t^-}^{(j)}}}_{=:B_j(t)}\Bigg)^{\theta_2} \Bigg(\underbrace{\frac{1}{\tau}\int_0^t s_j \frac{I}{I - N_{u^-}^{(j)}}\,\mathrm{d}u}_{=:A_j(t)}\Bigg)^{\theta_3} \cdot \mathbb{1}_{\left\{N_{t^-}^{(j)} < C_j\right\} \cap \left\{t \le \tau_j\right\}}\,.$$

Suppose that the sequence $(s_j)_{j\in\mathbb{N}}$ is bounded both downward by some $0 < s_{\mathrm{low}} \le 1$ and upward by an arbitrary constant s_{upp} (e.g., if a preset assortment of initial stress levels $s_1, \ldots, s_L \ge 1$ is consecutively repeated). If we assume that $\Theta \subset \mathbb{R}_+^3$, then the following holds for all $t \in \mathcal{I}$, $\theta \in \Theta$ and $\omega \in \Omega$:

$$\left| \frac{\partial^{p+q+r}}{\partial\theta_1^r \partial\theta_2^p \partial\theta_3^q} {}^{\times D}\lambda_\theta^{(j)}(t) \right| \le \max\{1, \theta_1\} \cdot \left[\ln\left(\frac{s_j I}{s_{\mathrm{low}}^2} \cdot \frac{\tau}{t}\right)\right]^{p+q} \cdot (s_j I)^{\theta_2+\theta_3}\,, \quad p, q, r \in \mathbb{N}_0\,, j \in \mathbb{N}\,. \tag{2.34}$$

If furthermore the parameter space Θ is bounded, there exists a constant C independent of $\theta \in \Theta$ and $j \in \mathbb{N}$ such that

$$\left| \frac{\partial^{p+q+r}}{\partial\theta_1^r \partial\theta_2^p \partial\theta_3^q} {}^{\times D}\lambda_\theta^{(j)}(t) \right| \le C \cdot \left[\ln\left(\frac{s_j I}{s_{\mathrm{low}}^2} \cdot \frac{\tau}{t}\right)\right]^{p+q}\,, \quad p, q, r \in \mathbb{N}_0\,, j \in \mathbb{N}\,. \tag{2.35}$$

Under these assumptions, differentiation of arbitrary order with respect to $\theta \in \Theta$ and integration with respect to $t \in \mathcal{I}$ are interchangeable, that is,

$$\frac{\mathrm{d}^p}{\mathrm{d}\theta^p} {}^{\times D}\Lambda_\theta^{(j)}(t) = \int_0^t \frac{\mathrm{d}^p}{\mathrm{d}\theta^p} {}^{\times D}\lambda_\theta^{(j)}(u)\, du\,, \quad p \in \mathbb{N}\,,$$

and we have

$$\left| \frac{\partial^{p+q+r}}{\partial\theta_1^r \partial\theta_2^p \partial\theta_3^q} {}^{\times D}\Lambda_\theta^{(j)}(t) \right| \le C\tau \sum_{k=0}^{p+q} \frac{(p+q)!}{k!} \left(\ln\left(\frac{s_j I}{s_{\mathrm{low}}^2}\right)\right)^k\,, \tag{2.36}$$

where s_j can be replaced by s_{upp} whenever a uniform bound is desired.

In preparation for the following proof, we first apply Lemma 2.14 to obtain a useful corollary on the integrability of the derived bounds.

Corollary B.1.1 (Integrability of Majorants for the Intensity Partial Derivatives in the Basquin Load Sharing Model with Multiplicative Damage Accumulation)**.**
In the situation of Lemma 2.15, let $C > 0$, $x \in \mathcal{I} = [0, \tau]$ and $0 < \varepsilon \leq 1$. Then,

$$\begin{aligned}
&\int_0^x C \cdot \left[\ln \left(\frac{s_j I}{s_{\text{low}}^2} \cdot \frac{\tau}{u} \right) \right]^{p+q} \cdot \left(s_{\text{low}} \frac{u}{\tau} \right)^{\varepsilon - 1} \mathrm{d}u \\
&\quad = \frac{C x^\varepsilon}{\varepsilon^{1+p+q}} \left(\frac{s_{\text{low}}}{\tau} \right)^{\varepsilon-1} \sum_{k=0}^{p+q} \frac{(p+q)!}{k!} \left(\varepsilon \ln \left(\frac{\tau s_j I}{x s_{\text{low}}^2} \right) \right)^k , \qquad p, q \in \mathbb{N}_0 , j \in \mathbb{N} .
\end{aligned}$$

Proof. An application of Lemma 2.14 with $q = \varepsilon - 1$ and $p + q$ in place of p directly yields the desired identity:

$$\begin{aligned}
&\int_0^x C \cdot \left[\ln \left(\frac{s_j I}{s_{\text{low}}^2} \cdot \frac{\tau}{u} \right) \right]^{p+q} \cdot \left(s_{\text{low}} \frac{u}{\tau} \right)^{\varepsilon - 1} \mathrm{d}u \\
&\quad = (-1)^{p+q} C \int_0^x \left[\ln \left(\underbrace{\frac{s_{\text{low}}^2}{\tau s_j I}}_{=:\kappa} \cdot u \right) \right]^{p+q} \cdot \left(s_{\text{low}} \frac{u}{\tau} \right)^{\varepsilon - 1} \mathrm{d}u \\
&\quad = (-1)^{p+q} C \int_0^x \left[\ln (\kappa u) \right]^{p+q} \cdot \left(\frac{s_j I}{s_{\text{low}}} \cdot \kappa u \right)^{\varepsilon - 1} \mathrm{d}u \qquad \Big| v = \kappa u \\
&\quad = (-1)^{p+q} \frac{C}{\kappa} \left(\frac{s_j I}{s_{\text{low}}} \right)^{\varepsilon - 1} \int_0^{\kappa x} (\ln v)^{p+q} \, v^{\varepsilon - 1} \, \mathrm{d}v \qquad \Big| \text{apply Lemma 2.14} \\
&\quad = (-1)^{p+q} \frac{C}{\kappa} \left(\frac{s_j I}{s_{\text{low}}} \right)^{\varepsilon - 1} \left[\frac{(\kappa x)^\varepsilon}{\varepsilon} \sum_{k=0}^{p+q} \left(\frac{-1}{\varepsilon} \right)^{p+q-k} \frac{(p+q)!}{k!} (\ln(\kappa x))^k \right] \\
&\quad = \frac{C x^\varepsilon}{\varepsilon^{1+p+q}} \left(\frac{s_j I}{s_{\text{low}}} \cdot \kappa \right)^{\varepsilon - 1} \sum_{k=0}^{p+q} \frac{(p+q)!}{k!} (-\varepsilon \ln(\kappa x))^k \\
&\quad = \frac{C x^\varepsilon}{\varepsilon^{1+p+q}} \left(\frac{s_{\text{low}}}{\tau} \right)^{\varepsilon - 1} \sum_{k=0}^{p+q} \frac{(p+q)!}{k!} \left(\varepsilon \ln \left(\frac{\tau s_j I}{x s_{\text{low}}^2} \right) \right)^k .
\end{aligned} \tag{B.1}$$

Notably, the integrand is non-negative (this becomes evident in the proofs of Lemma 2.15 and Corollary 2.17). Therefore, the term from Equation (B.1) increases monotonically in x and attains its maximum on $\mathcal{I}$ at τ. $\square$

Proof of Lemma 2.15. We start by noting that ${}^{\times\mathrm{D}}\lambda_\theta^{(j)}(t)$ is linear in the argument θ_1. Consequently, the parameter θ_1 vanishes at differentiation and all partial derivatives where we differentiate twice or more with respect to θ_1 then amount to zero. Formally, we obtain[60]:

$$\theta_1 \frac{\partial}{\partial \theta_1} {}^{\times\mathrm{D}}\lambda_\theta^{(j)}(t) = {}^{\times\mathrm{D}}\lambda_\theta^{(j)}(t) , \tag{B.2}$$

$$\frac{\partial^2}{\partial \theta_1^2} {}^{\times\mathrm{D}}\lambda_\theta^{(j)}(t) = 0 . \tag{B.3}$$

To derive the bounds from Equation (2.34), we now consider the partial derivatives of

[60]One might be tempted to write $\frac{1}{\theta_1}$ on the right-hand side of Equation (B.2), but then problems with $\theta_1 = 0$ might arise.

arbitrary order with respect to θ_2 and θ_3. By writing

$$^{\times\mathrm{D}}\lambda_\theta^{(j)}(t) = \theta_1 \exp\left[\theta_2 \ln B_j(t) + \theta_3 \ln A_j(t)\right] \cdot \mathbb{1}_{\left\{N_{t^-}^{(j)} < C_j\right\} \cap \left\{t \leq \tau_j\right\}}$$

and applying the chain rule, we have for $p, q \in \mathbb{N}_0$:

$$\frac{\partial^{p+q}}{\partial\theta_2^p \partial\theta_3^q} {}^{\times\mathrm{D}}\lambda_\theta^{(j)}(t) = \left(\ln B_j(t)\right)^p \left(\ln A_j(t)\right)^q \cdot {}^{\times\mathrm{D}}\lambda_\theta^{(j)}(t)\,. \tag{B.4}$$

Recall that, for each fixed θ, ${}^{\times\mathrm{D}}\lambda_\theta^{(j)}(t)$ can be bounded by giving uniform bounds for $A_j(t)$ and $B_j(t)$ according to Lemma 2.12. More precisely, we have for all $t \in \mathcal{I}$ and $\omega \in \Omega$:

$$0 \leq A_j(t) \leq s_j I\,, \quad s_{\text{low}} \leq B_j(t) \leq s_j I\,, \quad j \in \mathbb{N}\,. \tag{B.5}$$

We now recall the representation from Equation (2.37) found in Remark 2.16, where the rescaled initial stress levels $\tilde{s}_j = \frac{s_j}{s_{\text{low}}}$ were introduced to ensure $\tilde{s}_j \geq 1$. This proves helpful in view of Equation (B.4), as it causes the upper bounds for the logarithms involved to be positive. Since $s_{\text{low}} \leq 1$ by assumption, we have $\tilde{s}_j \geq s_j$ and observe:

$$\underbrace{\ln(s_{\text{low}})}_{\leq 0} \leq \ln B_j(t) \leq \ln\left(s_j I\right) \leq \underbrace{\ln\left(\tilde{s}_j I\right)}_{\geq 0}, \tag{B.6}$$

so that by utilizing $\max\{a, b\} \leq a + b$ for $a, b \geq 0$, we get

$$\left|\ln B_j(t)\right| \leq \max\left\{-\ln\left(s_{\text{low}}\right), \ln\left(\tilde{s}_j I\right)\right\} \leq \ln\left(\tilde{s}_j I\right) - \ln\left(s_{\text{low}}\right) = \ln\left(\frac{s_j I}{s_{\text{low}}^2}\right). \tag{B.7}$$

While the calculations from Equation (B.6) can be performed in the same way with A_j instead of B_j to achieve identical results for the upper bound of $\ln\left(A_j(t)\right)$, this procedure fails to deliver a proper lower bound because of $A_j(0) = 0$ and hence $\ln\left(A_j(t)\right) \to -\infty$ for $t \downarrow 0$. However, closer inspection of $A_j(t)$ yields:

$$\ln A_j(t) = \ln\left(\frac{1}{\tau}\int_0^t s_j \underbrace{\frac{I}{I - N_{u^-}^{(j)}}}_{\geq 1}\, du\right) \geq \ln\left(s_j \frac{t}{\tau}\right) \geq \ln\left(s_{\text{low}} \frac{t}{\tau}\right), \tag{B.8}$$

so that the right-hand side of Equation (B.8) takes the place of $\ln(s_{\text{low}})$ in Equation (B.7) when A_j is considered instead of B_j:

$$\left|\ln\left(A_j(t)\right)\right| \leq \ln\left(\tilde{s}_j I\right) - \ln\left(s_{\text{low}} \frac{t}{\tau}\right) = \ln\left(\frac{s_j I}{s_{\text{low}}^2} \cdot \frac{\tau}{t}\right).$$

In tandem with Lemma 2.12, for which the requirement $\Theta \subset \mathbb{R}_+^3$ is needed[61], we then

[61] Technically, $\theta_2 \geq 0$ need not be assumed here because of the lower bound for $B_j(t)$ given in Equation (B.5), but the resulting model would no longer conform to our interpretation of a load sharing model. However, $\theta_1 < 0$ conflicts with the non-negativity of the intensity function, while $\theta_3 < 0$ leads to unbounded partial derivatives due to $A_j(0) = 0$.

obtain by virtue of $\tau \geq t$:

$$\begin{aligned}\left|\frac{\partial^{p+q}}{\partial\theta_2^p\partial\theta_3^q}{}^{\times\mathrm{D}}\lambda_\theta^{(j)}(t)\right| &= |\ln B_j(t)|^p\,|\ln A_j(t)|^q \cdot {}^{\times\mathrm{D}}\lambda_\theta^{(j)}(t)\\ &\leq \left[\ln\left(\frac{s_jI}{s_{\mathrm{low}}^2}\right)\right]^p\left[\ln\left(\frac{s_jI}{s_{\mathrm{low}}^2}\cdot\frac{\tau}{t}\right)\right]^q \cdot \theta_1\,(s_jI)^{\theta_2+\theta_3}\\ &\leq \theta_1\cdot\left[\ln\left(\frac{s_jI}{s_{\mathrm{low}}^2}\cdot\frac{\tau}{t}\right)\right]^{p+q}\cdot(s_jI)^{\theta_2+\theta_3}\,. \end{aligned}\tag{B.9}$$

From Equations (B.2) and (B.3), we get that the parameter θ_1 vanishes at differentiation with respect to θ_1. Thus,

$$\left|\frac{\partial^{p+q+1}}{\partial\theta_1\partial\theta_2^p\partial\theta_3^q}{}^{\times\mathrm{D}}\lambda_\theta^{(j)}(t)\right| \leq 1\cdot\left[\ln\left(\frac{s_jI}{s_{\mathrm{low}}^2}\cdot\frac{\tau}{t}\right)\right]^{p+q}\cdot(s_jI)^{\theta_2+\theta_3}\,, \tag{B.10}$$

$$\left|\frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}{}^{\times\mathrm{D}}\lambda_\theta^{(j)}(t)\right| = 0\,, \qquad r\in\mathbb{N}\setminus\{1\}\,. \tag{B.11}$$

Finally, we conclude by combining Equations (B.9) through (B.11):

$$\left|\frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}{}^{\times\mathrm{D}}\lambda_\theta^{(j)}(t)\right| \leq \max\{1,\theta_1\}\cdot\left[\ln\left(\frac{s_jI}{s_{\mathrm{low}}^2}\cdot\frac{\tau}{t}\right)\right]^{p+q}\cdot(s_jI)^{\theta_2+\theta_3}\,,$$

where $p,q,r\in\mathbb{N}_0$ and $j\in\mathbb{N}$, hence proving Equation (2.34). To show Equation (2.35) of Lemma 2.15, note that we can write

$$\begin{aligned}&\max\{1,\theta_1\}\cdot\left[\ln\left(\frac{s_jI}{s_{\mathrm{low}}^2}\cdot\frac{\tau}{t}\right)\right]^{p+q}\cdot(s_jI)^{\theta_2+\theta_3}\\ &\quad=\underbrace{\left[\ln\left(\frac{s_jI}{s_{\mathrm{low}}^2}\cdot\frac{\tau}{t}\right)\right]^{p+q}}_{=\text{const w.r.t. }\theta}\cdot\underbrace{\max\{1,\theta_1\}\cdot(s_jI)^{\theta_2+\theta_3}}_{=:C(\theta)}\\ &\quad=\text{const}\cdot C(\theta)\,,\end{aligned}$$

where C is a continuous function of $\theta=(\theta_1,\theta_2,\theta_3)$. Thereby, the image $C(\Theta)$ of any bounded parameter space Θ is also bounded. We suggestively denote this bound by $C(\Theta)$ and proceed to the proof of the remaining part of Lemma 2.15. In order to apply the measure theoretic version of the Leibniz integral rule (cf. Lemma 16.2 of Bauer 2001) and differentiate under the integral sign, the uniform majorant given by (2.35) needs to be integrable on $\mathcal{I}$ as a function of t. We set $\varepsilon=1$, $x=\tau$ and use Corollary B.1.1 to calculate the integral explicitly[62]:

$$\int_0^t \underbrace{\max\{1,\theta_1\}\cdot\left[\ln\left(\frac{s_jI}{s_{\mathrm{low}}^2}\cdot\frac{\tau}{u}\right)\right]^{p+q}\cdot(s_jI)^{\theta_2+\theta_3}}_{>0\text{ for all }u\in\mathcal{I}}\,\mathrm{d}u$$

[62] In the first step we transition from t to τ in order to derive a bound that is uniform w.r.t. t. This step can be omitted if such a bound is not desired.

$$
\begin{aligned}
&\leq \int_0^\tau \max\{1,\theta_1\} \cdot \left[\ln\left(\frac{s_j I}{s_{\mathrm{low}}^2}\cdot\frac{\tau}{u}\right)\right]^{p+q} \cdot (s_j I)^{\theta_2+\theta_3}\,\mathrm{d}u \\
&\leq C(\Theta)\int_0^\tau \left[\ln\left(\frac{s_j I}{s_{\mathrm{low}}^2}\cdot\frac{\tau}{u}\right)\right]^{p+q} \mathrm{d}u \\
&= C(\Theta)\tau \sum_{k=0}^{p+q} \frac{(p+q)!}{k!}\left(\ln\left(\frac{s_j I}{s_{\mathrm{low}}^2}\right)\right)^k < \infty\,. \qquad \text{(B.12)}
\end{aligned}
$$

By assumption, the sequence $(s_j)_{j\in\mathbb{N}}$ is bounded above by some constant s_{upp}. From this,

$$
\frac{s_{\mathrm{upp}} I}{s_{\mathrm{low}}^2} \geq \frac{s_j I}{s_{\mathrm{low}}^2} = \underbrace{\frac{s_j}{s_{\mathrm{low}}}}_{\geq 1}\cdot\underbrace{\frac{1}{s_{\mathrm{low}}}}_{\geq 1}\cdot I \geq 1\,,
$$

and hence the summands of Equation (B.12) are non-negative. From here, the triangle inequality yields the final statement of Equation (2.36):

$$
\begin{aligned}
\left|\frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}{}^{\times\mathrm{D}}\Lambda_\theta^{(j)}(t)\right| &= \left|\int_0^t \frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}{}^{\times\mathrm{D}}\lambda_\theta^{(j)}(u)\,\mathrm{d}u\right| \\
&\leq \int_0^t \max\{1,\theta_1\}\cdot\left[\ln\left(\frac{s_j I}{s_{\mathrm{low}}^2}\cdot\frac{\tau}{u}\right)\right]^{p+q}\cdot(s_j I)^{\theta_2+\theta_3}\,\mathrm{d}u \\
&\leq C(\Theta)\tau\sum_{k=0}^{p+q}\frac{(p+q)!}{k!}\left(\ln\left(\frac{s_j I}{s_{\mathrm{low}}^2}\right)\right)^k \\
&\leq C(\Theta)\tau\sum_{k=0}^{p+q}\frac{(p+q)!}{k!}\left(\ln\left(\frac{s_{\mathrm{upp}} I}{s_{\mathrm{low}}^2}\right)\right)^k < \infty\,,
\end{aligned}
$$

and hence completes the proof. □

Corollary 2.17 (Extension of Lemma 2.15)**.**
Let again ${}^{\times D}\lambda_\theta^{(j)}(t)$ be the conditional intensity function of the Basquin load sharing model with multiplicative damage accumulation given in Definition 2.8 of Subsection 2.3.2 and suppose that the sequence $(s_j)_{j\in\mathbb{N}}$ is bounded as in Lemma 2.15. Moreover, we assume that $\Theta\subset\mathbb{R}_+^2\times(-1,\infty)$. Then, for each $\theta\in\Theta$ with $-1<\theta_3<0$, the following holds for all $t\in\mathcal{I}$ and $\omega\in\Omega$:

$$
\left|\frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}{}^{\times D}\lambda_\theta^{(j)}(t)\right| \leq \max\{1,\theta_1\}\cdot\left[\ln\left(\frac{s_j I}{s_{\mathrm{low}}^2}\cdot\frac{\tau}{t}\right)\right]^{p+q}\cdot(s_j I)^{\theta_2}\left(s_{\mathrm{low}}\frac{t}{\tau}\right)^{\theta_3}, \quad p,q,r\in\mathbb{N}_0\,,j\in\mathbb{N}\,. \tag{2.38}
$$

If further $0<\varepsilon\leq 1$ exists so that $\Theta\subset\pi_3^{-1}([-1+\varepsilon,\infty))$ (i.e., the third parameter θ_3 is bounded away from -1) and Θ is bounded, then a constant C indepedent of $\theta\in\Theta$ and $j\in\mathbb{N}$ can be found such that, for all $\theta\in\Theta$,

$$
\left|\frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}{}^{\times D}\lambda_\theta^{(j)}(t)\right| \leq C\cdot\left[\ln\left(\frac{s_j I}{s_{\mathrm{low}}^2}\cdot\frac{\tau}{t}\right)\right]^{p+q}\cdot\left(s_{\mathrm{low}}\frac{t}{\tau}\right)^{\varepsilon-1}, \quad p,q,r\in\mathbb{N}_0\,,j\in\mathbb{N}\,. \tag{2.39}
$$

Under these assumptions, differentiation of arbitrary order with respect to $\theta \in \Theta$ and integration with respect to $t \in \mathcal{I}$ are also interchangeable. In addition, the following bound applies:

$$\left| \frac{\partial^{p+q+r}}{\partial\theta_1^r \partial\theta_2^p \partial\theta_3^q} {}^{\times D}\Lambda_\theta^{(j)}(t) \right| \leq \frac{C\tau s_{\mathrm{low}}^{\varepsilon-1}}{\varepsilon^{1+p+q}} \sum_{k=0}^{p+q} \frac{(p+q)!}{k!} \left(\varepsilon \ln \left(\frac{s_j I}{s_{\mathrm{low}}^2} \right) \right)^k , \tag{2.40}$$

where replacing s_j by s_{upp} yields a bound that is uniform with respect to $j \in \mathbb{N}$.

Proof. Let $-1 < \theta_3 < 0$. While the derivatives of the conditional intensity function from Equation (B.4) as well as the bounds for $\ln A_j(t)$ and $\ln B_j(t)$ remain unchanged, the uniform bounds for ${}^{\times \mathrm{D}}\lambda_\theta^{(j)}(t)$ provided by Lemma 2.12 are no longer valid here. However, similar to Equation (B.8), we observe

$$A_j(t) \geq s_{\mathrm{low}} \frac{t}{\tau} \quad \Longrightarrow \quad A_j(t)^{\theta_3} \leq \left(s_{\mathrm{low}} \frac{t}{\tau} \right)^{\theta_3} ,$$

and therefore we obtain, as before, that

$${}^{\times \mathrm{D}}\lambda_\theta^{(j)}(t) \leq \theta_1 \left(s_j I\right)^{\theta_2} \left(s_{\mathrm{low}} \frac{t}{\tau} \right)^{\theta_3} .$$

From here, Equation (2.38) follows immediately. If we suppose that Θ is bounded, the image of Θ under the continuous function $\max\{1, \theta_1\} \cdot (s_j I)^{\theta_2}$ is again bounded by some constant C'. Let C'' be the constant given in Equation (2.35) when $\Theta \cap \mathbb{R}_+^3$ is considered, and set $C := \max\{C', C''\}$. Moreover, $s_{\mathrm{low}} \leq 1$ implies that $s_{\mathrm{low}} \frac{t}{\tau} \leq 1$ and hence

$$\left(s_{\mathrm{low}} \frac{t}{\tau} \right)^{\theta_3} \leq \underbrace{\left(s_{\mathrm{low}} \frac{t}{\tau} \right)^{\varepsilon-1}}_{\geq 1, \text{ as } \varepsilon \leq 1} \tag{B.13}$$

as long as $\theta_3 > \varepsilon - 1$. With Equation (B.13) and bearing in mind the constant C above, Equation (2.39) is then a direct consequence of combining Equations (2.35) and (2.38); note that Equation (2.39) conforms with Equation (2.35) in the case $\varepsilon = 1$. To complete the proof, we only need to show that the right-hand side of Equation (2.39) is integrable. Like in Equation (B.12), this can be inferred from Corollary B.1.1, where $x = \tau$ is considered:

$$\begin{aligned} &\int_0^\tau C \cdot \left[\ln \left(\frac{s_j I}{s_{\mathrm{low}}^2} \cdot \frac{\tau}{u} \right) \right]^{p+q} \cdot \left(s_{\mathrm{low}} \frac{u}{\tau} \right)^{\varepsilon-1} \mathrm{d}u \\ &\quad = \frac{C\tau s_{\mathrm{low}}^{\varepsilon-1}}{\varepsilon^{1+p+q}} \sum_{k=0}^{p+q} \frac{(p+q)!}{k!} \left(\varepsilon \ln \left(\frac{s_j I}{s_{\mathrm{low}}^2} \right) \right)^k < \infty . \end{aligned}$$

Finally, this term also serves as a bound for the derivatives of the cumulative intensity function, compare Equation (2.36). This completes the proof of Corollary 2.17. $\square$

Glivenko-Cantelli Type Convergence Theorems for Derivative Cumulative Intensities

A standard result, based on a Glivenko-Cantelli argument[63], is that averages of cumulative intensities almost surely converge uniformly on compact intervals such as $\mathcal{I}$. However, the required monotonicity is often lost as soon as derivatives with respect to the model parameter θ are involved. In the particular case of the Basquin load sharing model with multiplicative damage accumulation, a related result can still be inferred from the Inequalities (2.39) and (2.40), namely, the almost sure uniform convergence of aggregated derivatives of the cumulative intensities (as functions of t for any fixed $\theta \in \Theta$) on the compact interval $\mathcal{I}$. Before formulating this corollary, we first provide insight into the much simpler Glivenko-Cantelli approach. For simplicity, we operate here only within the framework of Kopperschmidt and Stute from Section 3.

Lemma B.1.2 (Glivenko-Cantelli for Cumulative Intensities)**.**
In the framework of Definition 3.2 and Assumptions 3.5, for all $\theta \in \Theta$ holds almost surely:

$$\overline{\Lambda}_\theta^{(n)} \quad \longrightarrow \quad \mathbb{E}\Lambda_\theta \quad \textit{uniformly on } \mathcal{I} \textit{ as } n \to \infty.$$

Proof. The advantage of the Glivenko-Cantelli approach shown here is that it remains valid for (almost) arbitrary cumulative intensities and thus does not depend on the specific model. We do, however, require the continuity of the expected cumulative intensity, which can be inferred directly from Assumptions 3.5, see Remark 3.6 for details. If moreover $\mathbb{E}\left(\Lambda_\theta(\tau)\right) < \infty$ holds for each $\theta \in \Theta$ (e.g., due to condition ($\widetilde{\mathrm{C}}3$)), then for any given $\varepsilon > 0$ we can choose $0 = t_0 < t_1 < \ldots < t_m = \tau$ such that

$$\left|\mathbb{E}\left(\Lambda_\theta(t_{i+1})\right) - \mathbb{E}\left(\Lambda_\theta(t_i)\right)\right| < \frac{\varepsilon}{2}\,, \quad \text{for all } i = 0, 1, \ldots, m-1\,. \tag{B.14}$$

By the strong law of large numbers, for all $i = 0, 1, \ldots, m-1$ we have with probability 1:

$$\overline{\Lambda}_\theta^{(n)}(t_i) \longrightarrow \mathbb{E}\left(\Lambda_\theta(t_i)\right) \quad \text{as } n \to \infty\,.$$

Hence, there exists $n_0 \in \mathbb{N}$ such that $\mathbb{P}_{\theta^*}$-almost surely for all $n \geq n_0$ simultaneously holds (note that m is finite):

$$\left|\overline{\Lambda}_\theta^{(n)}(t_i) - \mathbb{E}\left(\Lambda_\theta(t_i)\right)\right| < \frac{\varepsilon}{2}\,, \quad \text{for all } i = 0, 1, \ldots, m-1\,. \tag{B.15}$$

The monotonicity of the aggregate process $\overline{\Lambda}_\theta^{(n)}$ yields - again $\mathbb{P}_{\theta^*}$-almost surely - that

$$\overline{\Lambda}_\theta^{(n)}(t) \in \left[\overline{\Lambda}_\theta^{(n)}(t_i),\, \overline{\Lambda}_\theta^{(n)}(t_{i+1})\right]\,, \quad \text{for all } t \in [t_i, t_{i+1}]\,, \tag{B.16}$$

from where combining Equations (B.15) and (B.16) in a first step and then exploiting Equation (B.14) as well as the monotonicity of the expected cumulative intensity in a second step grants

$$\overline{\Lambda}_\theta^{(n)}(t) \in \left(\mathbb{E}\left(\Lambda_\theta(t_i)\right) - \frac{\varepsilon}{2},\, \mathbb{E}\left(\Lambda_\theta(t_{i+1})\right) + \frac{\varepsilon}{2}\right)$$

[63] Named after the Glivenko-Cantelli theorem on the uniform convergence of the empirical distribution function, see Klenke 2020, p. 129.

$$\subset \left(\mathbb{E}\left(\Lambda_\theta(t)\right) - \varepsilon,\, \mathbb{E}\left(\Lambda_\theta(t)\right) + \varepsilon\right) \quad \text{for all } t \in [t_i, t_{i+1}]\ .$$

Therefore, with probability 1 we obtain for all $n \geq n_0$ that

$$\sup_{t \in \mathcal{I}} \left| \overline{\Lambda}_\theta^{(n)}(t) - \mathbb{E}\left(\Lambda_\theta(t)\right) \right| \leq \max_{i \in \{0,1,\ldots,m-1\}} \sup_{t \in [t_i, t_{i+1}]} \left| \overline{\Lambda}_\theta^{(n)}(t) - \mathbb{E}\left(\Lambda_\theta(t)\right) \right| < \varepsilon\,,$$

and thus $\mathbb{P}_{\theta^*}$-almost surely:

$$\left\| \overline{\Lambda}_\theta^{(n)} - \mathbb{E}\Lambda_\theta \right\|_\infty \longrightarrow 0 \quad \text{as } n \to \infty\,.$$

□

The Glivenko-Cantelli approach can also be applied to the averaged counting process, as explained in the following remark.

Remark B.1.3 (Further Implications of Lemma B.1.2)**.**
The Glivenko-Cantelli argument in the proof of Lemma B.1.2 implies that

$$\left\| \overline{N}^{(n)} - \mathbb{E}\Lambda_{\theta^*} \right\|_\infty \longrightarrow 0 \quad \text{as } n \to \infty$$

holds $\mathbb{P}_{\theta^*}$-almost surely as well. For this, consider the case $\theta = \theta^*$ and replace each instance of $\overline{\Lambda}_{\theta^*}^{(n)}$ by $\overline{N}^{(n)}$. Since $\mathbb{E}\overline{N}^{(n)} = \mathbb{E}\overline{\Lambda}_{\theta^*}^{(n)} = \mathbb{E}\Lambda_{\theta^*}$ according to the Doob-Meyer decomposition and by the monotonicity of the aggregated counting process, all proof steps can easily be adopted.
Moreover, it follows that, with probability 1,

$$\left\| \overline{M}^{(n)} \right\|_\infty = \left\| \overline{N}^{(n)} - \overline{\Lambda}_{\theta^*}^{(n)} \right\|_\infty \leq \left\| \overline{N}^{(n)} - \mathbb{E}\Lambda_{\theta^*} \right\|_\infty + \left\| \overline{\Lambda}_{\theta^*}^{(n)} - \mathbb{E}\Lambda_{\theta^*} \right\|_\infty \longrightarrow 0 \quad \text{as } n \to \infty\,.$$

We now extend Lemma B.1.2 to include derivatives with respect to the parameter θ. However, additional requirements must be placed on the model for this to work and we therefore consider only the Basquin load sharing model with damage accumulation. It is worth mentioning that Kopperschmidt and Stute claim that this convergence holds without further restrictions "for averages of derivative processes" (see Kopperschmidt and Stute 2013, p. 1293), although their proposed solution fails since Glivenko-Cantelli arguments cannot be applied here. The following corollary provides a stronger version of Lemma 3.16 from Subsection 3.3.

Corollary B.1.4 (Uniform Convergence of Averages of Derivative Cumulative Intensities in the Basquin Load Sharing Model with Damage Accumulation)**.**
In the situation of Corollary 2.17, for all $\theta \in \Theta$ and $p, q, r \in \mathbb{N}_0$ holds with probability 1*:*

$$\frac{1}{n}\sum_{j=1}^{n} \frac{\partial^{p+q+r}}{\partial\theta_1^r \partial\theta_2^p \partial\theta_3^q} {}^{\times\mathrm{D}}\Lambda_\theta^{(j)} \quad \longrightarrow \quad \mathbb{E}\frac{\partial^{p+q+r}}{\partial\theta_1^r \partial\theta_2^p \partial\theta_3^q} {}^{\times\mathrm{D}}\Lambda_\theta \quad \textit{uniformly on } \mathcal{I} \textit{ as } n \to \infty.$$

In particular, for all $p \in \mathbb{N}_0$, we have $\mathbb{P}_{\theta^*}$-almost surely that

$$\frac{\mathrm{d}^p}{\mathrm{d}\theta^p} \overline{{}^{\times\mathrm{D}}\Lambda}_\theta^{(n)} \quad \longrightarrow \quad \mathbb{E}\frac{\mathrm{d}^p}{\mathrm{d}\theta^p} {}^{\times\mathrm{D}}\Lambda_\theta \quad \textit{uniformly on } \mathcal{I} \textit{ as } n \to \infty. \tag{B.17}$$

Proof. In order to shorten the notation, we will omit the model identifier ${}^{\times\mathrm{D}}$ for the time being. The special case $p = 0$ of Equation (B.17) (i.e., in the absence of any derivatives) was shown in Lemma B.1.2. For the general case, accounting also for differentiation with respect to θ, we will show the equicontinuity of the aggregate partial derivatives. Let $x, y \in \mathcal{I}$ and without loss of generality assume $x \leq y$. Then, for all $p, q, r \in \mathbb{N}_0$ and $j \in \mathbb{N}$ we observe

$$\begin{aligned}\left|\frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}\,\Lambda_\theta^{(j)}(y) - \frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}\,\Lambda_\theta^{(j)}(x)\right| &= \left|\int_x^y \frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}\,\lambda_\theta^{(j)}(u)\,\mathrm{d}u\right| \\ &\leq \sup_{u\in(x,y)}\left|\frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}\,\lambda_\theta^{(j)}(u)\right|\cdot|y-x|\ ,\end{aligned}$$

and plugging in the upper bound from Equation (2.39) yields for $x > 0$:

$$\begin{aligned}\sup_{u\in(x,y)}\left|\frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}\,\lambda_\theta^{(j)}(u)\right| &\leq \sup_{u\in(x,y)} C\cdot\underbrace{\left[\ln\left(\frac{s_jI}{s_{\text{low}}^2}\cdot\frac{\tau}{u}\right)\right]^{p+q}}_{\text{decreasing in } u}\cdot\underbrace{\left(s_{\text{low}}\frac{u}{\tau}\right)^{\varepsilon-1}}_{\text{decreasing in } u} \\ &\leq C\cdot\left[\ln\left(\frac{s_jI}{s_{\text{low}}^2}\cdot\frac{\tau}{x}\right)\right]^{p+q}\cdot\left(s_{\text{low}}\frac{x}{\tau}\right)^{\varepsilon-1} \\ &\leq C\cdot\left[\ln\left(\frac{s_{\text{upp}}I}{s_{\text{low}}^2}\cdot\frac{\tau}{x}\right)\right]^{p+q}\cdot\left(s_{\text{low}}\frac{x}{\tau}\right)^{\varepsilon-1} =: L(p,q,x) < \infty\,.\end{aligned}$$

Since $L(p, q, x)$ does not depend on j, we obtain by virtue of the triangle inequality:

$$\begin{aligned}&\left|\frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}\,\overline{\Lambda}_\theta^{(n)}(y) - \frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}\,\overline{\Lambda}_\theta^{(n)}(x)\right| \\ &\quad\leq \frac{1}{n}\sum_{j=1}^n\left|\frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}\,\Lambda_\theta^{(j)}(y) - \frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}\,\Lambda_\theta^{(j)}(x)\right| \\ &\quad\leq L(p,q,x)\cdot|y-x|\ .\end{aligned}\tag{B.18}$$

In order to prove (uniform) equicontinuity, we need to show that for each $\varepsilon > 0$ there exists $\delta > 0$ such that

$$\left|\frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}\,\overline{\Lambda}_\theta^{(n)}(y) - \frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}\,\overline{\Lambda}_\theta^{(n)}(x)\right| < \varepsilon \tag{B.19}$$

whenever $|x - y| < \delta$, $x, y \in [0, \tau]$. This seemingly follows from Equation (B.18), but problems arise in the neighbourhood of 0, where the local Lipschitz constant $L(p, q, x)$ tends to ∞ and therefore cannot be bounded. However, (pointwise) equicontinuity can still be ensured according to Corollary 2.17, as we have by virtue of Corollary B.1.1:

$$\begin{aligned}\left|\frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}\Lambda_\theta^{(j)}(x)\right| &= \left|\int_0^x \frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}\lambda_\theta^{(j)}(u)\,\mathrm{d}u\right| \leq \int_0^x\left|\frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}\lambda_\theta^{(j)}(u)\right|\mathrm{d}u \\ &\leq \int_0^x C\cdot\left[\ln\left(\frac{s_jI}{s_{\text{low}}^2}\cdot\frac{\tau}{u}\right)\right]^{p+q}\cdot\left(s_{\text{low}}\frac{u}{\tau}\right)^{\varepsilon-1}\mathrm{d}u \qquad \Big|\text{apply Cor. B.1.1}\end{aligned}$$

$$\leq \frac{Cx^{\varepsilon}}{\varepsilon^{1+p+q}} \left(\frac{s_{\text{low}}}{\tau}\right)^{\varepsilon-1} \sum_{k=0}^{p+q} \frac{(p+q)!}{k!} \left(\varepsilon \ln\left(\frac{\tau s_j I}{x s_{\text{low}}^2}\right)\right)^k$$

$$\leq \underbrace{\frac{Cx^{\varepsilon}}{\varepsilon^{1+p+q}} \left(\frac{s_{\text{low}}}{\tau}\right)^{\varepsilon-1} \sum_{k=0}^{p+q} \frac{(p+q)!}{k!} \left(\varepsilon \ln\left(\frac{\tau s_{\text{upp}} I}{x s_{\text{low}}^2}\right)\right)^k}_{\text{independent of } j} \to 0 \quad \text{as} \quad x \to 0\,,$$

due to the factor x again dominating the natural powers of the logarithm, compare Lemma 2.14. Consequently, we can find - with a reasoning analogous to the above - some $\delta > 0$ so that

$$\left|\frac{\partial^{p+q+r}}{\partial\theta_1^r \partial\theta_2^p \partial\theta_3^q} \overline{\Lambda}_{\theta}^{(n)}(x)\right| < \frac{\varepsilon}{2}\,, \quad \text{whenever } |x| < \delta,\, x \in [0,\tau]\,. \tag{B.20}$$

Now let

$$L := \sup_{x \in \left[\frac{\delta}{2}, \tau\right]} L(p,q,x) < \infty\,, \qquad \text{and choose} \quad 0 < \tilde{\delta} < \min\left\{\frac{\delta}{2}, \frac{\varepsilon}{L}\right\}\,.$$

Then, for any $x, y \in \left[\frac{\delta}{2}, \tau\right]$ with $|x - y| < \tilde{\delta}$, we have by Equation (B.18):

$$\left|\frac{\partial^{p+q+r}}{\partial\theta_1^r \partial\theta_2^p \partial\theta_3^q} \overline{\Lambda}_{\theta}^{(n)}(y) - \frac{\partial^{p+q+r}}{\partial\theta_1^r \partial\theta_2^p \partial\theta_3^q} \overline{\Lambda}_{\theta}^{(n)}(x)\right| \leq L \cdot |y - x| < \varepsilon\,.$$

If $|x - y| < \tilde{\delta}$ but $|x| < \frac{\delta}{2}$, then $|y| \leq |x| + |y - x| < \delta$. Hence,

$$\begin{aligned}
&\left|\frac{\partial^{p+q+r}}{\partial\theta_1^r \partial\theta_2^p \partial\theta_3^q} \overline{\Lambda}_{\theta}^{(n)}(y) - \frac{\partial^{p+q+r}}{\partial\theta_1^r \partial\theta_2^p \partial\theta_3^q} \overline{\Lambda}_{\theta}^{(n)}(x)\right| \\
&\quad \leq \left|\frac{\partial^{p+q+r}}{\partial\theta_1^r \partial\theta_2^p \partial\theta_3^q} \overline{\Lambda}_{\theta}^{(n)}(x)\right| + \left|\frac{\partial^{p+q+r}}{\partial\theta_1^r \partial\theta_2^p \partial\theta_3^q} \overline{\Lambda}_{\theta}^{(n)}(y)\right| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon
\end{aligned}$$

by Equation (B.20). The case $|y| < \frac{\delta}{2}$ gives the same result for symmetry reasons and thus uniform equicontinuity ensues. Importantly, δ does also not depend on the particular realization and since $[0,\tau]$ is compact, the $\mathbb{P}_{\theta^*}$-almost sure pointwise convergence of the aggregate partial derivatives (again due to the strong law of large numbers) then implies their uniform convergence. The proof of this conclusion is simple:
Let $\varepsilon > 0$, so that by virtue of the equicontinuity there exists $\delta > 0$ such that Equation (B.19) holds whenever $|x - y| < \delta$ (this is even deterministically true). As $[0,\tau]$ is compact, there exist some $k \in \mathbb{N}$ and $x_1, \ldots, x_k \in [0,\tau]$ with

$$[0,\tau] \subset \mathrm{B}_{\delta}(x_1) \cup \ldots \cup \mathrm{B}_{\delta}(x_k)\,.$$

The strong law of large numbers provides $\mathbb{P}_{\theta^*}$-almost surely for $j \in \{1,\ldots,k\}$:

$$\frac{\partial^{p+q+r}}{\partial\theta_1^r \partial\theta_2^p \partial\theta_3^q} \overline{\Lambda}_{\theta}^{(n)}(x_j) \quad \longrightarrow \quad \mathbb{E} \frac{\partial^{p+q+r}}{\partial\theta_1^r \partial\theta_2^p \partial\theta_3^q} \Lambda_{\theta}(x_j) \quad \text{as } n \to \infty,$$

and thus

$$\mathbb{P}_{\theta^*}\left(\limsup_{n\to\infty}\left\{\left|\frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}\overline{\Lambda}_\theta^{(n)}\left(x_j\right)-\mathbb{E}\frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}\Lambda_\theta\left(x_j\right)\right|>\varepsilon\right\}\right)=0\,.$$

By Lemma 4.18, this also holds simultaneously for all $j\in\{1,\ldots,k\}$, that is:

$$\begin{aligned}
&\mathbb{P}_{\theta^*}\left(\limsup_{n\to\infty}\left\{\max_{j=1,\ldots,k}\left|\frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}\overline{\Lambda}_\theta^{(n)}\left(x_j\right)-\mathbb{E}\frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}\Lambda_\theta\left(x_j\right)\right|>\varepsilon\right\}\right)\\
&\quad=\mathbb{P}_{\theta^*}\left(\limsup_{n\to\infty}\bigcup_{j=1}^{k}\left\{\left|\frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}\overline{\Lambda}_\theta^{(n)}\left(x_j\right)-\mathbb{E}\frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}\Lambda_\theta\left(x_j\right)\right|>\varepsilon\right\}\right)=0
\end{aligned}$$

Finally, since the uniform continuity implied by Equation (B.19) carries over to the limit function, we can compute:

$$\begin{aligned}
&\mathbb{P}_{\theta^*}\left(\limsup_{n\to\infty}\left\{\sup_{x\in[0,\tau]}\left|\frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}\overline{\Lambda}_\theta^{(n)}\left(x\right)-\mathbb{E}\frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}\Lambda_\theta\left(x\right)\right|>3\varepsilon\right\}\right)\\
&\quad=\mathbb{P}_{\theta^*}\left(\limsup_{n\to\infty}\left\{\max_{j=1,\ldots,k}\sup_{x\in\mathrm{B}_\delta(x_j)}\left|\frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}\overline{\Lambda}_\theta^{(n)}\left(x\right)-\mathbb{E}\frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}\Lambda_\theta\left(x\right)\right|>3\varepsilon\right\}\right)\\
&\quad\leq\mathbb{P}_{\theta^*}\left(\limsup_{n\to\infty}\left\{\max_{j=1,\ldots,k}\underbrace{\sup_{x\in\mathrm{B}_\delta(x_j)}\left|\frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}\overline{\Lambda}_\theta^{(n)}\left(x\right)-\frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}\overline{\Lambda}_\theta^{(n)}\left(x_j\right)\right|}_{<\,\varepsilon\text{ by equicontinuity and the choice of }\delta.}>\varepsilon\right\}\right)\\
&\qquad+\mathbb{P}_{\theta^*}\left(\limsup_{n\to\infty}\left\{\max_{j=1,\ldots,k}\cancel{\sup_{x\in\mathrm{B}_\delta(x_j)}}\left|\frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}\overline{\Lambda}_\theta^{(n)}\left(x_j\right)-\mathbb{E}\frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}\Lambda_\theta\left(x_j\right)\right|>\varepsilon\right\}\right)\\
&\qquad+\mathbb{P}_{\theta^*}\left(\limsup_{n\to\infty}\left\{\max_{j=1,\ldots,k}\sup_{x\in\mathrm{B}_\delta(x_j)}\left|\mathbb{E}\frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}\Lambda_\theta\left(x_j\right)-\mathbb{E}\frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}\Lambda_\theta\left(x\right)\right|>\varepsilon\right\}\right)\\
&\quad=\mathbb{P}_{\theta^*}\left(\limsup_{n\to\infty}\left\{\max_{j=1,\ldots,k}\left|\frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}\overline{\Lambda}_\theta^{(n)}\left(x_j\right)-\mathbb{E}\frac{\partial^{p+q+r}}{\partial\theta_1^r\partial\theta_2^p\partial\theta_3^q}\Lambda_\theta\left(x_j\right)\right|>\varepsilon\right\}\right)=0\,,
\end{aligned}$$

and this is $\mathbb{P}_{\theta^*}$-almost sure uniform convergence on $\mathcal{I}=[0,\tau]$, hence finishing the proof. □

B.2. Tightness on the Space $C(K)$

In the statistical analysis of counting processes and derived quantities like estimators and hypothesis tests, statisticians frequently face the challenge that the underlying distributions remain unknown, or at least that their computation is hardly feasible. The same then holds true for related characteristics of interest: One can easily imagine having a sequence of estimators that depend on the number of repetitions of an experiment and whose standard deviation is unknown. Under these circumstances, it is often helpful to resort to the asymptotic properties of the random variables involved - in this case, the sequence of estimators. The research on the asymptotics of such sequences of random variables or measures is itself a broad field whose best-known results have found their way into basic statistics courses. As they lend themselves to the derivation of asymptotic distributions, we will encounter some of these results (such as the central limit theorem) throughout this thesis. However, before answering the question of an asymptotic distribution, one has of course to establish whether a sequence of probability measures converges at all to some measure, or has at least one convergent subsequence, since the space of distribution functions is not compact (cf. Shiryaev 2016, pp. 383–384). Hence, one introduces the term *relative compactness*: a family Π of probability measures is called relatively compact if every sequence of elements from Π contains a convergent subsequence whose limit is not necessarily an element of Π (see Billingsley 1968, pp. 35–37 for a profound discussion). So, when we consider the space of probability measures[64] on a metric space endowed with its Borel σ-algebra, a characterization of its (relatively) compact subsets is crucial to achieve convenient asymptotic results. In the case of a complete and separable metric space, the theorem of Prohorov (a whole section is devoted to this fundamental result in Billingsley 1968, pp. 35–41) relates relative compactness to the notion of tightness. It offers therein an entry point to the subject, which (largely) omits the topological foundations. We can thus dive straight into the topic by stating the definition of tightness for a family of probability measures and defer further motivation for the time being.

Definition B.2.1 (Tight Family of Probability Measures; Billingsley 1968, p. 37)**.**
Let S be a metric space with metric d and let $\mathcal{B}_d$ denote the Borel σ-algebra on S induced by d. A family Π of probability measures on $(S, \mathcal{B}_d)$ is said to be tight *if for every $\varepsilon > 0$, there exists a compact set $K = K(\varepsilon) \in \mathcal{B}_d$ such that*

$$\mathbb{P}(K) > 1 - \varepsilon\,, \quad \textit{for all}\ \mathbb{P} \in \Pi\,.$$

In the aforementioned practice of statistical analysis, we often deal with sequences of random variables taking values in a metric space. Since the associated image measures form a family of probability measures, we can extend the notion of tightness to such sequences.

Definition B.2.2 (Tight Sequences of Random Variables; Billingsley 1968, p. 57)**.**
In the situation of Definition B.2.1, let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables taking values in the measurable space $(S, \mathcal{B}_d)$. The sequence $(X_n)_{n\in\mathbb{N}}$ is said to be tight*, if the corresponding family of image measures $\Pi = \{\mathbb{P}^{X_n} : n \in \mathbb{N}\}$ is tight, that is, if for all $\varepsilon > 0$, there exists a compact set K such that*

$$\mathbb{P}^{X_n}(K) = \mathbb{P}(X_n \in K) > 1 - \varepsilon\,, \quad \textit{for all}\ n \in \mathbb{N}.$$

In the following, we lay special emphasis on the space of continuous functions defined on a compact topological space and mapping into a normed vector space, which will take the role of the metric space $(S, \mathcal{B}_\mathrm{d})$ in the Definitions B.2.1 and B.2.2.

Definition B.2.3 (Banach Space of Continuous Functions $C(\mathcal{X}, \mathcal{Y})$; cf. Dudley 2002, pp. 51–53)**.**
For any compact topological space $\mathcal{X}$ and normed vector space $\mathcal{Y}$ with norm $\|\cdot\|$, let $C(\mathcal{X}, \mathcal{Y})$ be the space of all $\mathcal{Y}$-valued continuous functions on $\mathcal{X}$. The supremum norm $\|\cdot\|_\infty$ given by

$$\|f\|_\infty = \sup_{x\in\mathcal{X}} \|f(x)\|\,, \quad f \in C(\mathcal{X}, \mathcal{Y})\,,$$

defines a norm on $C(\mathcal{X}, \mathcal{Y})$ with respect to which $C(\mathcal{X}, \mathcal{Y})$ becomes a Banach space.

[64] This space can be topologized by the Prohorov metric, see Ethier and Kurtz 1986, pp. 96–103. Since we obtain a metric space this way, the notions of compactness and sequential compactness are equivalent, so that compactness ensures the existence of at least one convergent subsequence.

The reason for focusing on the space $C(\mathcal{X}, \mathcal{Y})$ resides in the following observation: Any continuous[65] stochastic process with index set $\mathcal{X}$ and state space $\mathcal{Y}$ can be seen as a $C(\mathcal{X}, \mathcal{Y})$-valued random variable, so that Definition B.2.2 allows us to relate the property of tightness to sequences of continuous stochastic processes. As we are primarily interested in continuous-parameter processes, we naturally restrict the index set $\mathcal{X}$ to be a compact subset of $\mathbb{R}^d$ for some $d \in \mathbb{N}$ and write $\mathcal{X} = K \subset \mathbb{R}^d$. Moreover, we consider only the case $\mathcal{Y} = \mathbb{R}^{\tilde{d}}$, with the specific dimension $\tilde{d}$ being largely irrelevant. We therefore use the abbreviated notation $C(K)$ in place of $C(K, \mathbb{R}^{\tilde{d}})$ and neglect the image space whenever it is evident from the context. Note that $C(K)$ is complete by default as a Banach space, and due to the Weierstrass approximation theorem - a consequence of the more general Stone-Weierstrass theorem - it is also separable (see Dudley 2002, p. 54 for details). Hence, Prohorov's theorem is applicable, so that a characterization of tightness on $C(K)$ simultaneously deals with relative compactness. A major benefit of this equivalence lies in the fact that such a characterization requires knowledge only of the compact subsets of the space $C(K)$ itself (as opposed to the relatively compact subsets of the space of probability measures on $C(K)$). The theorem of Arzelà-Ascoli provides that knowledge and thus forms the basis of the following theorem.

Theorem B.2.4 (Tight Family of Probability Measures on $C(K)$; Billingsley 1968, p. 55)**.**
Let $\Pi = \{\mathbb{P}_n : n \in \mathbb{N}\}$ *be a sequence of probability measures on the Banach space* $C(K)$*. The family* Π *is tight if and only if these two conditions hold:*

(i) There exists a $\theta_0 \in K$ *such that for each* $\eta > 0$*, there exists an* $a > 0$ *with*

$$\mathbb{P}_n\left(\{f \,:\, \|f(\theta_0)\| > a\}\right) \leq \eta\,, \quad \textit{for all}\; n \in \mathbb{N}\,.$$

(ii) For each $\eta > 0$ *and* $\varepsilon > 0$*, there exists a* $\delta > 0$ *and an* $n_0 \in \mathbb{N}$ *such that*

$$\mathbb{P}_n\left(\left\{f \,:\, \sup_{\|x-y\|<\delta} \|f(x) - f(y)\| \geq \varepsilon\right\}\right) \leq \eta\,, \quad \textit{for all}\; n \geq n_0\,.$$

Proof. The proof comes as a direct application of Arzelà-Ascoli's theorem and can be found in Billingsley 1968, p. 55 for the special case $K = [0,1] \subset \mathbb{R}$. As a side result, one finds that condition (i) must hold for arbitrary $\theta_0 \in K$, which is why only $\theta_0 = 0$ is considered in the proof quoted above[66]. □

Applying Theorem B.2.4 to a sequence of image measures like in the Definition B.2.2 yields a useful corollary.

Corollary B.2.5 (Tight Sequences of $C(K)$-Valued Random Variables; Billingsley 1968, p. 58)**.**
A sequence $(X_n)_{n\in\mathbb{N}}$ *of random variables taking values in the space* $C(K)$ *is tight if and only if these two conditions hold:*

(i) There exists a $\theta_0 \in K$ *such that for each* $\eta > 0$*, there exists an* $a > 0$ *with*

$$\mathbb{P}\left(\|X_n(\theta_0)\| > a\right) \leq \eta\,, \quad \textit{for all}\; n \in \mathbb{N}\,. \tag{B.21}$$

[65] Formally, Definition A.2 only covers the case where the index set $\mathcal{I}$ is an interval, but can be easily extended to general topological spaces.

[66] Although this observation is of less interest to us, condition (i) simply states that the sequence $\{\mathbb{P}_n \circ \pi_{\theta_0}^{-1} : n \in \mathbb{N}\}$ of probability measures on $\mathcal{B}(\mathbb{R})$ is tight, where $\pi_{\theta_0}(f) = f(\theta_0)$.

(ii) For each $\eta > 0$ and $\varepsilon > 0$, there exists a $\delta > 0$ and an $n_0 \in \mathbb{N}$ such that

$$\mathbb{P}\left(\sup_{\|x-y\|<\delta} \|X_n(x) - X_n(y)\| \geq \varepsilon\right) \leq \eta\,, \quad \textit{for all } n \geq n_0\,. \tag{B.22}$$

Corollary B.2.5 stipulates that in order to be tight, the continuous stochastic processes X_n need to - in a probabilistic sense - be locally bounded and not oscillate too violently (cf. Billingsley 1968, p. 58). With this practical interpretation in mind, a reasonable intent to adduce the concept of tightness can be illustrated by asking (and partially answering) the following question:
Suppose that we are given a sequence $(X_n)_{n\in\mathbb{N}}$ of continuous stochastic processes (i.e., random variables taking values in the space $C(K)$ for some compact subset $K \subset \mathbb{R}^d$). Moreover, we assume that we have a sequence $(\theta_n)_{n\in\mathbb{N}}$ of random variables converging in K towards θ_0. Under what conditions can we then conclude that, in some sense[67],

$$X_n(\theta_n) - X_n(\theta_0) \longrightarrow 0 \quad (n \to \infty) \tag{B.23}$$

holds? Since $\theta_n \to \theta_0$ by assumption, it should suffice - heuristically speaking - if the random functions X_n do not oscillate excessively. However, this heuristic coincides precisely with the above interpretation of Equation (B.22). It is therefore no surprise that the tightness of the sequence $(X_n)_{n\in\mathbb{N}}$ would in fact ensure the convergence (in probability) from Equation (B.23), at least given that $\theta_n \to \theta_0$ almost surely as $n \to \infty$. Indeed, the property of tightness seems unnecessarily potent and we often conclude that proving tightness is more challenging than directly validating the above convergence, as seen in Subsection 3.3. To address this concern, one often resorts to the easily verifiable tightness criterion of Kolmogorov, that we formulate here for the case where $K = [0,1]^d$ is the d-dimensional unit cube.

Theorem B.2.6 (Kolmogorov's Tightness Criterion; cf. Billingsley 1968, p. 95)**.**
The sequence $(X_n)_{n\in\mathbb{N}}$ of random variables taking values in the space $C\big([0,1]^d\big)$ is tight if it satisfies these two conditions:

(i) There exists a $\theta_0 \in [0,1]^d$ such that for each $\eta > 0$, there exists an $a > 0$ with

$$\mathbb{P}\left(\|X_n(\theta_0)\| > a\right) \leq \eta\,, \quad \textit{for all } n \in \mathbb{N}\,. \tag{B.24}$$

(ii) There exist constants $\nu > 0$ and $\alpha > 0$ such that

$$\mathbb{E}\,\|X_n(x) - X_n(y)\|^2 \leq \nu\,\|x-y\|^{d+\alpha}\,, \quad \textit{for all } x, y \in [0,1]^d \textit{ and } n \in \mathbb{N}\,. \tag{B.25}$$

Proof. The one-dimensional case $d = 1$ is contained in Theorem 12.3 of Billingsley 1968, p. 95, where one chooses $\gamma = 2$ and $F = id$. The cited source also serves as the main reference for tightness tools and techniques in the article Kopperschmidt and Stute 2013, p. 1289. While they also refer to Bickel and Wichura 1971 for multiparameter processes, we will later find that their condition is not sufficient to derive a moment condition independent of the dimension d (cf. the exponent from Equation (B.25)). We will instead draw on the multi-dimensional generalization from Kunita 1990, p. 38. Adapted to our

[67] Note that we have not yet specified the type of convergence, neither for the sequence $(\theta_n)_{n\in\mathbb{N}}$, nor the following Equation (B.23).

situation, their associated Theorem 1.4.7 replaces conditions (i) and (ii) above with the following:

(i*) There exists a constant $\eta > 0$ such that

$$\mathbb{E}\,\|X_n(x)\|^2 \leq \eta\,, \quad \text{for all } x \in [0,1]^d \text{ and } n \in \mathbb{N}\,.$$

(ii*) There exist constants $\nu > 0$ and $\alpha_1, \ldots, \alpha_d > 0$ with $\sum_{i=1}^d \alpha_i^{-1} < 1$ such that

$$\mathbb{E}\,\|X_n(x) - X_n(y)\|^2 \leq \nu \left(\sum_{i=1}^d |x_i - y_i|^{\alpha_i} \right)\,, \quad \text{for all } x, y \in [0,1]^d \text{ and } n \in \mathbb{N}\,.$$

If in condition (ii*) we choose $\alpha_i = d + \alpha$ for $i = 1, \ldots, d$, then

$$\nu \left(\sum_{i=1}^d |x_i - y_i|^{\alpha_i} \right) = \nu\,\|x - y\|_{d+\alpha}^{d+\alpha}\,,$$

so condition (ii) implies (ii*) by the equivalence of norms on $\mathbb{R}^d$. Furthermore, condition (i*) is only used to obtain condition (i) by virtue of the Markov inequality. Overall, the statement from Theorem 1.4.7 therefore holds even if conditions (i) and (ii) are used instead. The full proof can be found in Kunita 1990, pp. 38–39. We refrain from reproducing their proof, because it is rather technical and requires a variant of Kolmogorov's continuity criterion (cf. Kunita 1990, pp. 31–35), which lends its name to the tightness criterion we consider here. Additionally, the techniques used revolve around Hölder continuity, which is of no further relevance to us. □

The primary purpose of considering tightness was to determine conditions under which the convergence in Equation (B.23) could be guaranteed. With Kolmogorov's tightness criterion in mind, we now conclude the subsection by stating that this convergence basically holds by virtue of a straightforward moment condition and leave behind the more abstract notion of tightness in the process.

Lemma B.2.7.
Let $(X_n)_{n\in\mathbb{N}}$ be a tight sequence of random variables taking values in the space $C(K)$. Furthermore, let $(\theta_n)_{n\in\mathbb{N}} \subset K$ be another sequence of random variables that converges $\mathbb{P}$-almost surely to $\theta_0 \in K$. Then,

$$X_n(\theta_n) - X_n(\theta_0) \xrightarrow{\mathbb{P}} 0 \quad (n \to \infty)\,. \tag{B.26}$$

Proof. In order to prove Lemma B.2.7, we need to show that for all $\varepsilon > 0$ we have

$$\lim_{n\to\infty} \mathbb{P}\left(\|X_n(\theta_n) - X_n(\theta_0)\| \geq \varepsilon\right) = 0\,. \tag{B.27}$$

For any $\eta > 0$, there exist $\delta > 0$ and $n_0 \in \mathbb{N}$ with

$$\begin{aligned} &\mathbb{P}\left(\sup_{\theta \in \mathrm{B}_\delta(\theta_0)} \|X_n(\theta) - X_n(\theta_0)\| \geq \varepsilon \right) \\ &\quad \leq \mathbb{P}\left(\sup_{\|\theta - \theta'\| < \delta} \|X_n(\theta) - X_n(\theta')\| \geq \varepsilon \right) \leq \frac{\eta}{2}\,, \quad \text{for all } n \geq n_0\,, \end{aligned} \tag{B.28}$$

because Inequality (B.22) holds according to the prerequisite. The $\mathbb{P}$-almost sure convergence of $(\theta_n)_{n\in\mathbb{N}}$ to θ_0 implies that, given $\delta > 0$, there exists $n_1 \in \mathbb{N}$ such that

$$\mathbb{P}\left(\|\theta_n - \theta_0\| < \delta \text{ for all } n \geq n_1\right) \geq 1 - \frac{\eta}{2}\,, \tag{B.29}$$

see Bauer 1996, p. 32 for a reference on this elementary property. For $n \geq \max\{n_0, n_1\}$, the combination of Equations (B.28) and (B.29) yields:

$$\begin{aligned}
&\mathbb{P}\left(\|X_n(\theta_n) - X_n(\theta_0)\| \geq \varepsilon\right) \\
&\quad = \mathbb{P}\Big(\Big\{\|X_n(\theta_n) - X_n(\theta_0)\| \geq \varepsilon\Big\} \cap \underbrace{\Big\{\|\theta_m - \theta_0\| < \delta \text{ for all } m \geq n_1\Big\}}_{\text{holds for } m=n, \text{ since } n \geq n_1.}\Big) \\
&\qquad + \mathbb{P}\left(\Big\{\|X_n(\theta_n) - X_n(\theta_0)\| \geq \varepsilon\Big\} \cap \Big\{\|\theta_m - \theta_0\| < \delta \text{ for all } m \geq n_1\Big\}^{\complement}\right) \\
&\quad \leq \mathbb{P}\left(\Big\{\|X_n(\theta_n) - X_n(\theta_0)\| \geq \varepsilon\Big\} \cap \Big\{\|\theta_n - \theta_0\| < \delta\Big\}\right) \\
&\qquad + \underbrace{\mathbb{P}\left(\Big\{\|\theta_m - \theta_0\| < \delta \text{ for all } m \geq n_1\Big\}^{\complement}\right)}_{\leq \frac{\eta}{2} \text{ by virtue of Equation (B.29).}} \\
&\quad \leq \underbrace{\mathbb{P}\left(\sup_{\theta \in \mathrm{B}_\delta(\theta_0)} \|X_n(\theta) - X_n(\theta_0)\| \geq \varepsilon\right)}_{\leq \frac{\eta}{2} \text{ by Equation (B.28), since } n \geq n_0.} + \frac{\eta}{2} \leq \eta\,.
\end{aligned}$$

Since η was chosen arbitrarily, convergence in Equation (B.27) thus follows and so does the assertion. □

We end this detour into the topic of tightness by noting that the results shown are of secondary use to us. In Kopperschmidt and Stute 2013, the authors attempt to apply Kolmogorov's tightness criterion from Theorem B.2.6 and the associated Lemma B.2.7 to derive the asymptotic normality of their proposed minimum distance estimator. We discuss in Remark 3.27 why this approach fails and the tightness of the processes involved cannot be shown by virtue of Kolmogorov's criterion. However, to point out the flaws in their reasoning, a formulation of the statements used is imperative. Moreover, we can understand Lemma B.2.7 as the essential reason for Kopperschmidt's consideration of tightness, in the sense that an alternative approach to proving convergence in Equation (B.23) leads to the desired asymtptotic normality even without tightness.

B.3. Proof of the Strong Consistency of the Minimum Distance Estimator

The proof presented here is based on the abbreviated version in Kopperschmidt and Stute 2013, pp. 1284–1288, which in turn goes back to Kopperschmidt 2005, pp. 63–80. Furthermore, we incorporate additional details provided by the author in his master's thesis, Jakubzik 2017, pp. 83–99. The proof can be roughly divided into two parts: First, we invoke the strong law of large numbers for U-statistics to derive some general convergence statements. Then, these intermediate results are combined to show that the minimum distance estimator $\hat{\theta}_n$ converges to the true parameter θ^* with probability one. Throughout the proof of not only the consistency but also the asymptotic normality of the minimum distance estimator, we will deal with pathwise integrals with respect to various stochastic processes, namely the counting processes $N^{(i)}$, the cumulative intensities $\Lambda_\theta^{(i)}$ and the innovation martingales $M^{(i)} = N^{(i)} - \Lambda_{\theta^*}^{(i)}$. In doing so, we often need to determine the expectation of such integrals. We cite a useful lemma for this purpose, given and proved in Kopperschmidt 2005, pp. 130–139.

B.3.1. Expectations of Pathwise Integrals

We follow the work of Kopperschmidt 2005 and start by specifying the framework. Let X, Y and $Z = Z^+ - Z^-$ be stochastic processes with common index space $\mathcal{I} = [0, \tau] \subset \mathbb{R}$ and the following properties:

(i) The processes Z^+ and Z^- are almost surely non-negative, right-continuous and non-decreasing.

(ii) Almost surely $|\mathcal{J}_X| < \infty$ and $|\mathcal{J}_Y| < \infty$ applies, where $\mathcal{J}_X$ and $\mathcal{J}_Y$ denote the (random) sets of discontinuities of X and Y, respectively.

(iii) If μ_{Z^+} and μ_{Z^-} denote the (random) Borel measures on $\mathcal{I}$ induced by Z^+ and Z^-, respectively, then almost surely holds:

$$\mu_{Z^+}(\mathcal{J}_X) = \mu_{Z^-}(\mathcal{J}_X) = 0\,. \tag{B.30}$$

(iv) The following expectations exists:

$$\mathbb{E}\left[\int_{\mathcal{I}} |XY|\,\mathrm{d}\big(Z^+ + Z^-\big)\right] < \infty\,. \tag{B.31}$$

In this setting, the following lemma takes place.

Lemma B.3.1 (Expectations of Pathwise Integrals; Lemma A.2.3 of Kopperschmidt 2005, p. 134)**.**

(i) Assume that X_t is independent of $\sigma(\{Y_s, Z_s : s \in \mathcal{I}, s \leq t\})$ for each $t \in \mathcal{I}$. If

$$\mathbb{E}\left[\int_{\mathcal{I}} |\mathbb{E}(X)\,Y|\,\mathrm{d}\big(Z^+ + Z^-\big)\right] < \infty\,, \qquad \mathbb{E}\Big(\sup_{t\in\mathcal{I}} |X_t|\Big) < \infty\,, \tag{B.32}$$

and for all $t \in \mathcal{I}$ we have

$$\mathbb{P}(t \in \mathcal{J}_X) = \mathbb{P}(X_t - X_{t-} \neq 0) = 0\,, \tag{B.33}$$

then

$$\mathbb{E}\left[\int_{\mathcal{I}} XY\,\mathrm{d}Z\right] = \mathbb{E}\left[\int_{\mathcal{I}} \mathbb{E}(X)\,Y\,\mathrm{d}Z\right].$$

(ii) Assume that Z_t is independent of $\sigma(\{X_s, Y_s : s \in \mathcal{I}, s \leq t\})$ for each $t \in \mathcal{I}$. If

$$\mathbb{E}\left[\int_{\mathcal{I}} |XY|\,\mathrm{d}\big(\mathbb{E}Z^+ + \mathbb{E}Z^-\big)\right] < \infty\,, \qquad \mathbb{E}(Z_t) < \infty \quad \textit{for all } t \in \mathcal{I}\,, \tag{B.34}$$

and almost surely we have

$$\mu_{Z^+}(\mathcal{J}_Y) = \mu_{Z^-}(\mathcal{J}_Y) = 0\,, \tag{B.35}$$

then

$$\mathbb{E}\left[\int_{\mathcal{I}} XY \mathrm{d}Z\right] = \mathbb{E}\left[\int_{\mathcal{I}} XY \mathrm{d}\mathbb{E}Z\right].$$

Proof. A heuristic idea of proof for Lemma B.3.1 is easily obtained by looking at Riemann approximations of the integrals involved. Since we only sketch the essential approach, we limit ourselves to part (i). Let $(S_n)_{n\in\mathbb{N}}$ be a sequence of partitions of $\mathcal{I}$, so that

$$S_n = \{0 = s_{n,0} < s_{n,1} < \ldots < s_{n,q_n} = \tau\}$$

satisfies

$$\max_{i=1,\ldots,q_n} (s_{n,i} - s_{n,i-1}) \longrightarrow 0 \quad (n \to \infty),$$

where $(q_n)_{n\in\mathbb{N}} \subset \mathbb{N}$ is an increasing sequence. Then, for sufficiently large $n \in \mathbb{N}$,

$$\begin{aligned}
\mathbb{E}\left[\int_{\mathcal{I}} XY\,dZ\right] &\approx \mathbb{E}\left[\sum_{i=1}^{n} X_{s_{n,i-1}} Y_{s_{n,i-1}} \left(Z_{s_{n,i}} - Z_{s_{n,i-1}}\right)\right] \\
&= \sum_{i=1}^{n} \mathbb{E}\left[X_{s_{n,i-1}} Y_{s_{n,i-1}} \left(Z_{s_{n,i}} - Z_{s_{n,i-1}}\right)\right] \\
&= \sum_{i=1}^{n} \mathbb{E}\left[\mathbb{E}\Big(X_{s_{n,i-1}} Y_{s_{n,i-1}} \left(Z_{s_{n,i}} - Z_{s_{n,i-1}}\right) \mid \sigma\left(\{Y_s\,, Z_s \,:\, s \leq s_{n,i}\}\right)\Big)\right] \\
&= \sum_{i=1}^{n} \mathbb{E}\Big[Y_{s_{n,i-1}} \left(Z_{s_{n,i}} - Z_{s_{n,i-1}}\right) \underbrace{\mathbb{E}\Big(X_{s_{n,i-1}} \mid \sigma\left(\{Y_s\,, Z_s \,:\, s \leq s_{n,i}\}\right)\Big)}_{=\mathbb{E}(X_{s_{n,i-1}})}\Big] \\
&= \sum_{i=1}^{n} \mathbb{E}\left[\mathbb{E}\left(X_{s_{n,i-1}}\right) Y_{s_{n,i-1}} \left(Z_{s_{n,i}} - Z_{s_{n,i-1}}\right)\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} \mathbb{E}\left(X_{s_{n,i-1}}\right) Y_{s_{n,i-1}} \left(Z_{s_{n,i}} - Z_{s_{n,i-1}}\right)\right] \\
&\approx \mathbb{E}\left[\int_{\mathcal{I}} \mathbb{E}(X) Y\,dZ\right],
\end{aligned}$$

by virtue of the tower property. However, we cannot ensure the convergence of the Riemann sum to the integral here because we did not account for the discontinuities of X, Y and Z. The difficulty lies in finding a suitable partition of $\mathcal{I}$. In his Lemma A.2.2 on pages 131 to 134, Kopperschmidt 2005 addresses this issue and validates the desired convergence of the constructed (Darboux-)Riemann sum to the corresponding integral. Lemma A.2.3 then emerges as a corollary of this lemma, basically applying the above heuristic in a mathematically rigorous way. The formal proof is technical and is omitted here, but can be found in Kopperschmidt 2005, pp. 134–139. □

B.3.2. Implications of the Strong Law of Large Numbers for U-Statistics

We now begin the first part of the proof by stating the strong law of large numbers for U-statistics. For this purpose we have to introduce a few more terms: mainly the Polish spaces and the U-statistics themselves, but also the standard Borel spaces.

Definition B.3.2 (Polish Space; Klenke 2020, p. 209).
A topological space $(E, \mathcal{T})$ *is called a* Polish space *if it is separable and its topology is induced by a complete metric.*

The Polish space we encounter the most is the Banach space $\left(\mathbb{R}^d, \|\cdot\|\right)$ for some $d \in \mathbb{N}$. Another prominent example is the Skorokhod space of càdlàg functions.

Definition B.3.3 (Skorokhod Space; cf. Billingsley 1968, pp. 109–113).
Let $\mathcal{I} = [0, \tau] \subset \mathbb{R}$. *Let* $\mathcal{D}(\mathcal{I})$ *be the space of càdlàg functions on* $\mathcal{I}$*, that is, the space of functions* $x : \mathcal{I} \to \mathbb{R}$ *that are right-continuous and have left-hand limits:*

(i) For all $t \in [0, \tau)$, $x(t+) = \lim_{s \downarrow t} x(s)$ *exists and* $x(t+) = x(t)$.

(ii) For all $t \in (0, \tau]$, $x(t-) = \lim_{s \uparrow t} x(s)$ *exists.*

Let Γ *denote the class of strictly increasing, continuous mappings of* $\mathcal{I}$ *onto itself. For* $\gamma \in \Gamma$*, we define:*

$$\|\gamma\| := \sup_{s \neq t} \left| \log \frac{\gamma(t) - \gamma(s)}{t - s} \right| .$$

Then, a metric d on $\mathcal{D}(\mathcal{I})$ *is given by*

$$d(x, y) := \inf_{\gamma \in \Gamma} \max \left\{ \|\gamma\| , \sup_{t \in \mathcal{I}} |x(t) - y(\gamma(t))| \right\}. \tag{B.36}$$

If $\mathcal{T}_d$ *is the topology induced by d, the topological space* $(\mathcal{D}(\mathcal{I}), \mathcal{T}_d)$ *is a Polish space called* Skorokhod space.

Remark B.3.4 (Technical Details Regarding the Skorokhod Space).

(i) It is shown in Billingsley 1968, pp. 111–113, that d given by Equation (B.36) indeed defines a metric on $\mathcal{D}(\mathcal{I})$.

(ii) In the definition of the space $\mathcal{D}(\mathcal{I})$, the codomain $\mathbb{R}$ can be replaced by any metric space (E, d_E). Ethier and Kurtz 1986, pp. 121–122, show that if (E, d_E) is a Polish space, the same holds for $(\mathcal{D}(\mathcal{I}), \mathcal{T}_\mathrm{d})$. The case $E = \mathbb{R}$ is again discussed in Billingsley 1968, pp. 112–116.

(iii) The Skorokhod space is named after the topology introduced by Skorokhod, see Skorokhod 1956, p. 265. This topology is also induced by a more easily interpretable metric, which, however, is not complete (Billingsley 1968, pp. 111–112).

The introduction of the Skorokhod space enables a crucial change in perspective: we can now conceive of counting processes as well as their compensators and innovation martingales as $\mathcal{D}(\mathcal{I})$-valued random elements. Moreover, we will recognize that any random element in a Polish space can be treated like a real-valued random variable. We can therefore draw on the familiar repertoire of methods - such as the strong law of large numbers - when dealing with such stochastic processes. We concretize this train of thought with the help of (standard) Borel spaces.

Definition B.3.5 (Isomorphic Measurable Spaces; Klenke 2020, p. 208).
Two measurable spaces $(E, \mathcal{E})$ *and* $(E', \mathcal{E}')$ *are called* isomorphic *if there exists a bijective map* $\phi : E \to E'$ *such that* ϕ *is* $\mathcal{E} - \mathcal{E}'$*-measurable and the inverse map* ϕ^{-1} *is* $\mathcal{E}' - \mathcal{E}$*-measurable.*

Definition B.3.6 ((Standard) Borel Space; Klenke 2020, p. 208)**.**
A measurable space $(E, \mathcal{E})$ *is called a* (standard) Borel space *if there exists a Borel set* $B \in \mathcal{B}(\mathbb{R})$ *such that* $(E, \mathcal{E})$ *and* $(B, \mathcal{B}(B))$ *are isomorphic.*

We occasionally add the term "standard" to avoid confusion with measurable spaces equipped with the Borel σ-algebra induced by a topology (which are not necessarily Borel spaces in the sense of Definition B.3.6). However, an essential insight is that every Polish Space is in fact a Borel space.

Theorem B.3.7 (Polish Spaces as Standard Borel Spaces; Klenke 2020, p. 209)**.**
Let $(E, \mathcal{T})$ *be a Polish space and* $\mathcal{E} = \mathcal{B}(\mathcal{T})$ *the Borel* σ*-algebra induced by* $\mathcal{T}$*. Then* $(E, \mathcal{E})$ *is a standard Borel space.*

Proof. The result is a special case of the more general Theorem 2.12 from Parthasarathy 1967, p. 14. It states that two Borel sets of Polish spaces are isomorphic if and only if they have the same cardinality. For the detailed proof, see Parthasarathy 1967, pp. 7–14. □

We have now completed the necessary preparations to move on to U-statistics. They were first considered by Wassily Hoeffding, the chosen name indicating the *unbiased* nature of the statistic, see Hoeffding 1948, p. 293.

Definition B.3.8 (U-Statistic; cf. Hoeffding 1948, pp. 296–297 and Lee 1990, pp. 2, 7–8)**.**

Let $X = (X_n)_{n \in \mathbb{N}}$ *be a sequence of independent and identically distributed random variables taking values in a measurable space* $(E, \mathcal{E})$*. Furthermore, let* $m \in \mathbb{N}$ *and denote by* $E^m = E \times \ldots \times E$ *the* m*-fold Cartesian product of* E*. If we equip* E^m *with the corresponding product* σ*-algebra, that is,*

$$\mathcal{E}^m := \bigotimes_{i=1}^{m} \mathcal{E} \,,$$

then $(E^m, \mathcal{E}^m)$ *is a measurable space. For* $n \geq m$*, let* (n, m) *be the set of all injective mappings from* $\{1, \ldots, m\}$ *to* $\{1, \ldots, n\}$*:*

$$(n, m) := \{\tau : \{1, \ldots, m\} \to \{1, \ldots, n\} : \tau(i) \neq \tau(j) \text{ for } i \neq j \,, 1 \leq i, j \leq m\} \,.$$

Let $\psi : (E^m, \mathcal{E}^m) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *be a measurable function and define*

$$u_n\left((x_k)_{k \in \mathbb{N}}\right) := \frac{(n-m)!}{n!} \sum_{\tau \in (n,m)} \psi\left(x_{\tau(1)}, \ldots, x_{\tau(m)}\right) , \quad m \leq n \in \mathbb{N} \,. \tag{B.37}$$

Then, $U_n := u_n \circ X$ *is called a* U-statistic *with* m*-dimensional* kernel ψ*.*

In the literature, Equation (B.37) is expressed in various ways. For instance, the U-statistic U_n is commonly given as

$$U_n = \frac{(n-m)!}{n!} \sum \psi\left(X_{i_1}, \ldots, X_{i_m}\right) , \tag{B.38}$$

where the sum is extended over all permutations $(i_1, \ldots, i_m)$ of m different integers, $1 \leq i_j \leq n$ (see Hoeffding 1948, p. 293 and Kopperschmidt and Stute 2013, p. 1294).
We can now state the strong law of large numbers for U-statistics in a form suitable for our applications.

Theorem B.3.9 (Strong Law of Large Numbers for U-Statistics; cf. Lee 1990, p. 122). *In the situation of Definition B.3.8, assume that* $(E, \mathcal{E})$ *is a Polish space (i.e.,* $\mathcal{E} = \mathcal{B}(\mathcal{T})$ *as in Theorem B.3.7). Suppose that* $\mathbb{E}\left|\psi\left(X_1, \ldots, X_m\right)\right| < \infty$. *Then,*

$$U_n \longrightarrow \mathbb{E}\left(\psi\left(X_1, \ldots, X_m\right)\right) \quad (n \to \infty) \qquad \mathbb{P}\text{-almost surely.} \tag{B.39}$$

Proof. Several proofs are given in Lee 1990, pp. 122–131. One of these proofs is mentioned in Kopperschmidt and Stute 2013, p. 1294 and has been discussed in detail by the author, see Jakubzik 2017, pp. 53–64. It is carried out in 3 steps, where initially we assume that $(E, \mathcal{E}) = (B, \mathcal{B}(B))$ for some Borel set $B \in \mathcal{B}(\mathbb{R})$:

1. Define $\mathcal{F}_n := \sigma\left(U_j \,:\, j \geq n\right)$ and show that $\left(U_n, \mathcal{F}_n\right)_{n\in\mathbb{N}}$ is a reverse martingale (cf. Lee 1990, p. 112).

2. Apply Doob's second martingale convergence theorem (see Theorem 12.14 of Klenke 2020, p. 264) to obtain

$$U_n \longrightarrow U_\infty \quad (n \to \infty) \qquad \mathbb{P}\text{- almost surely}$$

   for some $\bigcap_{n\in\mathbb{N}} \mathcal{F}_n$-measurable random variable U_∞.

3. Use the 0-1 law of Hewitt-Savage (see Corollary 12.19 of Klenke 2020, p. 265) to conclude that

$$U_\infty = \mathbb{E}\left(\psi\left(X_1, \ldots, X_m\right)\right) \qquad \mathbb{P}\text{- almost surely.}$$

In the general case that $(E, \mathcal{E})$ is an arbitrary Polish space, according to Theorem B.3.7, we can find a Borel set $B \in \mathcal{B}(\mathbb{R})$ and an isomorphism $\phi : (E, \mathcal{E}) \to (B, \mathcal{B}(B))$. Defining the kernel $\tilde{\psi} : (B^m, \mathcal{B}(B)^m) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ by

$$\tilde{\psi}\left(x_1, \ldots, x_m\right) := \psi\left(\phi^{-1}(x_1), \ldots, \phi^{-1}(x_m)\right)$$

and applying the above steps to the sequence $\left(\phi(X_n)\right)_{n\in\mathbb{N}}$ instead of $\left(X_n\right)_{n\in\mathbb{N}}$ yields:

$$\begin{aligned} U_n &= \frac{(n-m)!}{n!} \sum_{\tau\in(n,m)} \psi\left(X_{\tau(1)}, \ldots, X_{\tau(m)}\right) \\ &= \frac{(n-m)!}{n!} \sum_{\tau\in(n,m)} \tilde{\psi}\left(\phi(X_{\tau(1)}), \ldots, \phi(X_{\tau(m)})\right) \\ &\overset{n\to\infty}{\longrightarrow} \quad \mathbb{E}\left(\tilde{\psi}\left(\phi(X_1), \ldots, \phi(X_m)\right)\right) = \mathbb{E}\left(\psi\left(X_1, \ldots, X_m\right)\right) \quad \mathbb{P}\text{-almost surely,} \end{aligned}$$

and hence the desired result. $\square$

Applying the strong law of large numbers for U-statistics immediately yields some simple limit results. For this, let $\mathrm{B}_r^\varepsilon(\theta)$ be the part of the r-ball centred in θ that does not belong to $\mathrm{B}_\varepsilon(\theta^*)$,

$$\mathrm{B}_r^\varepsilon(\theta) := \left\{\theta' \in \overline{\Theta} \setminus \mathrm{B}_\varepsilon(\theta^*) \,:\, \|\theta' - \theta\| < r\right\}.$$

The notation was chosen along the lines of Kopperschmidt and Stute 2013, p. 1284. See Figure 21 for an illustration of this set.

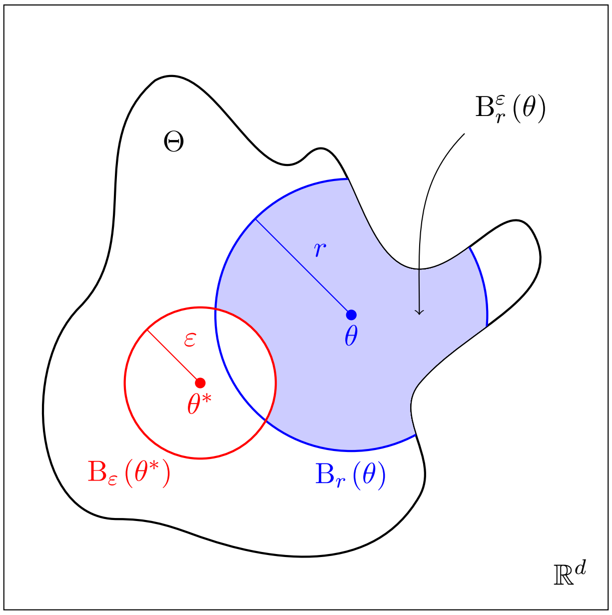


Figure 21: Illustration of the set $\mathrm{B}_r^\varepsilon(\theta)$. Note that by definition in Equation (3.6), $\mathrm{B}_r(\theta)$ consists only of the part of the r-ball around θ that belongs to $\overline{\Theta}$.

Lemma B.3.10 (Applications of the Strong Law of Large Numbers for U-Statistics; cf. Kopperschmidt and Stute 2013, p. 1285)**.**
Under the assumptions of Theorem 3.7, the following limits hold $\mathbb{P}$*-almost surely:*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \sup_{\theta'\in B_r^\varepsilon(\theta)} \left\| N^{(i)} - \Lambda_{\theta'}^{(i)} \right\|_{N^{(i)}}^2 = \mathbb{E}\left[\sup_{\theta'\in B_r^\varepsilon(\theta)} \| N - \Lambda_{\theta'} \|_N^2 \right], \tag{B.40}$$

$$\lim_{n\to\infty} \frac{1}{n(n-1)} \sum_{\substack{i,j=1\\ i\neq j}}^{n} \sup_{\theta'\in B_r^\varepsilon(\theta)} \left\| N^{(i)} - \Lambda_{\theta'}^{(i)} \right\|_{N^{(j)}}^2 = \mathbb{E}\left[\sup_{\theta'\in B_r^\varepsilon(\theta)} \left\| N^{(1)} - \Lambda_{\theta'}^{(1)} \right\|_{N^{(2)}}^2 \right], \tag{B.41}$$

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{n(n-1)} \sum_{\substack{i,j=1\\ i\neq j}}^{n} \sup_{\theta'\in B_r^\varepsilon(\theta)} \langle N^{(i)} - \Lambda_{\theta'}^{(i)}, N^{(j)} - \Lambda_{\theta'}^{(j)} \rangle_{N^{(i)}} \\ &\quad = \mathbb{E}\left[\sup_{\theta'\in B_r^\varepsilon(\theta)} \langle N^{(1)} - \Lambda_{\theta'}^{(1)}, N^{(2)} - \Lambda_{\theta'}^{(2)} \rangle_{N^{(1)}} \right], \end{aligned} \tag{B.42}$$

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{n(n-1)(n-2)} \sum_{\substack{i,j,k=1\\ i\neq j\neq k\neq i}}^{n} \sup_{\theta'\in B_r^\varepsilon(\theta)} \langle N^{(i)} - \Lambda_{\theta'}^{(i)}, N^{(j)} - \Lambda_{\theta'}^{(j)} \rangle_{N^{(k)}} \\ &\quad = \mathbb{E}\left[\sup_{\theta'\in B_r^\varepsilon(\theta)} \langle N^{(1)} - \Lambda_{\theta'}^{(1)}, N^{(2)} - \Lambda_{\theta'}^{(2)} \rangle_{N^{(3)}} \right], \end{aligned} \tag{B.43}$$

where $\varepsilon, r > 0$ *and* $\theta \in \Theta$*. These limits are also valid without the supremum, that is, they apply pointwise in* $\theta \in \Theta$*. In addition, the supremum can be replaced by an infimum.*

Because of the i.i.d. assumption, the choice of the particular indices in the above limits is arbitrary as long as they differ appropriately.

Proof. We first recall that the counting processes $N^{(i)}$ can be understood as random variables with values in the Skorokhod space $(\mathcal{D}(\mathcal{I}), \mathcal{T}_\mathrm{d})$. Moreover, by virtue of (C2), the associated cumulative intensities can likewise be viewed as random variables, that is,

$$\omega \longmapsto \Big[(t,\theta) \mapsto \Lambda_\theta^{(i)}(t,\omega)\Big] ,$$

mapping into the Banach space $C^0(\mathcal{I} \times \overline{\Theta})$, which is Polish (see Theorem (4.19) of Kechris 1995, p. 24). Since the product of Polish spaces is again Polish (cf. Klenke 2020, p. 305), the bivariate random variables $(N^{(i)}, \Lambda_\theta^{(i)})$ also take values in a Polish space. For each of the four equations (B.40) to (B.43), we construct a kernel ψ so that we can take the left-hand sides as U-statistics and apply the corresponding strong law of large numbers:

$$\begin{aligned}
\psi_1\left((N,\Lambda_\theta)\right) &:= \sup_{\theta' \in \mathrm{B}_r^\varepsilon(\theta)} \|N - \Lambda_{\theta'}\|_N^2 ,\\
\psi_2\Big((N^{(1)},\Lambda_\theta^{(1)}),(N^{(2)},\Lambda_\theta^{(2)})\Big) &:= \sup_{\theta' \in \mathrm{B}_r^\varepsilon(\theta)} \Big\|N^{(1)} - \Lambda_{\theta'}^{(1)}\Big\|_{N^{(2)}}^2 ,\\
\psi_3\Big((N^{(1)},\Lambda_\theta^{(1)}),(N^{(2)},\Lambda_\theta^{(2)})\Big) &:= \sup_{\theta' \in \mathrm{B}_r^\varepsilon(\theta)} \langle N^{(1)} - \Lambda_{\theta'}^{(1)}, N^{(2)} - \Lambda_{\theta'}^{(2)}\rangle_{N^{(1)}} ,\\
\psi_4\Big((N^{(1)},\Lambda_\theta^{(1)}),(N^{(2)},\Lambda_\theta^{(2)}),(N^{(3)},\Lambda_\theta^{(3)})\Big) &:= \sup_{\theta' \in \mathrm{B}_r^\varepsilon(\theta)} \langle N^{(1)} - \Lambda_{\theta'}^{(1)}, N^{(2)} - \Lambda_{\theta'}^{(2)}\rangle_{N^{(3)}} .
\end{aligned}$$

In order to apply Theorem B.3.9, we only need to check that

$$\begin{aligned}
&\mathbb{E}\left|\psi_1\left((N,\Lambda_\theta)\right)\right| && < \infty ,\\
&\mathbb{E}\Big|\psi_2\Big((N^{(1)},\Lambda_\theta^{(1)}),(N^{(2)},\Lambda_\theta^{(2)})\Big)\Big| && < \infty ,\\
&\mathbb{E}\Big|\psi_3\Big((N^{(1)},\Lambda_\theta^{(1)}),(N^{(2)},\Lambda_\theta^{(2)})\Big)\Big| && < \infty ,\\
&\mathbb{E}\Big|\psi_4\Big((N^{(1)},\Lambda_\theta^{(1)}),(N^{(2)},\Lambda_\theta^{(2)}),(N^{(3)},\Lambda_\theta^{(3)})\Big)\Big| && < \infty .
\end{aligned}$$

This is immediately clear from (C3), since for all $\theta' \in \Theta$ holds $\mathbb{P}$-almost surely:

$$\begin{aligned}
\Big|\langle N^{(i)} - \Lambda_{\theta'}^{(i)}, N^{(j)} - \Lambda_{\theta'}^{(j)}\rangle_{N^{(k)}}\Big| &= \left|\int_I \Big(N^{(i)} - \Lambda_{\theta'}^{(i)}\Big)\Big(N^{(j)} - \Lambda_{\theta'}^{(j)}\Big)\,\mathrm{d}N^{(k)}\right|\\
&\leq \int_I \underbrace{\Big|N^{(i)} - \Lambda_{\theta'}^{(i)}\Big|}_{\leq |N^{(i)}| + |\Lambda_{\theta'}^{(i)}| \leq 2C} \Big|N^{(j)} - \Lambda_{\theta'}^{(j)}\Big|\,\mathrm{d}N^{(k)}\\
&\leq 4C^2\Big(N_\tau^{(k)} - N_0^{(k)}\Big)\\
&\leq 4C^3 < \infty , \qquad \text{for all } i,j,k \in \mathbb{N} .
\end{aligned}$$

Since the bounds hold uniformly in θ', they are also valid for the supremum with respect to $\theta' \in \mathrm{B}_r^\varepsilon(\theta)$. The existence of the expectation then follows trivially. For the additional statements, one simply adjusts the kernels ψ_1 to ψ_4 appropriately by either replacing the supremum with an infimum or omitting it. $\square$

Kopperschmidt 2005, pp. 68–70 shows that here as elsewhere the weaker condition $(\widetilde{\mathrm{C}}3)$

suffices for finiteness of the above expectations. We demonstrate this using nearly the same techniques as before, but require Hölder's inequality in the final step.

$$
\begin{aligned}
&\mathbb{E}\left|\sup_{\theta'\in\mathrm{B}_r^\varepsilon(\theta)} \langle N^{(i)} - \Lambda_{\theta'}^{(i)}, N^{(j)} - \Lambda_{\theta'}^{(j)}\rangle_{N^{(k)}}\right| \qquad\qquad \text{(B.44)}\\
&\quad\leq \mathbb{E}\left[\sup_{\theta'\in\mathrm{B}_r^\varepsilon(\theta)} \left|\langle N^{(i)} - \Lambda_{\theta'}^{(i)}, N^{(j)} - \Lambda_{\theta'}^{(j)}\rangle_{N^{(k)}}\right|\right]\\
&\quad\leq \mathbb{E}\left[\sup_{\theta'\in\mathrm{B}_r^\varepsilon(\theta)} \int \left|N^{(i)} - \Lambda_{\theta'}^{(i)}\right| \left|N^{(j)} - \Lambda_{\theta'}^{(j)}\right| \,\mathrm{d}N^{(k)}\right]\\
&\quad\leq \mathbb{E}\left[\sup_{\theta'\in\mathrm{B}_r^\varepsilon(\theta)} \int \left(\left|N^{(i)}\right| + \left|\Lambda_{\theta'}^{(i)}\right|\right)\left(\left|N^{(j)}\right| + \left|\Lambda_{\theta'}^{(j)}\right|\right) \,\mathrm{d}N^{(k)}\right]\\
&\quad= \mathbb{E}\left[\sup_{\theta'\in\mathrm{B}_r^\varepsilon(\theta)} \int N^{(i)}N^{(j)} + N^{(i)}\Lambda_{\theta'}^{(j)} + \Lambda_{\theta'}^{(i)}N^{(j)} + \Lambda_{\theta'}^{(i)}\Lambda_{\theta'}^{(j)} \,\mathrm{d}N^{(k)}\right]\\
&\quad\leq \mathbb{E}\left[\sup_{\theta'\in\mathrm{B}_r^\varepsilon(\theta)} \left(N_\tau^{(i)}N_\tau^{(j)} + N^{(i)}\Lambda_{\theta'}^{(j)}(\tau) + \Lambda_{\theta'}^{(i)}(\tau)N_\tau^{(j)} + \Lambda_{\theta'}^{(i)}(\tau)\Lambda_{\theta'}^{(j)}(\tau)\right) N_\tau^{(k)}\right]\\
&\quad\leq \mathbb{E}\left[\sup_{\theta'\in\mathrm{B}_r^\varepsilon(\theta)} N_\tau^{(i)}N_\tau^{(j)}N_\tau^{(k)}\right] + \mathbb{E}\left[\sup_{\theta'\in\mathrm{B}_r^\varepsilon(\theta)} N_\tau^{(i)}\Lambda_{\theta'}^{(j)}(\tau)N_\tau^{(k)}\right]\\
&\qquad + \mathbb{E}\left[\sup_{\theta'\in\mathrm{B}_r^\varepsilon(\theta)} \Lambda_{\theta'}^{(i)}(\tau)N_\tau^{(j)}N_\tau^{(k)}\right] + \mathbb{E}\left[\sup_{\theta'\in\mathrm{B}_r^\varepsilon(\theta)} \Lambda_{\theta'}^{(i)}(\tau)\Lambda_{\theta'}^{(j)}(\tau)N_\tau^{(k)}\right]\\
&\quad\leq \left(\mathbb{E}\left[\left(N_\tau^{(i)}\right)^3\right]\cdot\mathbb{E}\left[\left(N_\tau^{(j)}\right)^3\right]\cdot\mathbb{E}\left[\left(N_\tau^{(k)}\right)^3\right]\right)^{\frac{1}{3}}\\
&\qquad + \left(\mathbb{E}\left[\left(N_\tau^{(i)}\right)^3\right]\cdot\mathbb{E}\left[\sup_{\theta'\in\mathrm{B}_r^\varepsilon(\theta)}\left(\Lambda_{\theta'}^{(j)}(\tau)\right)^3\right]\cdot\mathbb{E}\left[\left(N_\tau^{(k)}\right)^3\right]\right)^{\frac{1}{3}}\\
&\qquad + \left(\mathbb{E}\left[\sup_{\theta'\in\mathrm{B}_r^\varepsilon(\theta)}\left(\Lambda_{\theta'}^{(i)}(\tau)\right)^3\right]\cdot\mathbb{E}\left[\left(N_\tau^{(j)}\right)^3\right]\cdot\mathbb{E}\left[\left(N_\tau^{(k)}\right)^3\right]\right)^{\frac{1}{3}}\\
&\qquad + \left(\mathbb{E}\left[\sup_{\theta'\in\mathrm{B}_r^\varepsilon(\theta)}\left(\Lambda_{\theta'}^{(i)}(\tau)\right)^3\right]\cdot\mathbb{E}\left[\sup_{\theta'\in\mathrm{B}_r^\varepsilon(\theta)}\left(\Lambda_{\theta'}^{(j)}(\tau)\right)^3\right]\cdot\mathbb{E}\left[\left(N_\tau^{(k)}\right)^3\right]\right)^{\frac{1}{3}} < \infty\,,
\end{aligned}
$$

as a consequence of Equations (3.7) and (3.8) as well as the compactness of $\overline{\Theta}$.
We now use Lemma B.3.1 to simplify the pointwise version of Equation (B.43).

Lemma B.3.11 (cf. Kopperschmidt and Stute 2013, p. 1285)**.**
Under the assumptions of Theorem 3.7, we have

$$
\mathbb{E}\left[\langle N^{(1)} - \Lambda_\theta^{(1)}, N^{(2)} - \Lambda_\theta^{(2)}\rangle_{N^{(3)}}\right] = \|\mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_\theta\|_{\mathbb{E}\Lambda_{\theta^*}}^2\,. \qquad \text{(B.45)}
$$

Proof. We apply Lemma B.3.1 three times in succession, using first part (i) twice and then part (ii) once. We have $Z = Z^+ = N^{(3)}$ and thus $Z^- = 0$ across all these applications. For $X = N^{(1)} - \Lambda_\theta^{(1)}$ and $Y = N^{(2)} - \Lambda_\theta^{(2)}$, the integrability conditions of Equations (B.31) and (B.32) are again met by assumption ($\widetilde{\mathrm{C}}3$) which is implied by (C3). The discontinuities of X and Y correspond to those of $N^{(1)}$ and $N^{(2)}$, respectively, since the cumulative intensities are (absolutely) continuous according to (M2). These jump points of $N^{(1)}$ and $N^{(2)}$ in $\mathcal{I}$ are in turn represented by the associated simple point processes $T^{(1)}$ and $T^{(2)}$, so that we receive:

$$\mathcal{J}_X = \left\{T_i^{(1)} \,:\, i \in \mathbb{N}\right\} \cap \mathcal{I}\,, \qquad \mathcal{J}_Y = \left\{T_i^{(2)} \,:\, i \in \mathbb{N}\right\} \cap \mathcal{I}\,.$$

By Remark A.6 (iii), almost surely only finitely many jumps can occur in $\mathcal{I}$, so that $|\mathcal{J}_X| < \infty$ and $|\mathcal{J}_Y| < \infty$ follows. Now let $T^{(3)}$ denote the simple point process associated with $Z = N^{(3)}$. Due to the continuous finite-dimensional distribution of the stochastically independent point processes $T^{(1)}$, $T^{(2)}$ and $T^{(3)}$, with probability 1 the paths of $N^{(1)}$, $N^{(2)}$ and $N^{(3)}$ have no common discontinuity points, see Lemma A.37. This implies both the conditions from Equation (B.30) and Equation (B.33), so all preconditions of Lemma B.3.1 are fulfilled. The application of part (i) yields:

$$\begin{aligned}\mathbb{E}\left[\langle N^{(1)} - \Lambda_\theta^{(1)}, N^{(2)} - \Lambda_\theta^{(2)}\rangle_{N^{(3)}}\right] &= \mathbb{E}\left[\langle \mathbb{E}(N^{(1)} - \Lambda_\theta^{(1)}), N^{(2)} - \Lambda_\theta^{(2)}\rangle_{N^{(3)}}\right] \\ &= \mathbb{E}\left[\langle \mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_\theta, N^{(2)} - \Lambda_\theta^{(2)}\rangle_{N^{(3)}}\right]. \qquad \text{(B.46)}\end{aligned}$$

Repeating the above steps - that $\mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_\theta$ is a continuous function according to Remark 3.6 further facilitates checking the premises of Lemma B.3.1 - results in

$$\begin{aligned}\mathbb{E}\left[\langle \mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_\theta, N^{(2)} - \Lambda_\theta^{(2)}\rangle_{N^{(3)}}\right] &= \mathbb{E}\left[\langle \mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_\theta, \mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_\theta\rangle_{N^{(3)}}\right] \\ &= \mathbb{E}\left[\|\mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_\theta\|^2_{N^{(3)}}\right]. \qquad \text{(B.47)}\end{aligned}$$

We conclude the proof by applying part (ii) of Lemma B.3.1 to obtain:

$$\begin{aligned}\mathbb{E}\left[\|\mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_\theta\|^2_{N^{(3)}}\right] &= \mathbb{E}\left[\|\mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_\theta\|^2_{\mathbb{E}\Lambda_{\theta^*}}\right] \\ &= \|\mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_\theta\|^2_{\mathbb{E}\Lambda_{\theta^*}}\,, \qquad \text{(B.48)}\end{aligned}$$

and combining Equations (B.46) through (B.48) yields the desired result. □

Before we move on to the next part of the proof, we give an extension of Lemma B.3.11 that accounts for the supremum in Equation (B.43).

Lemma B.3.12 (cf. Kopperschmidt and Stute 2013, p. 1286).
For given $\varepsilon > 0$ and $\delta > 0$, and for each $\theta \in \overline{\Theta} \setminus B_\varepsilon(\theta^*)$, there exists $r > 0$ such that

$$\left|\mathbb{E}\left[\sup_{\theta' \in B_r^\varepsilon(\theta)} \langle N^{(1)} - \Lambda_{\theta'}^{(1)}, N^{(2)} - \Lambda_{\theta'}^{(2)}\rangle_{N^{(3)}}\right] - \inf_{\theta' \in B_r^\varepsilon(\theta)} \|\mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_{\theta'}\|^2_{\mathbb{E}\Lambda_{\theta^*}}\right| \leq \delta\,, \qquad \text{(B.49)}$$

$$\left|\mathbb{E}\left[\sup_{\theta' \in B_r^\varepsilon(\theta)} \langle N^{(1)} - \Lambda_{\theta'}^{(1)}, N^{(2)} - \Lambda_{\theta'}^{(2)}\rangle_{N^{(3)}}\right] - \|\mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_\theta\|^2_{\mathbb{E}\Lambda_{\theta^*}}\right| \leq \delta\,, \qquad \text{(B.50)}$$

$$\left|\mathbb{E}\left[\inf_{\theta' \in B_r^\varepsilon(\theta)} \langle N^{(1)} - \Lambda_{\theta'}^{(1)}, N^{(2)} - \Lambda_{\theta'}^{(2)}\rangle_{N^{(3)}}\right] - \inf_{\theta' \in B_r^\varepsilon(\theta)} \|\mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_{\theta'}\|^2_{\mathbb{E}\Lambda_{\theta^*}}\right| \leq \delta\,, \qquad \text{(B.51)}$$

$$\left|\mathbb{E}\left[\inf_{\theta' \in B_r^\varepsilon(\theta)} \langle N^{(1)} - \Lambda_{\theta'}^{(1)}, N^{(2)} - \Lambda_{\theta'}^{(2)}\rangle_{N^{(3)}}\right] - \|\mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_\theta\|^2_{\mathbb{E}\Lambda_{\theta^*}}\right| \leq \delta\,. \qquad \text{(B.52)}$$

Proof. By the dominated convergence theorem using condition $(\widetilde{\mathrm{C}}3)$, the process

$$\theta \mapsto \langle N^{(1)} - \Lambda_{\theta}^{(1)}, N^{(2)} - \Lambda_{\theta}^{(2)} \rangle_{N^{(3)}}$$

is continuous due to the continuity condition (C2). Let $\varepsilon > 0$ and $\delta > 0$ be given. For any $\theta \in \overline{\Theta} \setminus \mathrm{B}_{\varepsilon}(\theta^*)$, we thus have as $r \downarrow 0$:

$$\begin{aligned}
\sup_{\theta' \in \mathrm{B}_r^{\varepsilon}(\theta)} \langle N^{(1)} - \Lambda_{\theta'}^{(1)}, N^{(2)} - \Lambda_{\theta'}^{(2)} \rangle_{N^{(3)}} \quad &\downarrow \quad \langle N^{(1)} - \Lambda_{\theta}^{(1)}, N^{(2)} - \Lambda_{\theta}^{(2)} \rangle_{N^{(3)}} \,, \\
\inf_{\theta' \in \mathrm{B}_r^{\varepsilon}(\theta)} \langle N^{(1)} - \Lambda_{\theta'}^{(1)}, N^{(2)} - \Lambda_{\theta'}^{(2)} \rangle_{N^{(3)}} \quad &\uparrow \quad \langle N^{(1)} - \Lambda_{\theta}^{(1)}, N^{(2)} - \Lambda_{\theta}^{(2)} \rangle_{N^{(3)}} \,.
\end{aligned}$$

The application of Beppo Levi's monotone convergence theorem yields in conjunction with Lemma B.3.11:

$$\begin{aligned}
&\lim_{r \downarrow 0} \mathbb{E}\left[\sup_{\theta' \in \mathrm{B}_r^{\varepsilon}(\theta)} \langle N^{(1)} - \Lambda_{\theta'}^{(1)}, N^{(2)} - \Lambda_{\theta'}^{(2)} \rangle_{N^{(3)}} \right] \\
&\quad = \mathbb{E}\left[\lim_{r \downarrow 0} \sup_{\theta' \in \mathrm{B}_r^{\varepsilon}(\theta)} \langle N^{(1)} - \Lambda_{\theta'}^{(1)}, N^{(2)} - \Lambda_{\theta'}^{(2)} \rangle_{N^{(3)}} \right] \\
&\quad = \mathbb{E}\left[\langle N^{(1)} - \Lambda_{\theta}^{(1)}, N^{(2)} - \Lambda_{\theta}^{(2)} \rangle_{N^{(3)}} \right] = \| \mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_{\theta} \|_{\mathbb{E}\Lambda_{\theta^*}}^2 \,,
\end{aligned}$$

and likewise

$$\lim_{r \downarrow 0} \mathbb{E}\left[\inf_{\theta' \in \mathrm{B}_r^{\varepsilon}(\theta)} \langle N^{(1)} - \Lambda_{\theta'}^{(1)}, N^{(2)} - \Lambda_{\theta'}^{(2)} \rangle_{N^{(3)}} \right] = \| \mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_{\theta} \|_{\mathbb{E}\Lambda_{\theta^*}}^2 \,.$$

Consequently, there exists $r_1 > 0$ such that for all $0 < r < r_1$ holds simultaneously:

$$\begin{aligned}
\left| \mathbb{E}\left[\sup_{\theta' \in \mathrm{B}_r^{\varepsilon}(\theta)} \langle N^{(1)} - \Lambda_{\theta'}^{(1)}, N^{(2)} - \Lambda_{\theta'}^{(2)} \rangle_{N^{(3)}} \right] - \| \mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_{\theta} \|_{\mathbb{E}\Lambda_{\theta^*}}^2 \right| \qquad &\leq \delta \,, \\
\left| \mathbb{E}\left[\inf_{\theta' \in \mathrm{B}_r^{\varepsilon}(\theta)} \langle N^{(1)} - \Lambda_{\theta'}^{(1)}, N^{(2)} - \Lambda_{\theta'}^{(2)} \rangle_{N^{(3)}} \right] - \| \mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_{\theta} \|_{\mathbb{E}\Lambda_{\theta^*}}^2 \right| \qquad &\leq \delta \,,
\end{aligned}$$

which proves Equations (B.50) and (B.52). The rest of the proof proceeds analogously, exploiting the continuity of the process

$$\theta \mapsto \| \mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_{\theta} \|_{\mathbb{E}\Lambda_{\theta^*}}^2 \,, \tag{B.53}$$

which can again be attributed to the continuity condition (C2) by means of the dominated convergence theorem: For any $\theta \in \overline{\Theta}$, we have by the moment condition $(\widetilde{\mathrm{C}}3)$ that

$$\begin{aligned}
\| \mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_{\theta} \|_{\mathbb{E}\Lambda_{\theta^*}}^2 &= \int_{\mathcal{I}} \left(\mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_{\theta} \right)^2 \, \mathrm{d}\mathbb{E}\Lambda_{\theta^*} \\
&\leq 2\mathbb{E}\Big[\sup_{\theta \in \overline{\Theta}} \left(\Lambda_{\theta}(\tau) \right)^2 \Big] \int_{\mathcal{I}} \mathrm{d}\mathbb{E}\Lambda_{\theta^*} \\
&= 2\mathbb{E}\Big[\sup_{\theta \in \overline{\Theta}} \left(\Lambda_{\theta}(\tau) \right)^2 \Big] \mathbb{E}\Lambda_{\theta^*}(\tau) < \infty \,,
\end{aligned}$$

and hence the continuity in Equation (B.53) results. As before, we observe for $r \downarrow 0$:

$$\inf_{\theta' \in \mathrm{B}_r^\varepsilon(\theta)} \|\mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_{\theta'}\|^2_{\mathbb{E}\Lambda_{\theta^*}} \quad \uparrow \quad \|\mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_{\theta}\|^2_{\mathbb{E}\Lambda_{\theta^*}} \,. \tag{B.54}$$

If we then choose $r_2 > 0$, such that for all $0 < r < r_2$ holds simultaneously:

$$\begin{aligned}
\left| \mathbb{E}\left[\sup_{\theta' \in \mathrm{B}_r^\varepsilon(\theta)} \langle N^{(1)} - \Lambda_{\theta'}^{(1)}, N^{(2)} - \Lambda_{\theta'}^{(2)} \rangle_{N^{(3)}} \right] - \|\mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_{\theta}\|^2_{\mathbb{E}\Lambda_{\theta^*}} \right| \qquad &\leq \frac{\delta}{2} \,, \\
\left| \mathbb{E}\left[\inf_{\theta' \in \mathrm{B}_r^\varepsilon(\theta)} \langle N^{(1)} - \Lambda_{\theta'}^{(1)}, N^{(2)} - \Lambda_{\theta'}^{(2)} \rangle_{N^{(3)}} \right] - \|\mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_{\theta}\|^2_{\mathbb{E}\Lambda_{\theta^*}} \right| \qquad &\leq \frac{\delta}{2} \,, \\
\left| \inf_{\theta' \in \mathrm{B}_r^\varepsilon(\theta)} \|\mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_{\theta'}\|^2_{\mathbb{E}\Lambda_{\theta^*}} - \|\mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_{\theta}\|^2_{\mathbb{E}\Lambda_{\theta^*}} \right| \qquad &\leq \frac{\delta}{2} \,,
\end{aligned}$$

from where Equations (B.49) and (B.51) follow by the triangle inequality. □

B.3.3. Further Steps of Proof

We begin the second part of the proof by determining the limit of the Cramér-von Mises distance used to define the minimum distance estimator $\hat{\theta}_n$. We will see that, $\mathbb{P}$-almost surely,

$$\left\| \overline{N}^{(n)} - \overline{\Lambda}_\theta^{(n)} \right\|^2_{\overline{N}^{(n)}} \longrightarrow \|\mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_{\theta}\|^2_{\mathbb{E}\Lambda_{\theta^*}} \quad (n \to \infty) \,. \tag{B.55}$$

By definition, $\hat{\theta}_n$ minimizes the left-hand side of Equation (B.55), whereas θ^* minimizes its right-hand side. We then want to conclude that the almost sure convergence carries over to the sequence $\big(\hat{\theta}_n\big)_{n \in \mathbb{N}}$ of minimum distance estimators. A few more steps are needed for this, but we start with a proof of Equation (B.55).

Lemma B.3.13 (Limit of the Cramér-von Mises Distance; Kopperschmidt and Stute 2013, p. 1286)**.**
For each $\theta \in \Theta$, $\mathbb{P}$-almost surely holds:

$$\left\| \overline{N}^{(n)} - \overline{\Lambda}_\theta^{(n)} \right\|^2_{\overline{N}^{(n)}} \longrightarrow \|\mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_{\theta}\|^2_{\mathbb{E}\Lambda_{\theta^*}} \quad (n \to \infty) \,.$$

Proof. Upon closer examination of the Cramér-von Mises distance, it becomes apparent that it is suitable for an application of Lemma B.3.10:

$$\begin{aligned}
\left\| \overline{N}^{(n)} - \overline{\Lambda}_\theta^{(n)} \right\|^2_{\overline{N}^{(n)}} &= \int_{\mathcal{I}} \left[\overline{N}^{(n)} - \overline{\Lambda}_\theta^{(n)} \right]^2 \mathrm{d}\overline{N}^{(n)} \\
&= \frac{1}{n} \sum_{k=1}^n \int_{\mathcal{I}} \left[\frac{1}{n} \sum_{i=1}^n N^{(i)} - \frac{1}{n} \sum_{i=1}^n \Lambda_\theta^{(i)} \right]^2 \mathrm{d}N^{(k)} \\
&= \frac{1}{n^3} \sum_{k=1}^n \int_{\mathcal{I}} \left[\sum_{i=1}^n \left(N^{(i)} - \Lambda_\theta^{(i)} \right) \right]^2 \mathrm{d}N^{(k)} \\
&= \frac{1}{n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \int_{\mathcal{I}} \left(N^{(i)} - \Lambda_\theta^{(i)} \right) \left(N^{(j)} - \Lambda_\theta^{(j)} \right) \mathrm{d}N^{(k)} \\
&= \frac{1}{n^3} \sum_{i,j,k=1}^n \langle N^{(i)} - \Lambda_\theta^{(i)}, N^{(j)} - \Lambda_\theta^{(j)} \rangle_{N^{(k)}} \,.
\end{aligned}$$

We split this sum into sub-sums in order to apply the pointwise versions of Equations (B.40) through (B.43):

$$\begin{aligned}
&\frac{1}{n^3}\sum_{i,j,k=1}^{n}\langle N^{(i)}-\Lambda_\theta^{(i)},N^{(j)}-\Lambda_\theta^{(j)}\rangle_{N^{(k)}}\\
&\quad=\underbrace{\frac{1}{n^3}\sum_{i=1}^{n}\left\|N^{(i)}-\Lambda_\theta^{(i)}\right\|_{N^{(i)}}^2}_{=\mathcal{O}(n)\text{ by Eq. (B.40)}} && \text{(B.56)}\\
&\quad+\underbrace{\frac{1}{n^3}\sum_{\substack{i,k=1\\ i\neq k}}^{n}\left\|N^{(i)}-\Lambda_\theta^{(i)}\right\|_{N^{(k)}}^2}_{=\mathcal{O}(n^2)\text{ by Eq. (B.41)}} && \text{(B.57)}\\
&\quad+\underbrace{\frac{2}{n^3}\sum_{\substack{i,j=1\\ i\neq j}}^{n}\langle N^{(i)}-\Lambda_\theta^{(i)},N^{(j)}-\Lambda_\theta^{(j)}\rangle_{N^{(i)}}}_{=\mathcal{O}(n^2)\text{ by Eq. (B.42)}} && \text{(B.58)}\\
&\quad+\frac{1}{n^3}\sum_{\substack{i,j,k=1\\ i\neq j\neq k\neq i}}^{n}\langle N^{(i)}-\Lambda_\theta^{(i)},N^{(j)}-\Lambda_\theta^{(j)}\rangle_{N^{(k)}}\,. && \text{(B.59)}
\end{aligned}$$

By virtue of the standardizing factor n^{-3}, applying Equations (B.40), (B.41) and (B.42) to (B.56), (B.57) and (B.58) thus yields almost sure convergence of the corresponding terms to 0. For the remaining summand in Equation (B.59), Equation (B.43) and Lemma B.3.11 provide almost surely:

$$\begin{aligned}
&\frac{1}{n^3}\sum_{\substack{i,j,k=1\\ i\neq j\neq k\neq i}}^{n}\langle N^{(i)}-\Lambda_\theta^{(i)},N^{(j)}-\Lambda_\theta^{(j)}\rangle_{N^{(k)}}\\
&=\underbrace{\frac{n(n-1)(n-2)}{n^3}}_{\to 1\ (n\to\infty)}\cdot\frac{1}{n(n-1)(n-2)}\sum_{\substack{i,j,k=1\\ i\neq j\neq k\neq i}}^{n}\langle N^{(i)}-\Lambda_\theta^{(i)},N^{(j)}-\Lambda_\theta^{(j)}\rangle_{N^{(k)}}\\
&\longrightarrow \mathbb{E}\left[\langle N^{(1)}-\Lambda_\theta^{(1)},N^{(2)}-\Lambda_\theta^{(2)}\rangle_{N^{(3)}}\right]=\|\mathbb{E}\Lambda_{\theta^*}-\mathbb{E}\Lambda_\theta\|_{\mathbb{E}\Lambda_{\theta^*}}^2\qquad(n\to\infty)\,,
\end{aligned}$$

which proves the claimed convergence of the Cramér-von Mises distance. □

In calculating the minimum distance estimator, the infimum of the Cramér-von Mises distance is needed. As a result, we require a uniform version of Lemma B.3.13. The above computations will serve as a blueprint for the essential steps of its proof.

Lemma B.3.14 (Uniform Limit of the Cramér-von Mises Distance; Kopperschmidt and Stute 2013, p. 1287)**.**
For each $\varepsilon>0$, $\mathbb{P}$-almost surely holds:

$$\inf_{\theta\,:\,\|\theta-\theta^*\|\geq\varepsilon}\left\|\overline{N}^{(n)}-\overline{\Lambda}_\theta^{(n)}\right\|_{\overline{N}^{(n)}}^2\longrightarrow\inf_{\theta\,:\,\|\theta-\theta^*\|\geq\varepsilon}\|\mathbb{E}\Lambda_{\theta^*}-\mathbb{E}\Lambda_\theta\|_{\mathbb{E}\Lambda_{\theta^*}}^2\qquad(n\to\infty)\,.$$

Proof. In order to prove Lemma B.3.14, we will show that for all $\varepsilon>0$ and all $\delta>0$ we have $\mathbb{P}$-almost surely:

$$\limsup_{n\to\infty}\left|\inf_{\theta\,:\,\|\theta-\theta^*\|\geq\varepsilon}\left\|\overline{N}^{(n)}-\overline{\Lambda}_\theta^{(n)}\right\|^2_{\overline{N}^{(n)}}-\inf_{\theta\,:\,\|\theta-\theta^*\|\geq\varepsilon}\|\mathbb{E}\Lambda_{\theta^*}-\mathbb{E}\Lambda_\theta\|^2_{\mathbb{E}\Lambda_{\theta^*}}\right|\leq\delta\,. \tag{B.60}$$

Let therefore $\varepsilon>0$ and $\delta>0$ be given. By Lemma B.3.12, for each $\theta\in\overline{\Theta}\setminus\mathrm{B}_\varepsilon(\theta^*)$ there exists $r=r(\theta)>0$ such that the Equations (B.49) and (B.51) hold simultaneously. Then,

$$\bigcup_{\theta\in\overline{\Theta}\setminus\mathrm{B}_\varepsilon(\theta^*)}\mathrm{B}^\varepsilon_{r(\theta)}(\theta)=\overline{\Theta}\setminus\mathrm{B}_\varepsilon(\theta^*)$$

is an open cover of the compact set $\overline{\Theta}\setminus\mathrm{B}_\varepsilon(\theta^*)$. Consequently, there exists a finite subcover and we can find $\theta_1,\dots,\theta_q\in\overline{\Theta}\setminus\mathrm{B}_\varepsilon(\theta^*)$ such that

$$\bigcup_{p=1}^q\mathrm{B}^\varepsilon_{r(\theta_p)}(\theta_p)=\overline{\Theta}\setminus\mathrm{B}_\varepsilon(\theta^*)\ .$$

We set $r_p:=r(\theta_p)$ and deduce:

$$\begin{aligned}
&\left|\inf_{\theta\,:\,\|\theta-\theta^*\|\geq\varepsilon}\left\|\overline{N}^{(n)}-\overline{\Lambda}_\theta^{(n)}\right\|^2_{\overline{N}^{(n)}}-\inf_{\theta\,:\,\|\theta-\theta^*\|\geq\varepsilon}\|\mathbb{E}\Lambda_{\theta^*}-\mathbb{E}\Lambda_\theta\|^2_{\mathbb{E}\Lambda_{\theta^*}}\right|\\
&\quad=\left|\min_{1\leq p\leq q}\inf_{\theta\in\mathrm{B}^\varepsilon_{r_p}(\theta_p)}\left\|\overline{N}^{(n)}-\overline{\Lambda}_\theta^{(n)}\right\|^2_{\overline{N}^{(n)}}-\min_{1\leq p\leq q}\inf_{\theta\in\mathrm{B}^\varepsilon_{r_p}(\theta_p)}\|\mathbb{E}\Lambda_{\theta^*}-\mathbb{E}\Lambda_\theta\|^2_{\mathbb{E}\Lambda_{\theta^*}}\right|\\
&\quad\leq\max_{1\leq p\leq q}\left|\inf_{\theta\in\mathrm{B}^\varepsilon_{r_p}(\theta_p)}\left\|\overline{N}^{(n)}-\overline{\Lambda}_\theta^{(n)}\right\|^2_{\overline{N}^{(n)}}-\inf_{\theta\in\mathrm{B}^\varepsilon_{r_p}(\theta_p)}\|\mathbb{E}\Lambda_{\theta^*}-\mathbb{E}\Lambda_\theta\|^2_{\mathbb{E}\Lambda_{\theta^*}}\right|\,,
\end{aligned}\tag{B.61}$$

where we used that for any positive real numbers $a_1,\dots,a_q$ and $b_1,\dots,b_q$ we have:

$$\left|\min_{1\leq p\leq q}a_p-\min_{1\leq p\leq q}b_p\right|\leq\max_{1\leq p\leq q}|a_p-b_p|\ .$$

For this, without loss of generality, let $\min\limits_{1\leq p\leq q}a_p\geq\min\limits_{1\leq p\leq q}b_p$ and $p_0=\arg\min\limits_{1\leq p\leq q}b_p$, so that

$$\left|\min_{1\leq p\leq q}a_p-\min_{1\leq p\leq q}b_p\right|=\min_{1\leq p\leq q}a_p-b_{p_0}\leq a_{p_0}-b_{p_0}=|a_{p_0}-b_{p_0}|\leq\max_{1\leq p\leq q}|a_p-b_p|\ .$$

From Equation (B.61) we can conclude that Equation (B.60) is satisfied if for all $1\leq p\leq q$ holds $\mathbb{P}$-almost surely:

$$\limsup_{n\to\infty}\left|\inf_{\theta\in\mathrm{B}^\varepsilon_{r_p}(\theta_p)}\left\|\overline{N}^{(n)}-\overline{\Lambda}_\theta^{(n)}\right\|^2_{\overline{N}^{(n)}}-\inf_{\theta\in\mathrm{B}^\varepsilon_{r_p}(\theta_p)}\|\mathbb{E}\Lambda_{\theta^*}-\mathbb{E}\Lambda_\theta\|^2_{\mathbb{E}\Lambda_{\theta^*}}\right|\leq\delta\,. \tag{B.62}$$

Retracing the proof of Lemma B.3.13, we get $\mathbb{P}$-almost surely:

$$\begin{aligned}
\inf_{\theta\in\mathrm{B}^\varepsilon_{r_p}(\theta_p)}\left\|\overline{N}^{(n)}-\overline{\Lambda}_\theta^{(n)}\right\|^2_{\overline{N}^{(n)}}&=\inf_{\theta\in\mathrm{B}^\varepsilon_{r_p}(\theta_p)}\frac{1}{n^3}\sum_{i,j,k=1}^n\langle N^{(i)}-\Lambda_\theta^{(i)},N^{(j)}-\Lambda_\theta^{(j)}\rangle_{N^{(k)}}\\
&\leq\frac{1}{n^3}\sum_{i,j,k=1}^n\sup_{\theta\in\mathrm{B}^\varepsilon_{r_p}(\theta_p)}\langle N^{(i)}-\Lambda_\theta^{(i)},N^{(j)}-\Lambda_\theta^{(j)}\rangle_{N^{(k)}}
\end{aligned}$$

$$\leq \frac{1}{n(n-1)(n-2)} \sum_{i \neq j \neq k \neq i} \sup_{\theta \in \mathrm{B}^{\varepsilon}_{r_p}(\theta_p)} \langle N^{(i)} - \Lambda^{(i)}_{\theta}, N^{(j)} - \Lambda^{(j)}_{\theta} \rangle_{N^{(k)}}$$
$$+ \frac{1}{n^3} \mathcal{O}(n^2)$$
$$\longrightarrow \mathbb{E}\left[\sup_{\theta \in \mathrm{B}^{\varepsilon}_{r_p}(\theta_p)} \langle N^{(1)} - \Lambda^{(1)}_{\theta}, N^{(2)} - \Lambda^{(2)}_{\theta} \rangle_{N^{(3)}} \right] \quad (n \to \infty)\,,$$

where once again Lemma B.3.10 grants that the sub-sum over partially matching indices is of order n^2 and hence negligible. Analogously, the additional statement of Lemma B.3.10 regarding the infimum yields:

$$\begin{aligned}
\inf_{\theta \in \mathrm{B}^{\varepsilon}_{r_p}(\theta_p)} \left\| \overline{N}^{(n)} - \overline{\Lambda}^{(n)}_{\theta} \right\|^2_{\overline{N}^{(n)}} &= \inf_{\theta \in \mathrm{B}^{\varepsilon}_{r_p}(\theta_p)} \frac{1}{n^3} \sum_{i,j,k=1}^{n} \langle N^{(i)} - \Lambda^{(i)}_{\theta}, N^{(j)} - \Lambda^{(j)}_{\theta} \rangle_{N^{(k)}} \\
&\geq \frac{1}{n^3} \sum_{i,j,k=1}^{n} \inf_{\theta \in \mathrm{B}^{\varepsilon}_{r_p}(\theta_p)} \langle N^{(i)} - \Lambda^{(i)}_{\theta}, N^{(j)} - \Lambda^{(j)}_{\theta} \rangle_{N^{(k)}} \\
&\longrightarrow \mathbb{E}\left[\inf_{\theta \in \mathrm{B}^{\varepsilon}_{r_p}(\theta_p)} \langle N^{(1)} - \Lambda^{(1)}_{\theta}, N^{(2)} - \Lambda^{(2)}_{\theta} \rangle_{N^{(3)}} \right] \quad (n \to \infty)\,.
\end{aligned}$$

Since the radii $r_1, \ldots, r_q$ were chosen appropriately so that the Equations (B.49) and (B.51) are satisfied, the combination of the above estimates provides

$$\begin{aligned}
&\inf_{\theta \in \mathrm{B}^{\varepsilon}_{r_p}(\theta_p)} \|\mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_{\theta}\|^2_{\mathbb{E}\Lambda_{\theta^*}} - \delta \\
&\leq \mathbb{E}\left[\inf_{\theta \in \mathrm{B}^{\varepsilon}_{r_p}(\theta_p)} \langle N^{(1)} - \Lambda^{(1)}_{\theta}, N^{(2)} - \Lambda^{(2)}_{\theta} \rangle_{N^{(3)}} \right] \\
&\leq \liminf_{n \to \infty} \inf_{\theta \in \mathrm{B}^{\varepsilon}_{r_p}(\theta_p)} \left\| \overline{N}^{(n)} - \overline{\Lambda}^{(n)}_{\theta} \right\|^2_{\overline{N}^{(n)}} \\
&\leq \limsup_{n \to \infty} \inf_{\theta \in \mathrm{B}^{\varepsilon}_{r_p}(\theta_p)} \left\| \overline{N}^{(n)} - \overline{\Lambda}^{(n)}_{\theta} \right\|^2_{\overline{N}^{(n)}} \\
&\leq \mathbb{E}\left[\sup_{\theta \in \mathrm{B}^{\varepsilon}_{r_p}(\theta_p)} \langle N^{(1)} - \Lambda^{(1)}_{\theta}, N^{(2)} - \Lambda^{(2)}_{\theta} \rangle_{N^{(3)}} \right] \\
&\leq \inf_{\theta \in \mathrm{B}^{\varepsilon}_{r_p}(\theta_p)} \|\mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_{\theta}\|^2_{\mathbb{E}\Lambda_{\theta^*}} + \delta\,,
\end{aligned}$$

and in particular we obtain:

$$\begin{aligned}
-\delta &\leq \liminf_{n \to \infty} \left(\inf_{\theta \in \mathrm{B}^{\varepsilon}_{r_p}(\theta_p)} \left\| \overline{N}^{(n)} - \overline{\Lambda}^{(n)}_{\theta} \right\|^2_{\overline{N}^{(n)}} - \inf_{\theta \in \mathrm{B}^{\varepsilon}_{r_p}(\theta_p)} \|\mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_{\theta}\|^2_{\mathbb{E}\Lambda_{\theta^*}} \right) \\
&\leq \limsup_{n \to \infty} \left(\inf_{\theta \in \mathrm{B}^{\varepsilon}_{r_p}(\theta_p)} \left\| \overline{N}^{(n)} - \overline{\Lambda}^{(n)}_{\theta} \right\|^2_{\overline{N}^{(n)}} - \inf_{\theta \in \mathrm{B}^{\varepsilon}_{r_p}(\theta_p)} \|\mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_{\theta}\|^2_{\mathbb{E}\Lambda_{\theta^*}} \right) \leq \delta\,,
\end{aligned}$$

which in turn implies Equation (B.62), completing the proof. □

B.3.4. Completion of the Consistency Proof

Lemma B.3.14 paved the way for the proof of Theorem 3.7. We now combine the previous results to infer the consistency of the minimum distance estimator.

Proof of Theorem 3.7. Let $\varepsilon > 0$ be given. By definition of the minimum distance estimator $\hat{\theta}_n$, $\|\hat{\theta}_n - \theta^*\| \geq \varepsilon$ is valid only if the infimum of the function

$$\theta \longmapsto \left\| \overline{N}^{(n)} - \overline{\Lambda}_\theta^{(n)} \right\|_{\overline{N}^{(n)}}$$

over $\overline{\Theta} \setminus \mathrm{B}_\varepsilon(\theta^*)$ is smaller than its infimum over $\mathrm{B}_\varepsilon(\theta^*)$. We hence obtain the inclusions:

$$\begin{aligned}\left\{\|\hat{\theta}_n - \theta^*\| \geq \varepsilon\right\} &\subset \Bigg\{ \inf_{\theta\,:\,\|\theta-\theta^*\|\geq\varepsilon} \left\| \overline{N}^{(n)} - \overline{\Lambda}_\theta^{(n)} \right\|_{\overline{N}^{(n)}} < \underbrace{\inf_{\theta\,:\,\|\theta-\theta^*\|<\varepsilon} \left\| \overline{N}^{(n)} - \overline{\Lambda}_\theta^{(n)} \right\|_{\overline{N}^{(n)}}}_{\leq \left\| \overline{N}^{(n)} - \overline{\Lambda}_{\theta^*}^{(n)} \right\|_{\overline{N}^{(n)}}} \Bigg\} \\ &\subset \Bigg\{ \inf_{\theta\,:\,\|\theta-\theta^*\|\geq\varepsilon} \left\| \overline{N}^{(n)} - \overline{\Lambda}_\theta^{(n)} \right\|_{\overline{N}^{(n)}} < \left\| \overline{N}^{(n)} - \overline{\Lambda}_{\theta*}^{(n)} \right\|_{\overline{N}^{(n)}} \Bigg\}. \end{aligned} \tag{B.63}$$

According to Lemma B.3.14 and the identifiability condition (C1), $\mathbb{P}$-almost surely holds:

$$\inf_{\theta\,:\,\|\theta-\theta^*\|\geq\varepsilon} \left\| \overline{N}^{(n)} - \overline{\Lambda}_\theta^{(n)} \right\|_{\overline{N}^{(n)}} \longrightarrow \inf_{\theta\,:\,\|\theta-\theta^*\|\geq\varepsilon} \|\mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_\theta\|^2_{\mathbb{E}\Lambda_{\theta^*}} =: \delta > 0 \quad (n \to \infty),$$

while Lemma B.3.13 yields that

$$\left\| \overline{N}^{(n)} - \overline{\Lambda}_{\theta*}^{(n)} \right\|_{\overline{N}^{(n)}} \longrightarrow 0 \quad (n \to \infty). \tag{B.64}$$

The combination of these limits then ensures that the set considered in Equation (B.63) converges to a $\mathbb{P}$-null set. Formally, we first compute:

$$\begin{aligned} &\Bigg\{ \inf_{\theta\,:\,\|\theta-\theta^*\|\geq\varepsilon} \left\| \overline{N}^{(n)} - \overline{\Lambda}_\theta^{(n)} \right\|_{\overline{N}^{(n)}} < \left\| \overline{N}^{(n)} - \overline{\Lambda}_{\theta*}^{(n)} \right\|_{\overline{N}^{(n)}} \Bigg\} \\ &= \Bigg\{ \inf_{\theta\,:\,\|\theta-\theta^*\|\geq\varepsilon} \left\| \overline{N}^{(n)} - \overline{\Lambda}_\theta^{(n)} \right\|_{\overline{N}^{(n)}} < \left\| \overline{N}^{(n)} - \overline{\Lambda}_{\theta*}^{(n)} \right\|_{\overline{N}^{(n)}}, \left\| \overline{N}^{(n)} - \overline{\Lambda}_{\theta*}^{(n)} \right\|_{\overline{N}^{(n)}} \leq \frac{\delta}{2} \Bigg\} \\ &\quad \cup \Bigg\{ \inf_{\theta\,:\,\|\theta-\theta^*\|\geq\varepsilon} \left\| \overline{N}^{(n)} - \overline{\Lambda}_\theta^{(n)} \right\|_{\overline{N}^{(n)}} < \left\| \overline{N}^{(n)} - \overline{\Lambda}_{\theta*}^{(n)} \right\|_{\overline{N}^{(n)}}, \left\| \overline{N}^{(n)} - \overline{\Lambda}_{\theta*}^{(n)} \right\|_{\overline{N}^{(n)}} > \frac{\delta}{2} \Bigg\} \\ &\subset \Bigg\{ \inf_{\theta\,:\,\|\theta-\theta^*\|\geq\varepsilon} \left\| \overline{N}^{(n)} - \overline{\Lambda}_\theta^{(n)} \right\|_{\overline{N}^{(n)}} < \frac{\delta}{2} \Bigg\} \cup \left\{ \left\| \overline{N}^{(n)} - \overline{\Lambda}_{\theta*}^{(n)} \right\|_{\overline{N}^{(n)}} > \frac{\delta}{2} \right\}. \end{aligned} \tag{B.65}$$

For the second set from Equation (B.65), the $\mathbb{P}$-almost sure convergence given in Equation (B.64) yields:

$$\mathbb{P}\left(\limsup_{n\to\infty} \left\{ \left\| \overline{N}^{(n)} - \overline{\Lambda}_{\theta*}^{(n)} \right\|_{\overline{N}^{(n)}} > \frac{\delta}{2} \right\} \right) = 0. \tag{B.66}$$

To deal with the remaining set from Equation (B.65), we subtract δ on both sides and conclude:

$$\Bigg\{ \inf_{\theta\,:\,\|\theta-\theta^*\|\geq\varepsilon} \left\| \overline{N}^{(n)} - \overline{\Lambda}_\theta^{(n)} \right\|_{\overline{N}^{(n)}} < \frac{\delta}{2} \Bigg\}$$

$$= \Big\{ \inf_{\theta\,:\,\|\theta-\theta^*\|\geq\varepsilon} \left\|\overline{N}^{(n)} - \overline{\Lambda}_\theta^{(n)}\right\|_{\overline{N}^{(n)}} - \underbrace{\inf_{\theta\,:\,\|\theta-\theta^*\|\geq\varepsilon} \|\mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_\theta\|^2_{\mathbb{E}\Lambda_{\theta^*}}}_{=\delta \text{ by definition.}} < -\frac{\delta}{2} \Big\}$$

$$\subset \Big\{ \Big| \inf_{\theta\,:\,\|\theta-\theta^*\|\geq\varepsilon} \left\|\overline{N}^{(n)} - \overline{\Lambda}_\theta^{(n)}\right\|_{\overline{N}^{(n)}} - \inf_{\theta\,:\,\|\theta-\theta^*\|\geq\varepsilon} \|\mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_\theta\|^2_{\mathbb{E}\Lambda_{\theta^*}} \Big| > \frac{\delta}{2} \Big\}.$$

Reapplying Lemma B.3.14 then provides:

$$\mathbb{P}\left(\limsup_{n\to\infty} \Big\{ \Big| \inf_{\theta\,:\,\|\theta-\theta^*\|\geq\varepsilon} \left\|\overline{N}^{(n)} - \overline{\Lambda}_\theta^{(n)}\right\|_{\overline{N}^{(n)}} - \inf_{\theta\,:\,\|\theta-\theta^*\|\geq\varepsilon} \|\mathbb{E}\Lambda_{\theta^*} - \mathbb{E}\Lambda_\theta\|^2_{\mathbb{E}\Lambda_{\theta^*}} \Big| > \frac{\delta}{2} \Big\}\right) = 0\,. \tag{B.67}$$

From here, by consecutively employing Equations (B.63), (B.65), (B.66) and (B.67), we deduce that

$$\mathbb{P}\left(\limsup_{n\to\infty} \Big\{ \|\hat{\theta}_n - \theta^*\| \geq \varepsilon \Big\}\right) = 0\,,$$

which proves the consistency of the minimum distance estimator. □

B.4. Complementary Notes for the Proof of the Asymptotic Normality of the Minimum Distance Estimator

B.4.1. Applications of the Continuous Mapping Theorem

When studying the asymptotic behavior of estimators, we often deal with transformations of random variable sequences with known limits. For example, if g is a continuous function, the continuous mapping theorem (see Billingsley 1968, p. 31) states that it follows from $X_n \overset{\mathbb{P}}{\to} X$ that $g(X_n) \overset{\mathbb{P}}{\to} g(X)$. However, this apparently requires that the admissible domain of the function g includes the range of the X_n, which in practice may not be satisfied: Take for instance a sequence of random matrices in $\mathbb{R}^{d\times d}$ and consider the inversion mapping on the open subspace of invertible matrices, that is, the general linear group $\mathrm{GL}_d(\mathbb{R})$. While said mapping is continuous, the continuous mapping theorem cannot be applied unless the invertibility of the X_n can be ensured (which is not possible in general due to stochastic convergence). Nevertheless, if the limit X is (almost certainly) invertible, it is plausible that this should also hold for X_n as long as n is sufficiently large - at least in a stochastic sense. For the special case of X being deterministic, we show this intuition to be true in Corollary B.4.2. Before doing so, a more general result is formulated and proved in the following Lemma.

Lemma B.4.1 (Continuous Mapping Theorem for Functions of Restricted Domain)**.** *Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of random variables taking values in a metric space S. Suppose that*

$$A_n \xrightarrow{\mathbb{P}} A \quad \text{as } n\to\infty\,, \tag{B.68}$$

where $A\in S$ is deterministic. Furthermore, let T denote another metric space and for $S'\subset S$ consider a Borel-measurable function $g : S'\to T$. If A is an inner point of S' and g is continuous at A, then

$$\tilde{g}(A_n) := g(A_n)\cdot \mathbb{1}_{\{A_n\in S'\}} \xrightarrow{\mathbb{P}} g(A) \quad \text{as } n\to\infty\,. \tag{B.69}$$

Proof. First, it should be mentioned that the notation used in Equation (B.69) is not mathematically rigorous, but is considered an abbreviated version of the formally preferable case distinction

$$\tilde{g}(A_n) := \begin{cases} g(A_n) , & A_n \in S' , \\ 0 , & A_n \notin S' . \end{cases} \tag{B.70}$$

It can be seen from Equation (B.70) that $\tilde{g}$ extends the function g to S by setting $\tilde{g} \equiv 0$ on $S \setminus S'$. The particular value 0 is chosen arbitrarily and has no further meaning for us. Let d_S and d_T denote the metrics on S and T, respectively. Since A is an inner point of S', there exists $\delta > 0$ such that $\mathrm{B}_\delta(A) \subset S'$. Accordingly,

$$\mathbb{P}(A_n \notin S') \leq \mathbb{P}(A_n \notin \mathrm{B}_\delta(A)) = \mathbb{P}(d_S(A_n, A) \geq \delta) \overset{n\to\infty}{\longrightarrow} 0 , \tag{B.71}$$

by the assumption of Equation (B.68). Let $\varepsilon > 0$ and observe:

$$\begin{aligned} \mathbb{P}(d_T(\tilde{g}(A_n), g(A)) \geq \varepsilon) &= \mathbb{P}(d_T(\tilde{g}(A_n), g(A)) \geq \varepsilon , A_n \in S') \\ &\quad + \mathbb{P}(d_T(\tilde{g}(A_n), g(A)) \geq \varepsilon , A_n \notin S') \\ &\leq \mathbb{P}(d_T(g(A_n), g(A)) \geq \varepsilon , A_n \in S') + \mathbb{P}(A_n \notin S') . \end{aligned} \tag{B.72}$$

By Equation (B.71), the second summand of Equation (B.72) tends to 0 as $n \to \infty$. For the remainder, one proceeds as in the continuous mapping theorem by utilizing the continuity of g at A. For each given $\varepsilon > 0$, there again exists $\delta > 0$ such that for all $A' \in S'$ we have:

$$d_S(A', A) < \delta \quad \Longrightarrow \quad d_T(g(A'), g(A)) < \varepsilon .$$

Substituting the contrapositive into Equation (B.72) yields:

$$\begin{aligned} \mathbb{P}(d_T(g(A_n), g(A)) \geq \varepsilon , A_n \in S') &\leq \mathbb{P}(d_S(A_n, A) \geq \delta , A_n \in S') \\ &\leq \mathbb{P}(d_S(A_n, A) \geq \delta) \overset{n\to\infty}{\longrightarrow} 0 , \end{aligned}$$

and hence the desired result. □

We now return to the application of Lemma B.4.1 indicated earlier.

Corollary B.4.2 (Limit of an Inverse Matrix Sequence)**.**
Let $d \in \mathbb{N}$ and consider a sequence of random matrices $(A_n)_{n\in\mathbb{N}}$ in $\mathbb{R}^{d\times d}$. Suppose that $(A_n)_{n\in\mathbb{N}}$ converges in probability to an invertible (deterministic) matrix $A \in \mathbb{R}^{d\times d}$. Then,

$$A_n^{-1} \cdot \mathbb{1}_{\{A_n \text{ is invertible.}\}} \overset{\mathbb{P}}{\longrightarrow} A^{-1} \quad \textit{as } n \to \infty . \tag{B.73}$$

Proof. The proof boils down to a simple application of the continuous mapping theorem for functions of restricted domain. For this, let $S = T = \mathbb{R}^{d\times d}$, $S' = \mathrm{GL}_d(\mathbb{R})$ and g be the inversion mapping given by

$$g \,:\, \mathrm{GL}_d(\mathbb{R}) \longrightarrow \mathrm{GL}_d(\mathbb{R}) \,:\, A \longmapsto A^{-1} .$$

Since S and T are required to be metric spaces, we equip the $\mathbb{R}$-vector field $\mathbb{R}^{d\times d}$ with an arbitrary matrix norm to obtain a normed vector space. By the equivalence of norms, the particular choice of norm does not matter, so we opt for the max norm once again. In order to apply Lemma B.4.1, it suffices to verify that (a) $S' = \mathrm{GL}_d(\mathbb{R})$ is an open subset of $\mathbb{R}^{d\times d}$ (and hence every invertible matrix is an inner point of S') and (b) g is continuous (which implies Borel-measurability as well as continuity at A).

(a) We start by proving that the determinant as a function from $\mathbb{R}^{d\times d}$ to $\mathbb{R}$ is continuous. For each $1 \leq i, j \leq d$, the coefficient mapping

$$\pi_{ij} \,:\, \mathbb{R}^{d\times d} \longrightarrow \mathbb{R} \,:\, A = (a_{kl})_{1\leq k,l\leq d} \longmapsto a_{ij}$$

is Lipschitz continuous with Lipschitz constant 1:

$$\left|\pi_{ij}(A) - \pi_{ij}(A')\right| = \left|a_{ij} - a'_{ij}\right| \leq \max_{1\leq k,l\leq d} \left|a_{kl} - a'_{kl}\right| = \left\|A - A'\right\|_{\max} .$$

Then, the mapping

$$\pi \,:\, \mathbb{R}^{d\times d} \longrightarrow \mathbb{R}^{d^2} \,:\, A \longmapsto (\pi_{11}(A), \dots, \pi_{1d}(A), \pi_{21}(A), \dots, \pi_{dd}(A))$$

is continuous as well, since all its components are continuous. We recall that multivariate polynomials on $\mathbb{R}^{d^2}$ are continuous, and since the Leibniz formula shows that the determinant of a matrix is a polynomial in its d^2 components, this is also true for the determinant mapping. But then we have

$$\mathrm{GL}_d(\mathbb{R}) = \det{}^{-1}\left(\mathbb{R} \setminus \{0\}\right) ,$$

where $\mathbb{R} \setminus \{0\}$ is an open subset of $\mathbb{R}$, so we can conclude that the general linear group is an open subset of $\mathbb{R}^{d\times d}$.

(b) The continuity of the inversion mapping is a direct consequence of (a), where the continuity of the determinant was proved. For this, remember that for any $A \in \mathrm{GL}_d(\mathbb{R})$ it holds:

$$g(A) = A^{-1} = \frac{1}{\det A}\, \mathrm{adj}(A) ,$$

where $\mathrm{adj}(A)$ is the adjugate of A. The entries of this adjugate matrix consist of cofactors of A, which in turn can be represented as determinants of sub-matrices of A, called minors. The continuity of g then follows as a composition of continuous mappings.

□

B.4.2. Second Moment Bounds for Sums of Multi-Indexed Random Vectors

In the following lemmas, we consider multi-indexed random vectors under the assumption of square-integrability and the vanishing of certain mixed moments. The latter - and fairly specific - requirement is tailored to an application in proving the asymptotic normality of the minimum distance estimator in Chapter 3, where we exploit that the convergence in quadratic mean implies the convergence in probability due to the Markov inequality (cf. Brémaud 2020, p. 189). Since the original but flawed proof of Kopperschmidt and Stute 2013 utilizes the presented results in a similar way, we consequently find them formulated and proved in Kopperschmidt and Stute 2013, pp. 1295–1297, see Lemmas 15 to 17. For completeness, we reproduce here the detailed and revised proofs given in the master's thesis of the author, see Jakubzik 2017, pp. 144–148. They all require a simple inequality, which we would like to note beforehand.

Lemma B.4.3 (Auxiliary Inequality for the Square Norm of Sums).
For any norm $\|\cdot\|$ *on* $\mathbb{R}^d$ *and* $a_1, \ldots, a_l \in \mathbb{R}^d$ *it holds:*

$$\Big\| \sum_{j=1}^{l} a_j \Big\|^2 \leq 2^{l-1} \sum_{j=1}^{l} \|a_j\|^2 \ .$$

Proof. The lemma is a generalization of the inequality $(a+b)^2 \leq 2\left(a^2+b^2\right)$ obtained in the case $d=1$ and $l=2$. We observe for arbitrary $a, b \in \mathbb{R}$:

$$\begin{aligned}
& & 0 \leq (a-b)^2 &= a^2 - 2ab + b^2 \\
&\iff & 2ab &\leq a^2 + b^2 \\
&\iff & a^2 + 2ab + b^2 &\leq 2a^2 + 2b^2 \\
&\iff & (a+b)^2 &\leq 2\left(a^2+b^2\right) \ .
\end{aligned} \tag{B.74}$$

Because of the triangle equality, this easily extends to $d > 1$, as we have for $a, b \in \mathbb{R}^d$:

$$\|a+b\|^2 \leq (\|a\| + \|b\|)^2 \overset{(B.74)}{\leq} 2\left(\|a\|^2 + \|b\|^2\right) \ .$$

The statement for $l > 2$ follows by induction, with $l = 2$ serving as the base case: Suppose it holds for some $l \in \mathbb{N} \setminus \{1\}$. Then, we conclude:

$$\begin{aligned}
\Big\| \sum_{j=1}^{l+1} a_j \Big\|^2 &\leq \left(\|a_{l+1}\| + \Big\| \sum_{j=1}^{l} a_j \Big\| \right)^2 \leq 2\left(\|a_{l+1}\|^2 + \Big\| \sum_{j=1}^{l} a_j \Big\|^2 \right) \\
&\leq 2\left(\|a_{l+1}\|^2 + 2^{l-1} \sum_{j=1}^{l} \|a_j\|^2 \right) \leq 2\left(2^{l-1} \|a_{l+1}\|^2 + 2^{l-1} \sum_{j=1}^{l} \|a_j\|^2 \right) \\
&= 2^l \sum_{j=1}^{l+1} \|a_j\|^2 \ ,
\end{aligned}$$

and thus the statement holds for $l+1$, completing the proof. □

We will now successively state and prove the Lemmas 15 to 17 from Kopperschmidt and Stute 2013. Throughout this subsection, we will assume $\|\cdot\|$ to be the Euclidean norm on $\mathbb{R}^d$.

Lemma B.4.4 (Lemma 15 of Kopperschmidt and Stute 2013, p. 1295).
For $p, k \in \{1, \ldots, n\}$, *let* U_{pk} *be d-variate random vectors with* $\mathbb{E}\,\|U_{pk}\|^2 < \infty$. *If*

$$\mathbb{E}\left[U_{pk}^\top U_{ql}\right] = 0 \quad \textit{for } k \notin \{p,q,l\} \textit{ or } l \notin \{p,q,k\} \ , \tag{B.75}$$

then:

$$\mathbb{E}\Big\| \sum_{\substack{p,k=1 \\ p \neq k}}^{n} U_{pk} \Big\|^2 \leq 2 \sum_{k \gtrless p,q} \mathbb{E}\left[U_{pk}^\top U_{qk}\right] \ . \tag{B.76}$$

Proof. An application of Lemma B.4.3 in conjunction with the monotonicity and linearity of the expectation yields:

$$\mathbb{E}\Big\| \sum_{\substack{p,k=1\\ p\neq k}}^{n} U_{pk} \Big\|^2 = \mathbb{E}\Big\| \sum_{\substack{p,k=1\\ k<p}}^{n} U_{pk} + \sum_{\substack{p,k=1\\ k>p}}^{n} U_{pk} \Big\|^2 \leq \mathbb{E}\left[\Big\| \sum_{k<p}^{n} U_{pk} \Big\| + \Big\| \sum_{k>p}^{n} U_{pk} \Big\| \right]^2$$

$$\leq 2\mathbb{E}\Big\| \sum_{k<p} U_{pk} \Big\|^2 + 2\mathbb{E}\Big\| \sum_{k>p} U_{pk} \Big\|^2 . \tag{B.77}$$

By virtue of Equation (B.75), we then obtain:

$$\mathbb{E}\Big\| \sum_{k<p} U_{pk} \Big\|^2 = \mathbb{E}\left[\Big(\sum_{k<p} U_{pk} \Big)^\top \Big(\sum_{l<q} U_{ql} \Big) \right] = \sum_{k<p} \sum_{l<q} \mathbb{E}\left[U_{pk}^\top U_{ql} \right]$$

$$= \sum_{k<p,q} \mathbb{E}\left[U_{pk}^\top U_{qk} \right] , \tag{B.78}$$

since $k \neq l$ readily implies $k \notin \{p, q, l\}$ or $l \notin \{p, q, k\}$, which can be seen as follows:

- If $k < l$, then $k < q$ because of $l < q$ and thus $k \neq p, q, l$.
- If $k > l$, then $l < p$ because of $k < p$ and thus $l \neq p, q, k$.

Similarly, we receive:

$$\mathbb{E}\Big\| \sum_{k>p} U_{pk} \Big\|^2 = \sum_{k>p,q} \mathbb{E}\left[U_{pk}^\top U_{qk} \right] , \tag{B.79}$$

and substituting Equations (B.78) and (B.79) into Equation (B.77) finishes the proof. □

Lemma B.4.5 (Lemma 16 of Kopperschmidt and Stute 2013, p. 1295)**.**
For $p, k \in \{1, \ldots, n\}$, let U_{pk} be d-variate random vectors with $\mathbb{E}\, \|U_{pk}\|^2 < \infty$. If

$$\mathbb{E}\left[U_{pk}^\top U_{ql} \right] = 0 \qquad \textit{whenever one index differs from the rest,} \tag{B.80}$$

then:

$$\mathbb{E}\Big\| \sum_{\substack{p,k=1\\ p\neq k}}^{n} U_{pk} \Big\|^2 \leq 2 \sum_{\substack{p,k=1\\ p\neq k}}^{n} \mathbb{E}\, \|U_{pk}\|^2 . \tag{B.81}$$

Proof. We again start from Equation (B.77), which holds regardless of the assumptions given in Lemma B.4.4. Since the condition (B.80) implies (B.75), the Equations (B.78) and (B.79) remain valid. From there, we further compute:

$$\mathbb{E}\Big\| \sum_{k<p} U_{pk} \Big\|^2 = \sum_{k<p,q} \underbrace{\mathbb{E}\left[U_{pk}^\top U_{qk} \right]}_{=0 \text{ for } p\neq q\,.}$$

$$= \sum_{k<p} \mathbb{E}\left[U_{pk}^\top U_{pk} \right] = \sum_{k<p} \mathbb{E}\, \|U_{pk}\|^2 , \tag{B.82}$$

and likewise for $k > p$. The assertion then follows immediately. □

Lemma B.4.6 (Lemma 17 of Kopperschmidt and Stute 2013, p. 1296)**.**
For $p, k, i \in \{1, \dots, n\}$, let U_{pki} be d-variate random vectors with $\mathbb{E}\,\|U_{pki}\|^2 < \infty$. If

$$\mathbb{E}\left[U_{pki}^\top U_{qlj}\right] = 0 \qquad \textit{whenever } k,\ i,\ l \textit{ or } j \textit{ differs from the rest,} \tag{B.83}$$

then[68]*:*

$$\mathbb{E}\Bigg\| \sum_{\substack{p,k,i=1\\ p\neq k\neq i\neq p}}^{n} U_{pki} \Bigg\|^2 \leq 32 \sum \mathbb{E}\left[U_{pki}^\top U_{qki}\right] , \tag{B.84}$$

where the summation on the right takes place over all index combinations in which p and q have the same position relative to i and k, that is:

$$\begin{aligned} &p,q < k < i\,, \qquad &&p,q < i < k\,, \qquad &&k < p,q < i\,,\\ &i < p,q < k\,, \qquad &&k < i < p,q\,, \qquad &&i < k < p,q\,. \end{aligned}$$

Proof. Similar to Equation (B.77), applying Lemma B.4.3 in the first step yields:

$$\begin{aligned} \mathbb{E}\Bigg\| \sum_{\substack{p,k,i=1\\ p\neq k\neq i\neq p}}^{n} U_{pki} \Bigg\|^2 &= \mathbb{E}\Bigg\| \sum_{p<k<i} U_{pki} + \sum_{p<i<k} U_{pki} + \sum_{k<p<i} U_{pki} \\ &\qquad + \sum_{i<p<k} U_{pki} + \sum_{k<i<p} U_{pki} + \sum_{i<k<p} U_{pki} \Bigg\|^2 \\ &\leq 32 \Bigg[\mathbb{E}\Bigg\| \sum_{p<k<i} U_{pki} \Bigg\|^2 + \mathbb{E}\Bigg\| \sum_{p<i<k} U_{pki} \Bigg\|^2 + \mathbb{E}\Bigg\| \sum_{k<p<i} U_{pki} \Bigg\|^2 \\ &\qquad + \mathbb{E}\Bigg\| \sum_{i<p<k} U_{pki} \Bigg\|^2 + \mathbb{E}\Bigg\| \sum_{k<i<p} U_{pki} \Bigg\|^2 + \mathbb{E}\Bigg\| \sum_{i<k<p} U_{pki} \Bigg\|^2 \Bigg] . \end{aligned} \tag{B.85}$$

To bound the first expectation, we proceed as in the previous proofs and compute:

$$\begin{aligned} \mathbb{E}\Bigg\| \sum_{p<k<i} U_{pki} \Bigg\|^2 &= \mathbb{E}\left[\Bigg(\sum_{p<k<i} U_{pki} \Bigg)^\top \Bigg(\sum_{p<k<i} U_{pki} \Bigg) \right] \\ &= \mathbb{E}\left[\sum_{p<k<i} \sum_{q<l<j} U_{pki}^\top U_{qlj} \right] = \sum_{p<k<i} \sum_{q<l<j} \mathbb{E}\left[U_{pki}^\top U_{qlj}\right] . \end{aligned}$$

Whenever $i \neq j$, the expectation vanishes, since in this case either $i \notin \{p, k, q, l, j\}$ or $j \notin \{p, k, i, q, l\}$ follows (compare proof of Lemma B.4.4). Hence, we have:

$$\mathbb{E}\Bigg\| \sum_{p<k<i} U_{pki} \Bigg\|^2 = \sum_{p<k<i} \sum_{q<l<i} \mathbb{E}\left[U_{pki}^\top U_{qli}\right] .$$

[68] In Kopperschmidt and Stute 2013, the constant factor was miscalculated as 64 instead of 32.

By the same argument as above, only $k = l$ needs to be considered, so we can conclude:

$$\mathbb{E}\Big\| \sum_{p<k<i} U_{pki} \Big\|^2 = \sum_{p,q<k<i} \mathbb{E}\left[U_{pki}^\top U_{qki}\right] . \tag{B.86}$$

Performing the exact same steps for the remaining expectations of Equation (B.85), we receive:

$$\begin{aligned}
&\mathbb{E}\Big\| \sum_{p<i<k} U_{pki} \Big\|^2 = \sum_{p,q<i<k} \mathbb{E}\left[U_{pki}^\top U_{qki}\right] , \quad \mathbb{E}\Big\| \sum_{k<p<i} U_{pki} \Big\|^2 = \sum_{k<p,q<i} \mathbb{E}\left[U_{pki}^\top U_{qki}\right] , \\
&\mathbb{E}\Big\| \sum_{i<p<k} U_{pki} \Big\|^2 = \sum_{i<p,q<k} \mathbb{E}\left[U_{pki}^\top U_{qki}\right] , \quad \mathbb{E}\Big\| \sum_{k<i<p} U_{pki} \Big\|^2 = \sum_{k<i<p,q} \mathbb{E}\left[U_{pki}^\top U_{qki}\right] , \\
&\mathbb{E}\Big\| \sum_{i<k<p} U_{pki} \Big\|^2 = \sum_{i<k<p,q} \mathbb{E}\left[U_{pki}^\top U_{qki}\right] .
\end{aligned} \tag{B.87}$$

Substituting Equations (B.86) and (B.86) into Equation (B.85) then yields the desired result. □

In view of the moment condition in Kolmogorov's tightness criterion, see part (ii) of Theorem B.2.6, it is apparent that the Lemmas B.4.4 to B.4.6 are inherently connected to Kopperschmidt's proof approach. Nevertheless, we will also make use of basic L^2 techniques to infer several straightforward convergences. Since the corresponding lemma is irrelevant outside our adapted proof, we can resort to the stronger Assumptions 3.8 for its formulation.

Lemma B.4.7 (Limit Theorem for Joint Means of Multi-Indexed Random Vectors)**.** *Let $m \in \mathbb{N}$ and $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Suppose that $U_{i_1,\dots,i_m}$, $i_1, \dots, i_m \in \mathbb{N}$, are uniformly square-integrable real random vectors, that is, for some constant $C > 0$ it holds:*

$$\mathbb{E}\, \|U_{i_1,\dots,i_m}\|^2 \le C < \infty\,, \quad \textit{for all } i_1, \dots, i_m \in \mathbb{N}\,.$$

If

$$\mathbb{E}\left[U_{i_1,\dots,i_m}^\top U_{j_1,\dots,j_m}\right] = 0 \quad \textit{whenever all indices differ,} \tag{B.88}$$

then

$$\frac{1}{n^m} \sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} U_{i_1,\dots,i_m} \xrightarrow{\mathrm{L}^2} 0 \quad \textit{as } n \to \infty\,. \tag{B.89}$$

Proof. The proof is elementary. It relies on the fact that the number of summands with partially matching indices is of negligible order, whereas the summands with differing indices are zero according to Equation (B.88). We have:

$$\mathbb{E}\left\| \frac{1}{n^m} \sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} U_{i_1,\dots,i_m} \right\|^2 = \frac{1}{n^{2m}} \sum_{i_1,\dots,i_m} \sum_{j_1,\dots,j_m} \mathbb{E}\left(U_{i_1,\dots,i_m}^\top U_{j_1,\dots,j_m}\right)$$

Because of Equation (B.88), we need only consider the sub-sum over those index combinations where at least two indices match. In total, there are

$$\underbrace{n^{2m}}_{\substack{\text{total number}\\ \text{of combinations}}} \quad - \quad \underbrace{n(n-1)\cdot\ldots\cdot(n-2m+1)}_{\text{combinations with differing indices}} = \mathcal{O}\left(n^{2m-1}\right)$$

of these combinations. But for any such index combination, the Cauchy-Schwarz inequality yields:

$$\left|\mathbb{E}\left(U_{i_1,\ldots,i_m}^\top U_{j_1,\ldots,j_m}\right)\right| \leq \sqrt{\mathbb{E}\,\|U_{i_1,\ldots,i_m}\|^2\cdot\mathbb{E}\,\|U_{j_1,\ldots,j_m}\|^2} \leq \sqrt{C\cdot C} = C\,,$$

as the involved random vectors are uniformly square-integrable by assumption. Hence,

$$\frac{1}{n^{2m}}\sum_{i_1,\ldots,i_m}\sum_{j_1,\ldots,j_m}\mathbb{E}\left(U_{i_1,\ldots,i_m}^\top U_{j_1,\ldots,j_m}\right) \leq \frac{C}{n^{2m}}\mathcal{O}\left(n^{2m-1}\right) \longrightarrow 0 \quad \text{as } n\to\infty\,,$$

which proves the assertion. □

B.4.3. Other Minor Remarks

Remark B.4.8 (On the Usage of Non-Sub-Multiplicative Matrix Norms)**.**
In Chapter 3 we considered the max norm $\|\cdot\|_{\max}$ for matrices. This norm lacks the commonly demanded property of sub-multiplicativity, which is why the induced vector norm (i.e., the maximum absolute element of a vector) is not compatible with the matrix norm itself. A counterexample can easily be given:

$$2 = \left\|\begin{pmatrix}2\\2\end{pmatrix}\right\|_{\max} = \left\|\begin{pmatrix}1&1\\1&1\end{pmatrix}\begin{pmatrix}1\\1\end{pmatrix}\right\|_{\max} \not\leq \left\|\begin{pmatrix}1&1\\1&1\end{pmatrix}\right\|_{\max}\cdot\left\|\begin{pmatrix}1\\1\end{pmatrix}\right\|_{\max} = 1\cdot 1 = 1\,.$$

Here the problem could be solved by rescaling the matrix norm with the factor $d=2$, and in general the max norm on $\mathbb{R}^{k\times l}$ indeed only needs to be multiplied by $\sqrt{kl}$ to obtain a sub-multiplicative (and thus compatible) norm. Although this would hardly lead to significant complications, the rescaling factor can easily be forgotten and often inflates subsequent calculations. This raises the following question: If for sequences $(x_n)_{n\in\mathbb{N}}\subset\mathbb{R}^d$ and $(A_n)_{n\in\mathbb{N}}\subset\mathbb{R}^{d\times d}$ we have $x_n\to 0$ $(n\to\infty)$ and $\|A_n\|_{\max}\leq C$ for some $C>0$ and all $n\in\mathbb{N}$, does $\|A_nx_n\|\to 0$ $(n\to\infty)$ follow even though $\|A_nx_n\|\leq\|A_n\|_{\max}\cdot\|x_n\|$ cannot be ensured? Due to the equivalence of norms on finite dimensional spaces, the answer is yes. Suggestively, we denote the operator norm induced by the (arbitrary!) vector norm also with $\|\cdot\|$ and thus obtain for some constant $C_{\|\cdot\|}$:

$$\|A_nx_n\| \leq \|A_n\|\cdot\|x_n\| \leq C_{\|\cdot\|}\cdot\|A_n\|_{\max}\cdot\|x_n\| \leq C_{\|\cdot\|}\cdot C\cdot\|x_n\| \longrightarrow 0 \quad \text{as } n\to\infty\,. \tag{B.90}$$

Note that Equation (B.90) remains valid regardless of the matrix norm chosen. Consequently, it is always sufficient for convergence to prove that the sequence $(A_n)_{n\in\mathbb{N}}$ is bounded with respect to any - not necessarily sub-multiplicative - matrix norm.

Lemma B.4.9 (Law of Total Expectation)**.**
Let X be an integrable random variable defined on a probability space $(\Omega,\mathcal{F},\mathbb{P})$. For any event $A\in\mathcal{F}$ then holds

$$\mathbb{E}(X\,|\,\sigma(A)) = \mathbb{E}(X\,|\,A)\cdot\mathbb{1}_A + \mathbb{E}(X\,|\,A^\complement)\cdot\mathbb{1}_{A^\complement}\,,$$

where $\sigma(A) = \{\emptyset, A, A^{\complement}, \Omega\}$ *is the* σ*-algebra induced by* A*. In particular, if* $\mathbb{P}(A) > 0$*, the random variable* $\mathbb{E}(X \,|\, \sigma(A))$ *corresponds on* A *to the usual conditional expectation,*

$$\mathbb{E}(X \,|\, A) := \int X \,\mathrm{d}\mathbb{P}(\,\cdot\,|\, A) = \frac{1}{\mathbb{P}(A)} \int_A X \,\mathrm{d}\mathbb{P} = \frac{\mathbb{E}\left(X \cdot \mathbb{1}_A\right)}{\mathbb{P}(A)}\,.$$

The expectation $\mathbb{E}(X)$ *then satisfies the equation*

$$\mathbb{E}(X) = \mathbb{E}(X \,|\, A) \cdot \mathbb{P}(A) + \mathbb{E}(X \,|\, A^{\complement}) \cdot \mathbb{P}(A^{\complement})\,. \tag{B.91}$$

We refer to the identity (B.91) *as the* law of total expectation*.*

Proof. We first establish the stated representation of $\mathbb{E}(X \,|\, \sigma(A))$. For it to hold, the following must be proved:

$$\forall B \in \sigma(A) : \quad \int_B X \,\mathrm{d}\mathbb{P} = \int_B \mathbb{E}(X \,|\, A) \cdot \mathbb{1}_A + \mathbb{E}(X \,|\, A^{\complement}) \cdot \mathbb{1}_{A^{\complement}} \,\mathrm{d}\mathbb{P}\,. \tag{B.92}$$

For any such B, we compute:

$$\begin{aligned}
&\int_B \mathbb{E}(X \,|\, A) \cdot \mathbb{1}_A + \mathbb{E}(X \,|\, A^{\complement}) \cdot \mathbb{1}_{A^{\complement}} \,\mathrm{d}\mathbb{P} \\
&\quad = \mathbb{E}(X \,|\, A) \int_B \mathbb{1}_A \,\mathrm{d}\mathbb{P} + \mathbb{E}(X \,|\, A^{\complement}) \int_B \mathbb{1}_{A^{\complement}} \,\mathrm{d}\mathbb{P} \\
&\quad = \mathbb{E}(X \,|\, A) \cdot \mathbb{P}(A \cap B) + \mathbb{E}(X \,|\, A^{\complement}) \cdot \mathbb{P}(A \cap B) \\
&\quad = \int_A X \,\mathrm{d}\mathbb{P} \cdot \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} + \int_{A^{\complement}} X \,\mathrm{d}\mathbb{P} \cdot \frac{\mathbb{P}(A^{\complement} \cap B)}{\mathbb{P}(A^{\complement})}\,.
\end{aligned}$$

For a non-trivial event $A \in \mathcal{F}$, we have $|\sigma(A)| = 4$, so for B only four different events need to be considered. If $B = A$, we obtain:

$$\begin{aligned}
&\int_A X \,\mathrm{d}\mathbb{P} \cdot \frac{\mathbb{P}(A \cap A)}{\mathbb{P}(A)} + \int_{A^{\complement}} X \,\mathrm{d}\mathbb{P} \cdot \frac{\mathbb{P}(A^{\complement} \cap A)}{\mathbb{P}(A^{\complement})} \\
&\quad = \int_A X \,\mathrm{d}\mathbb{P} \cdot \underbrace{\frac{\mathbb{P}(A)}{\mathbb{P}(A)}}_{=1} + \int_{A^{\complement}} X \,\mathrm{d}\mathbb{P} \cdot \underbrace{\frac{\mathbb{P}(\emptyset)}{\mathbb{P}(A^{\complement})}}_{=0} = \int_A X \,\mathrm{d}\mathbb{P}\,,
\end{aligned}$$

and analogously the other cases yield the validity of Equation (B.92). From here, the tower property then provides:

$$\begin{aligned}
\mathbb{E}(X) &= \mathbb{E}\left(\mathbb{E}(X \,|\, \sigma(A))\right) = \mathbb{E}\Big(\mathbb{E}(X \,|\, A) \cdot \mathbb{1}_A + \mathbb{E}(X \,|\, A^{\complement}) \cdot \mathbb{1}_{A^{\complement}}\Big) \\
&= \mathbb{E}(X \,|\, A) \cdot \mathbb{P}(A) + \mathbb{E}(X \,|\, A^{\complement}) \cdot \mathbb{P}(A^{\complement})\,,
\end{aligned}$$

and thus proves the law of total expectation. $\square$

B.5. Calculations for the Additional Models under Consideration

The applicability of the statistical methods studied in this dissertation was only shown for the model ${}^{\times}\mathrm{D}$, the Basquin load sharing model with multiplicative damage accumulation. In most cases, however, they can also be applied to related models such as model ${}^{\times}\mathrm{E}$ (exponential damage accumulation) or model ${}^{\times}\mathrm{S}$ (shifted damage accumulation), and the proofs can be easily adapted. In order to facilitate future research, we recapitulate some of the major calculations from the main part of this thesis for both models. We refrain from further reasoning and focus on the comprehensibility of these calculations.

B.5.1. Basquin Load Sharing Model With Exponential Damage Accumulation

We first repeat the definition of the intensity function for model ${}^{\times}\mathrm{E}$.

Definition 2.10 (Basquin Load Sharing Model With Exponential Damage Accumulation)**.** *In the framework of Section 2.1 and under Assumptions 2.3, the Basquin load sharing model with* exponential *damage accumulation is given via the intensity process*

$$ {}^{\times\mathrm{E}}\lambda_\theta^{(j)}(t) := \theta_1 B_j(t)^{\theta_2} \exp\left(\theta_3 A_j(t)\right) \cdot \mathbb{1}_{\left\{N_{t-}^{(j)} < C_j\right\} \cap \left\{t \le \tau_j\right\}}, \qquad \theta = (\theta_1, \theta_2, \theta_3)^\top \in \mathbb{R}_+^3 . $$

We structure the computations presented here in three parts:
In part (i), we give the cumulative conditional hazard function and its inverse for the model ${}^{\times}\mathrm{E}$. In part (ii), we derive the probability for a positive sign of the standardized hazard transform like in Equation (2.61), which then yields the link function required in condition (D5) for the consistency of the 3-sign depth test. Furthermore, analogous to Lemma 4.24, we prove the monotonicity of this link function with respect to the "essential past" of the process, which here corresponds to the damage accumulation term $A_{j,i}$. The part (iii) contains the likelihood function for the model ${}^{\times}\mathrm{E}$.

(i) **(Inverse) cumulative conditional hazard function.** A key feature of the model ${}^{\times}\mathrm{D}$ was that both the cumulative conditional hazard function and its inverse could be calculated explicitly. They are given in Lemmas 2.24 and 2.27, respectively. As a minimum requirement, we demand to preserve this property when we change the model, and in fact the calculations performed for the intensity function ${}^{\times\mathrm{D}}\lambda_\theta^{(j)}$ can be adopted almost one-to-one. On $\left\{t \ge T_{i-1}^{(j)}\right\}$, we have for $i \in \{1, \dots, I_c\}$:

$$ {}^{\times\mathrm{E}}h_i^\theta\left(t \,\middle|\, T_{1:(i-1)}^{(j)}, s_j\right) = \theta_1 B_{j,i}^{\theta_2} \exp\left(\theta_3 A_j(t)\right) . $$

Moreover, Equation (2.56) shows that

$$ \frac{\mathrm{d}}{\mathrm{d}t} \theta_3 A_j(t) = \frac{\theta_3 B_{j,i}}{\tau} . $$

We can hence compute:

$$ \begin{aligned} {}^{\times\mathrm{E}}H_i^\theta\left(t \,\middle|\, T_{1:(i-1)}^{(j)}, s_j\right) &= \int_{T_{i-1}^{(j)}}^t {}^{\times\mathrm{E}}h_i^\theta\left(u \,\middle|\, T_{1:(i-1)}^{(j)}, s_j\right) \mathrm{d}u \\ &= \int_{T_{i-1}^{(j)}}^t \theta_1 B_{j,i}^{\theta_2} \exp\left(\theta_3 A_j(u)\right) \mathrm{d}u \end{aligned} $$

$$
\begin{aligned}
&= \theta_1 B_{j,i}^{\theta_2} \left[\frac{\tau}{\theta_3 B_{j,i}} \exp\left(\theta_3 A_j(u)\right) \right]_{u=T_{i-1}^{(j)}}^{t} \\
&= \frac{\tau \theta_1 B_{j,i}^{\theta_2 - 1}}{\theta_3} \left[\exp\left(\theta_3 A_j(t)\right) - \exp\left(\theta_3 \frac{A_{j,i}}{\tau} \right) \right] \\
&= \frac{\tau \theta_1 B_{j,i}^{\theta_2 - 1}}{\theta_3} \exp\left(\frac{\theta_3}{\tau} A_{j,i} \right) \left[\exp\left(\frac{\theta_3}{\tau} B_{j,i} \left(t - T_{i-1}^{(j)} \right) \right) - 1 \right] .
\end{aligned}
$$

Again, this function is invertible with respect to t on the interval $[T_{i-1}^{(j)}, \infty)$. The inverse function is given as follows:

$$
\left({}^{\times\mathrm{E}}H_i^{\theta} \right)^{-1} \left(u \,\middle|\, T_{1:(i-1)}^{(j)}, s_j \right) = T_{i-1}^{(j)} + \frac{\tau}{\theta_3 B_{j,i}} \ln \left(1 + \frac{\theta_3}{\tau \theta_1 B_{j,i}^{\theta_2 - 1}} \exp\left(-\frac{\theta_3}{\tau} A_{j,i} \right) u \right) .
$$

(ii) **Signs of the standardized hazard transforms.** In accordance with Equation (2.57), we obtain from (i):

$$
\begin{aligned}
&\mathbb{P}_{\theta^*} \left(R_{j,i}^{\theta} > u \,\middle|\, T_{1:(i-1)}^{(j)}, s_j \right) \\
&\quad = \exp \left(-\tau \frac{\theta_1^*}{\theta_3^*} B_{j,i}^{\theta_2^* - 1} \exp\left(\frac{\theta_3^*}{\tau} A_{j,i} \right) \left[\left(1 + \frac{\theta_3}{\tau \theta_1 B_{j,i}^{\theta_2 - 1}} \exp\left(-\frac{\theta_3}{\tau} A_{j,i} \right) u \right)^{\frac{\theta_3^*}{\theta_3}} - 1 \right] \right) .
\end{aligned}
$$

At $u = \ln(2)$, this is the conditional probability of a positive sign for the standardized hazard transform $\tilde{R}_{j,i}^{\theta}$. If we continue in the framework of Chapter 4 and define

$$
\kappa\left(l, i, \theta\right) = \frac{\theta_3}{\tau \theta_1 B_{l,i}^{\theta_2 - 1}} > 0 \, ,
$$

then here the link function $g_{l,i}$ can be written as

$$
\begin{aligned}
&g_{l,i}\left(\theta, \theta^*, x\right) \\
&\quad = \exp \left(-\frac{1}{\kappa\left(l, i, \theta^*\right)} \exp\left(\frac{\theta_3^*}{\tau} x \right) \left[\left(1 + \ln(2) \kappa\left(l, i, \theta\right) \exp\left(-\frac{\theta_3}{\tau} x \right) \right)^{\frac{\theta_3^*}{\theta_3}} - 1 \right] \right) .
\end{aligned}
$$

Retracing the proof of Lemma 4.24 tells us that it is sufficient to check the auxiliary function

$$
\gamma(x) := \exp\left(\frac{\theta_3^*}{\tau} x \right) \left[\left(1 + \ln(2) \kappa\left(l, i, \theta\right) \exp\left(-\frac{\theta_3}{\tau} x \right) \right)^{\frac{\theta_3^*}{\theta_3}} - 1 \right]
$$

to infer the monotonicity of $g_{l,i}\left(\theta, \theta^*, \cdot\right)$. Whether $g_{l,i}\left(\theta, \theta^*, \cdot\right)$ is non-decreasing or non-increasing once more depends on how θ and θ^* differ. According to the product rule, the derivative of γ with respect to x is given by

$$
\frac{\mathrm{d}}{\mathrm{d}x} \gamma(x) = \frac{\theta_3^*}{\tau} \exp\left(\frac{\theta_3^*}{\tau} x \right) \left[\left(1 + \ln(2) \kappa\left(l, i, \theta\right) \exp\left(-\frac{\theta_3}{\tau} x \right) \right)^{\frac{\theta_3^*}{\theta_3}} - 1 \right]
$$

$$
\begin{aligned}
&- \ln(2)\kappa\left(l,i,\theta\right)\frac{\theta_3^*}{\theta_3}\frac{\theta_3}{\tau}\exp\left(-\frac{\theta_3}{\tau}x\right)\exp\left(\frac{\theta_3^*}{\tau}x\right)\\
&\cdot\left(1+\ln(2)\kappa\left(l,i,\theta\right)\exp\left(-\frac{\theta_3}{\tau}x\right)\right)^{\frac{\theta_3^*-\theta_3}{\theta_3}},
\end{aligned}
$$

and introducing another auxiliary function via

$$
z(x) := 1 + \underbrace{\ln(2)\kappa\left(l,i,\theta\right)\exp\left(-\frac{\theta_3}{\tau}x\right)}_{>0} > 1
$$

allows the simplified representation

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}x}\gamma(x) &= \frac{\theta_3^*}{\tau}\exp\left(\frac{\theta_3^*}{\tau}x\right)\left[z(x)^{\frac{\theta_3^*}{\theta_3}} - 1 - (z(x)-1)\cdot z(x)^{\frac{\theta_3^*-\theta_3}{\theta_3}}\right]\\
&= \frac{\theta_3^*}{\tau}\exp\left(\frac{\theta_3^*}{\tau}x\right)\left[z(x)^{\frac{\theta_3^*-\theta_3}{\theta_3}} - 1\right].
\end{aligned}
$$

As $\frac{\theta_3^*}{\tau}\exp\left(\frac{\theta_3^*}{\tau}x\right) > 0$, we can further break down this equation in terms of its sign:

$$
\begin{aligned}
0 \gtreqless \frac{\mathrm{d}}{\mathrm{d}x}\gamma(x) \quad &\Longleftrightarrow \quad 0 \gtreqless z(x)^{\frac{\theta_3^*-\theta_3}{\theta_3}} - 1\\
&\Longleftrightarrow \quad 1 \gtreqless z(x)^{\frac{\theta_3^*-\theta_3}{\theta_3}}.
\end{aligned}
$$

Because of $z(x) > 1$ for all $x \in \mathbb{R}$ we furthermore receive:

$$
z(x)^{\frac{\theta_3^*-\theta_3}{\theta_3}}\begin{cases} >1\,, & \text{for } \theta_3^* > \theta_3,\\ =1\,, & \text{for } \theta_3^* = \theta_3,\\ <1\,, & \text{for } \theta_3^* < \theta_3.\end{cases}
$$

With the same arguments as before, $g_{l,i}\left(\theta,\theta^*,x\right)$ is thus strictly decreasing in x for $\theta_3^* > \theta_3$, strictly increasing for $\theta_3^* < \theta_3$, and constant in the case $\theta_3^* = \theta_3$.

(iii) **(Log-)likelihood function.** The likelihood function can be calculated by substituting the intensity process ${}^{\times\mathrm{E}}\lambda_\theta^{(j)}$ into Equation (5.3):

$$
\begin{aligned}
L_{\times\mathrm{E}}(\theta) &= \prod_{j=1}^{J}\left[\prod_{i=1}^{N_\tau^{(j)}} {}^{\times\mathrm{E}}\lambda_\theta^{(j)}\left(t_i^{(j)}\right)\exp\left(-\int_0^\tau {}^{\times\mathrm{E}}\lambda_\theta^{(j)}(u)\,\mathrm{d}u\right)\right]\\
&= \prod_{j=1}^{J}\left[\prod_{i=1}^{N_\tau^{(j)}} \theta_1 B_{j,i}^{\theta_2}\exp\left(\theta_3\frac{A_{j,i+1}}{\tau}\right)\exp\left(-\sum_{k=1}^{\tilde{C}_j}\frac{\tau\theta_1 B_{j,k}^{\theta_2-1}}{\theta_3}\right.\right.\\
&\qquad\left.\left.\cdot\left(\exp\left(\frac{\theta_3}{\tau}\tilde{A}_{j,k+1}\right) - \exp\left(\frac{\theta_3}{\tau}\tilde{A}_{j,k}\right)\right)\right)\right],
\end{aligned}
$$

where $\tilde{C}_j$ and $\tilde{A}_{j,k}$ are defined as in Theorem 5.4. The corresponding log-likelihood is given by:

$$l_{\times\mathrm{E}}(\theta) = \sum_{j=1}^{J} \left[\sum_{i=1}^{N_\tau^{(j)}} \left(\log(\theta_1) + \theta_2 \log\left(B_{j,i}\right) + \theta_3 \frac{A_{j,i+1}}{\tau} \right) \right.$$
$$\left. - \frac{\tau\theta_1}{\theta_3} \sum_{k=1}^{\tilde{C}_j} B_{j,k}^{\theta_2 - 1} \left(\exp\left(\frac{\theta_3}{\tau} \tilde{A}_{j,k+1} \right) - \exp\left(\frac{\theta_3}{\tau} \tilde{A}_{j,k} \right) \right) \right] ,$$

which again (compare Equation (5.19)) admits an expression of the form

$$l_{\times\mathrm{E}}\left(\left(\widehat{\theta}_1(\theta_2, \theta_3), \theta_2, \theta_3 \right)^\top \right) = \left[\log\left(\frac{\sum_{j=1}^J N_\tau^{(j)}}{G_2\left(\theta_2, \theta_3\right)} \right) - 1 \right] \sum_{j=1}^{J} N_\tau^{(j)} + G_1\left(\theta_2, \theta_3\right) .$$

B.5.2. Basquin Load Sharing Model With Shifted Damage Accumulation

We start again by repeating the definition of the intensity function for model ${}^\times S$.

Definition 2.9 (Basquin Load Sharing Model With Shifted Damage Accumulation)**.** *In the framework of Section 2.1 and under Assumptions 2.3, the Basquin load sharing model with* shifted *damage accumulation is given via the intensity process*

$${}^{\times\mathrm{S}}\lambda_\theta^{(j)}(t) := \theta_1 B_j(t)^{\theta_2} \left(1 + A_j(t)\right)^{\theta_3} \cdot \mathbb{1}_{\left\{N_{t^-}^{(j)} < C_j\right\} \cap \left\{t \le \tau_j\right\}} , \qquad \theta = (\theta_1, \theta_2, \theta_3)^\top \in \mathbb{R}_+^3 .$$

We proceed as for the model with exponential damage accumulation by retracing and adjusting the essential formulas within three parts.

(i) **(Inverse) cumulative conditional hazard function.** Applying any affine transformation to the damage accumulation term has virtually no effect on the calculation of the intensity integral. More precisely, adopting Equation (2.54) yields:

$$\begin{aligned}
{}^{\times\mathrm{S}}H_i^\theta\left(t \,\middle|\, T_{1:(i-1)}^{(j)}, s_j\right) &= \int_{T_{i-1}^{(j)}}^{t} \theta_1 B_{j,i}^{\theta_2} \left[1 + \frac{1}{\tau} \left(B_{j,i}\left(u - T_{i-1}^{(j)}\right) + A_{j,i} \right) \right]^{\theta_3} \mathrm{d}u \\
&= \theta_1 B_{j,i}^{\theta_2} \left(\frac{1}{\tau} \right)^{\theta_3} \int_{T_{i-1}^{(j)}}^{t} \left[\tau + B_{j,i}\left(u - T_{i-1}^{(j)}\right) + A_{j,i} \right]^{\theta_3} \mathrm{d}u \\
&= \frac{\theta_1 B_{j,i}^{\theta_2 - 1}}{\tau^{\theta_3}(\theta_3 + 1)} \left[\left(\tau + B_{j,i}\left(t - T_{i-1}^{(j)}\right) + A_{j,i} \right)^{\theta_3 + 1} - \left(\tau + A_{j,i}\right)^{\theta_3 + 1} \right] ,
\end{aligned}$$

with corresponding inverse function

$$\begin{aligned}
&\left({}^{\times\mathrm{S}}H_i^\theta \right)^{-1} \left(u \,\middle|\, T_{1:(i-1)}^{(j)} \right) \\
&\qquad = \frac{1}{B_{j,i}} \left[\left(\frac{\tau^{\theta_3}\left(\theta_3 + 1\right)}{\theta_1 B_{j,i}^{\theta_2 - 1}} u + \left(\tau + A_{j,i}\right)^{\theta_3 + 1} \right)^{\frac{1}{\theta_3 + 1}} - \left(\tau + A_{j,i}\right) + B_{j,i} T_{i-1}^{(j)} \right] .
\end{aligned}$$

(ii) **Signs of the standardized hazard transforms.** From step (i) we immediately get that

$$\begin{aligned}
&\mathbb{P}_{\theta^*}\left(R^{\theta}_{j,i} > u \,\middle|\, T^{(j)}_{1:(i-1)}, s_j\right) \\
&\quad = \exp\left(-\frac{\theta_1^* B_{j,i}^{\theta_2^*-1}}{\tau^{\theta_3^*}\left(\theta_3^*+1\right)}\left[\left(\frac{\tau^{\theta_3}\left(\theta_3+1\right)}{\theta_1 B_{j,i}^{\theta_2-1}}u + \left(\tau + A_{j,i}\right)^{\theta_3+1}\right)^{\frac{\theta_3^*+1}{\theta_3+1}} - \left(\tau + A_{j,i}\right)^{\theta_3^*+1}\right]\right),
\end{aligned}$$

which is nothing else than Equation (2.57) if we replace each instance of $\tau + A_{j,i}$ with $A_{j,i}$. Accordingly, if we consider a link function ${}^{\times\mathrm{S}}g_{l,i}$ for condition (D5) of Assumption 4.14, we have

$${}^{\times\mathrm{S}}g_{l,i}(\theta, \theta^*, x) = {}^{\times\mathrm{D}}g_{l,i}(\theta, \theta^*, \tau + x)\,,$$

where ${}^{\times\mathrm{D}}g_{l,i}$ is the corresponding link function for the model ${}^{\times}$D. Since the shift $x \mapsto \tau + x$ preserves the monotonicity properties of this link function, the statement of Lemma 4.24 remains valid.

(iii) **(Log-)likelihood function.** We already recognized in steps (i) and (ii) that shifting the damage accumulation term by 1 involves no significant changes in the resulting formulas. The same applies to the likelihood function, which is directly obtained from Theorem 5.4 by once again substituting $\tau + A_{j,i}$ for $A_{j,i}$ (and likewise for $\tilde{A}_{j,i}$):

$$\begin{aligned}
L_{\times\mathrm{S}}(\theta) = \prod_{j=1}^{J}\Bigg[\prod_{i=1}^{N_\tau^{(j)}} \theta_1 B_{j,i}^{\theta_2}\left(1 + \frac{A_{j,i+1}}{\tau}\right)^{\theta_3} \exp\Bigg(-\sum_{k=1}^{\tilde{C}_j} \frac{\tau\theta_1 B_{j,k}^{\theta_2-1}}{\theta_3} \\
\cdot \left(\left(\tau + \tilde{A}_{j,k+1}\right)^{\theta_3+1} - \left(\tau + \tilde{A}_{j,k}\right)^{\theta_3+1}\right)\Bigg)\Bigg]\,.
\end{aligned}$$

Again, the log-likelihood function can be represented in a way similar to Equation (5.16), but we omit the explicit specification here due to its repetitive nature.

C. Additional Tables and Figures

Table 7: List of recurring symbols with a fixed meaning within the thesis.

Symbol	Meaning
$\tau > 0$	deterministic constant, marks the end of an experiment
$\mathcal{I} \subset \mathbb{R}$	interval, typically $\mathcal{I} = [0, \tau]$ or $\mathcal{I} = [0, \infty)$
$N = (N_t)_{t \in \mathcal{I}}$	counting process over the interval $\mathcal{I}$
$(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in \mathcal{I}}, \mathbb{P})$	probability space consisting of sample space Ω, σ-algebra $\mathcal{F}$, filtration $\{\mathcal{F}_t\}_{t \in \mathcal{I}}$ and probability measure $\mathbb{P}$
$\mathbb{E}(\cdot)$	expected value
$\mathcal{B}(\cdot)$	Borel σ-algebra
$\mathcal{F}_t^N$	σ-algebra generated by the history of N up to time t
$\mathcal{G}_0$	σ-algebra, contains information about random external covariates
$\mathcal{F}_t^N \vee \mathcal{G}_0$	intrinsic filtration
$J \in \mathbb{N}$	number of processes or repetitions of an experiment
$N^{(1)}, \ldots, N^{(J)}$	independent copies of a counting process N
X or X_j	real-valued random variable of arbitrary dimension
$T^{(j)} = \big(T_i^{(j)}\big)_{i \in \mathbb{N}}$	simple point process associated with $N^{(j)}$, similar for T and N
$t_i^{(j)} = T_i^{(j)}(\omega)$	realization of $T_i^{(j)}$ at $\omega \in \Omega$
$\Lambda^{(j)} = \big(\Lambda^{(j)}(t)\big)_{t \in \mathcal{I}}$	compensator of $N^{(j)}$ given by the Doob-Meyer decomposition
M or $M^{(j)}$	martingale, typically satisfying $M^{(j)} = N^{(j)} - \Lambda^{(j)}$
$\lambda^{(j)} = \big(\lambda^{(j)}(t)\big)_{t \in \mathcal{I}}$	stochastic intensity or intensity process corresponding to $\Lambda^{(j)}$
$\Theta \subset \mathbb{R}^d$, $d \in \mathbb{N}$	parameter space and its dimension d
$\theta, \theta^* \in \Theta$	parameter of interest, θ^* denotes the true parameter
$\mathcal{M}$	parametric intensity-based model
$\lambda_\theta^{(j)}$ or $\Lambda_\theta^{(j)}$	parametric (cumulative) intensity
$1:n$	abbreviated notation for $(1, \ldots, n)$
f_i, S_i, h_i	(conditional) density, survival, hazard function of T_i, may depend on θ
λ^*	conditional intensity function, usually identified with λ
$W_i^{(j)} = T_i^{(j)} - T_{i-1}^{(j)}$	interarrival or waiting time
$\mathbb{1}_A : \Omega \to \{0, 1\}$	indicator function of a subset $A \subset \Omega$
$I_j \in \mathbb{N}$	covariate, number of components of the jth system; often $I_j \equiv I$
$s_j > 0$	(random) covariate, initial stress level in the jth system
$\tau_j \in [0, \tau]$	(random) covariate, the end of the jth experiment
$C_j \in \{1, \ldots, I_j\}$	(random) covariate, number of observable component failures for the jth system
δ_x	Dirac measure centred on x
$\mathcal{E}(\beta)$	exponential distribution with parameter β

Continues on the following page.

Table 7: List of recurring symbols with a fixed meaning within the thesis (continued).

Symbol	Meaning
$\mathcal{N}_d\left(\mu, \Sigma\right)$	d-dimensional normal distribution with mean vector μ and covariance matrix Σ
${}^{\mathrm{B}}\lambda_\theta^{(j)}(t)$	intensity process of the Basquin load sharing model without damage accumulation (model identifier B)
$B_j(t)$, $B_{j,i}$	load sharing term and its abbreviated notation
$A_j(t)$, $A_{j,i}$	damage accumulation term and its abbreviated notation
${}^{\times\mathrm{D}}\lambda_\theta^{(j)}(t)$	intensity process of the Basquin load sharing model with multiplicative damage accumulation (model identifier ${}^{\times}$D)
${}^{+}$D, ${}^{\times}$S, ${}^{\times}$E, ${}^{\times}\phi$	further model identifiers
s_{low}, s_{upp}	lower and upper bound for the initial stress level s_j
$\mathrm{supp}(\cdot)$	support of a measure or function
$I_c \in \{1, \ldots, I\}$	critical number of component failures; used with $C_j \equiv I_c$
π_i	ith coordinate projection (e.g., $\pi_2\big((\theta_1, \theta_2, \theta_3)^\top\big) = \theta_2$)
H_i	conditional cumulative hazard function of T_i, may depend on θ
$R_{j,i}^\theta$	hazard transform of $T_i^{(j)}$ at θ
$\tilde{R}_{j,i}^\theta$	standardized hazard transform of $T_i^{(j)}$ at θ
$g_{j,i}, g_{l,i}, \ldots$	link functions, typically in conjunction with hazard transforms
$\mathcal{H}_0$, $\mathcal{H}_1$	null and alternative hypothesis of a statistical hypothesis test
$\Theta_0, \Theta_1 \subset \Theta$	subsets of Θ defining $\mathcal{H}_0$ and $\mathcal{H}_1$
$\mathrm{L}^2\left(\Omega, \mathcal{F}, \mu\right)$	space of square-integrable random variables on $(\Omega, \mathcal{F}, \mu)$
$\langle \cdot, \cdot \rangle_\mu$	inner product on $\mathrm{L}^2\left(\Omega, \mathcal{F}, \mu\right)$
$\lVert \cdot \rVert_\mu$	norm induced by $\langle \cdot, \cdot \rangle_\mu$
$\overline{N}^{(n)}$, $\overline{\Lambda}_\theta^{(n)}$, $\overline{M}^{(n)}$	aggregate counting process, model compensator at θ, martingale
$\mathcal{J}_N$, $\mathcal{J}_{\overline{N}^{(n)}}$	set of time points belonging to the jumps of N or $\overline{N}^{(n)}$, respectively
$\hat{\theta}_n$	minimum distance estimator for θ^*
$C^0(\mathcal{X}) = C(\mathcal{X})$	space of continuous functions on $\mathcal{X}$
$C^k(\mathcal{X})$	space of k-times continuously differentiable functions on $\mathcal{X}$
$\overline{\Theta}$	closure of the parameter space Θ; similar for other sets
K	either a compact set or an integer (for K-sign depth)
$\mathrm{B}_r\left(\theta\right)$	open ball with radius r around θ (in $\overline{\Theta}$)
$\frac{\mathrm{d}}{\mathrm{d}\theta}$, D_θ	total derivative with respect to θ
$\frac{\partial}{\partial \theta_j}$	partial derivative with respect to θ_j
$\frac{\mathrm{d}^p}{\mathrm{d}\theta^p}$, D_θ^p	pth total derivative with respect to θ
α_n, β_n, γ_n	auxiliary parametric processes (vector-valued)
Φ_n, Φ_0, Ψ_n	auxiliary parametric processes (matrix-valued); Ψ_K is also used for a functional on $\mathcal{D}\left([0,1]\right)$
$\mathcal{D}\left(\mathcal{I}\right)$	Skorokhod space of càdlàg functions on $\mathcal{I}$

Continues on the following page.

Table 7: List of recurring symbols with a fixed meaning within the thesis (continued).

Symbol	Meaning
$\Sigma(\theta^*)$	asymptotic covariance matrix at θ^*
$\|\cdot\|_{\max}$	max norm (for matrices)
$\xrightarrow{\mathrm{d}}$	convergence in distribution
$\xrightarrow{\mathbb{P}}$	convergence in probability
$\xrightarrow{\mathrm{L}^2}$	convergence in quadratic mean
$\mathcal{O}(\cdot)$	Big O notation for the order of a function
$o(1)$, $o_{\mathbb{P}}(1)$	Bachmann-Landau notation: converges to 0 (in probability)
$\Psi_K(W)$	asymptotic distribution of the K-sign depth
$\Psi_K(\mathcal{W}^N(\theta))$	normalized K-sign depth at θ
$q_\alpha(\cdot)$	α-quantile of a given distribution
$\alpha \in (0,1)$	the level of a test or confidence region
$\eta \in \mathbb{N}$	total number of observations
$L \in \mathbb{N}$	number of different distributions or classes
$\leq_{\mathrm{acc}}$	total order for double-indexed point process transforms
$L(\theta)$	likelihood function at θ; often used with model indicator
$l(\theta)$	log-likelihood function at θ; often used with model indicator
$\tilde{C}_j$, $\tilde{T}_i^{(j)}$, $\tilde{A}_{j,i}$	special notation used only for the (log-)likelihood function
$\mathcal{C}_{J,1-\alpha}$	$(1-\alpha)$-confidence region for the true parameter θ^*
$\chi^2_{d,1-\alpha}$	$(1-\alpha)$-quantile of the χ^2-distribution with d degrees of freedom

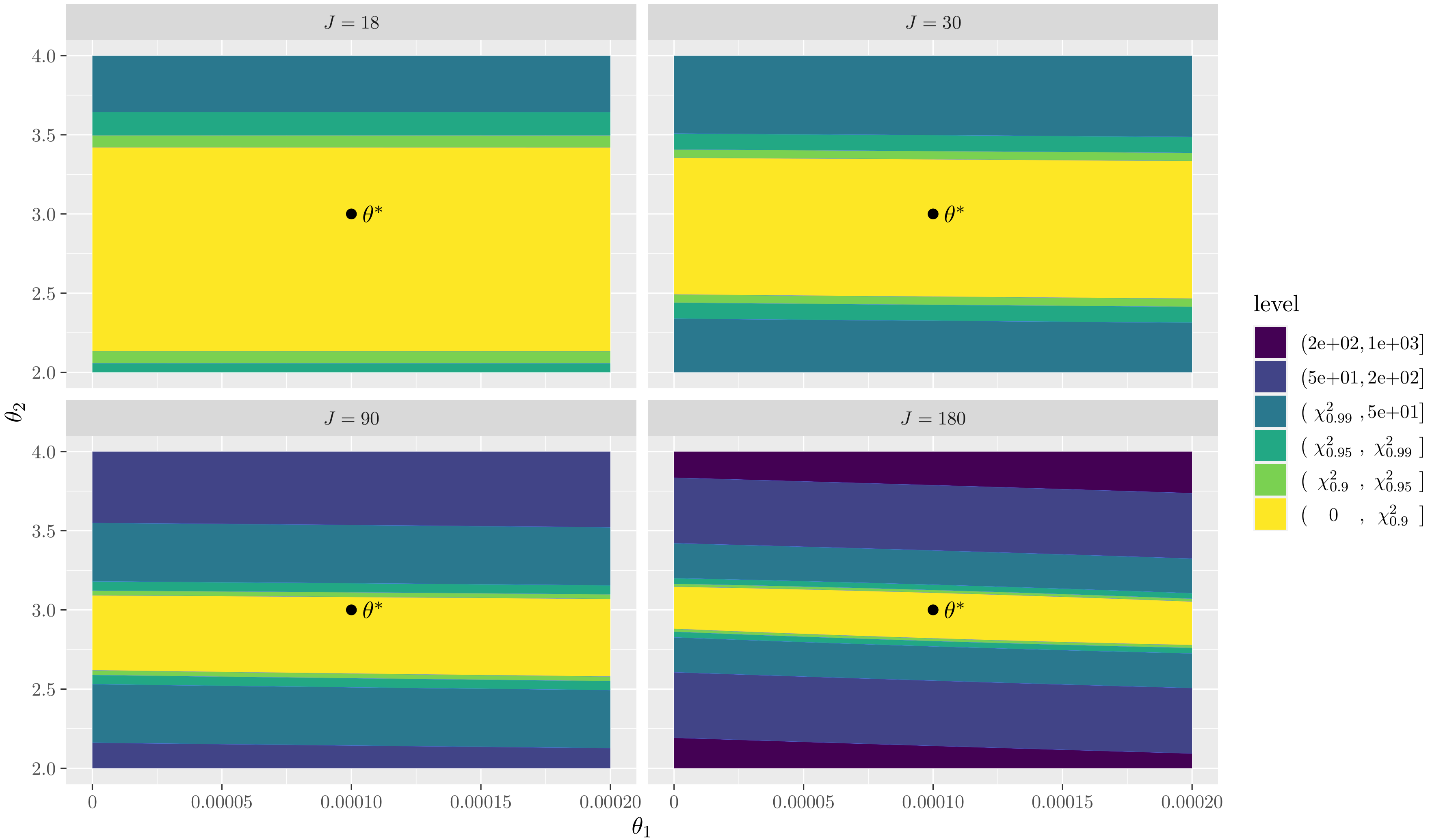


Figure 22: **Minimum Distance Estimator, θ_1-θ_2-direction.** Visualization of $\mathcal{C}^{(\mathrm{dist})}_{J,1-\alpha} \cap \{\theta_3 = \theta_3^*\}$, that is, the intersection of the $(1-\alpha)$-confidence region based on the minimum distance estimator with the hyperplane $\{\theta_3 = \theta_3^*\}$. The plot shows the contour lines of $J(\hat{\theta}_J - \theta)^\top \hat{\Sigma}_\Phi(\hat{\theta}_J)^{-1}(\hat{\theta}_J - \theta)$ as a function of θ for $J \in \{18, 30, 90, 180\}$ and the parameters & covariates given in Table 2. For the estimation of $\hat{\Sigma}_\Phi(\hat{\theta}_J)$, we chose $J_{\mathrm{sim}} = J$ and $n_{\mathrm{rep}} = 3000$ in each case. The intersections of the confidence regions for $\alpha \in \{0.01, 0.05, 0.1\}$ with the hyperplane can be identified from the levels; for example, the innermost yellow area corresponds to $\mathcal{C}^{(\mathrm{dist})}_{J,0.9} \cap \{\theta_3 = \theta_3^*\}$.

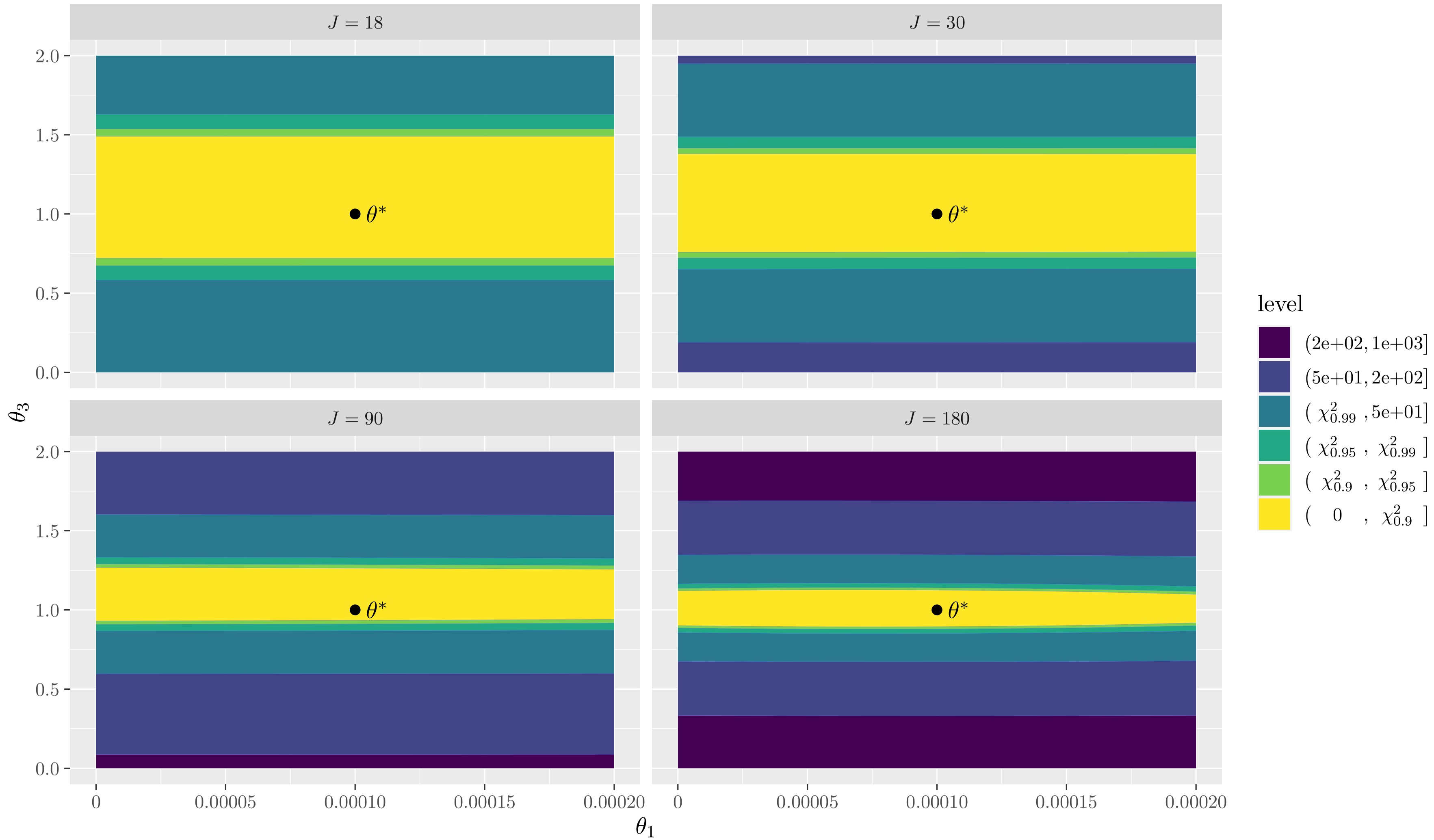


Figure 23: **Minimum Distance Estimator, θ_1-θ_3-direction.** Visualization of $\mathcal{C}^{(\mathrm{dist})}_{J,1-\alpha} \cap \{\theta_2 = \theta_2^*\}$, that is, the intersection of the $(1-\alpha)$-confidence region based on the minimum distance estimator with the hyperplane $\{\theta_2 = \theta_2^*\}$. The plot shows the contour lines of $J(\hat{\theta}_J - \theta)^\top \hat{\Sigma}_\Phi(\hat{\theta}_J)^{-1}(\hat{\theta}_J - \theta)$ as a function of θ for $J \in \{18, 30, 90, 180\}$ and the parameters & covariates given in Table 2. For the estimation of $\hat{\Sigma}_\Phi(\hat{\theta}_J)$, we chose $J_{\mathrm{sim}} = J$ and $n_{\mathrm{rep}} = 3000$ in each case. The intersections of the confidence regions for $\alpha \in \{0.01, 0.05, 0.1\}$ with the hyperplane can be identified from the levels; for example, the innermost yellow area corresponds to $\mathcal{C}^{(\mathrm{dist})}_{J,0.9} \cap \{\theta_2 = \theta_2^*\}$.

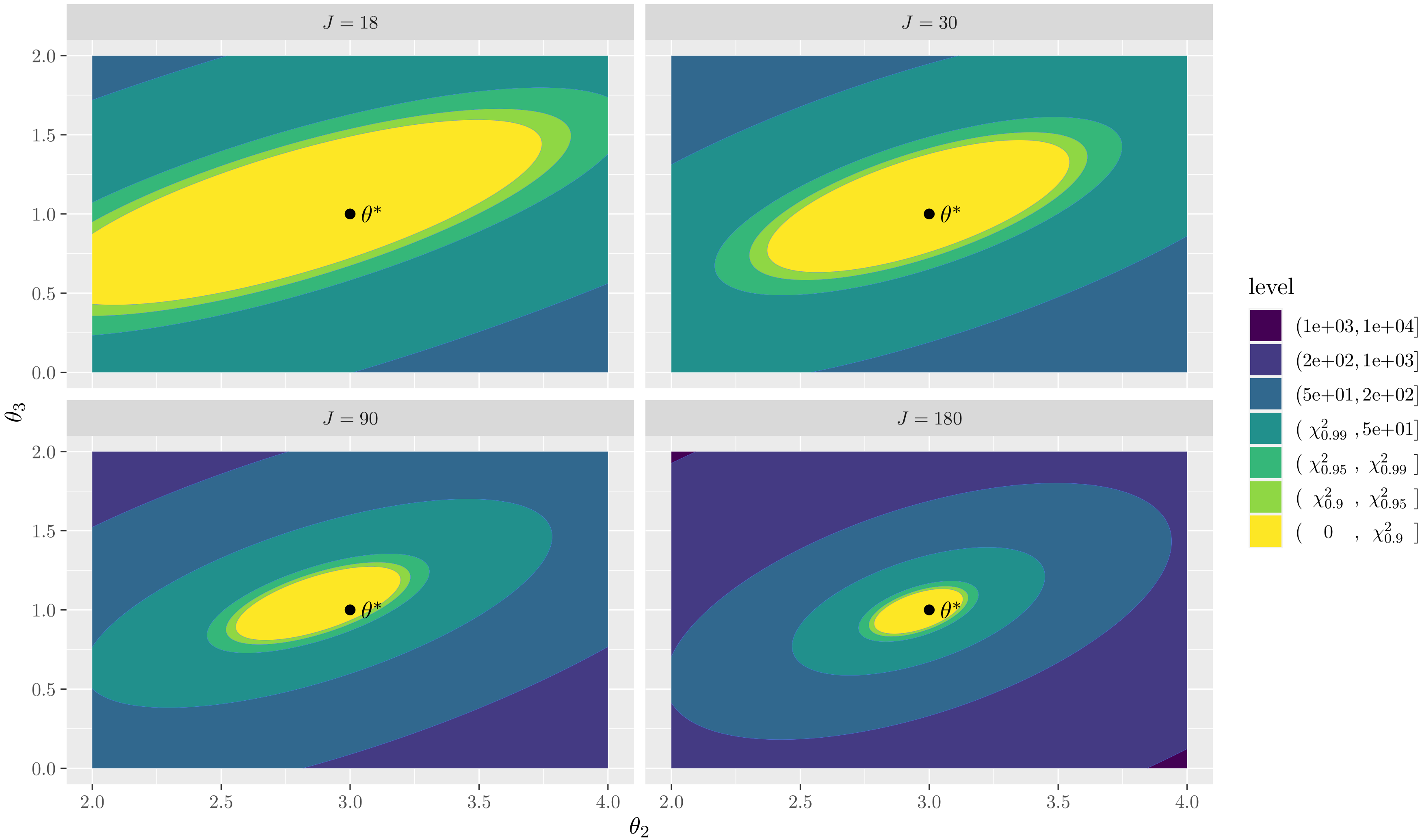


Figure 24: **Minimum Distance Estimator, θ_2-θ_3-direction.** Visualization of $\mathcal{C}^{(\text{dist})}_{J,1-\alpha} \cap \{\theta_1 = \theta_1^*\}$, that is, the intersection of the $(1-\alpha)$-confidence region based on the minimum distance estimator with the hyperplane $\{\theta_1 = \theta_1^*\}$. The plot shows the contour lines of $J(\hat{\theta}_J - \theta)^\top \hat{\Sigma}_\Phi(\hat{\theta}_J)^{-1}(\hat{\theta}_J - \theta)$ as a function of θ for $J \in \{18, 30, 90, 180\}$ and the parameters & covariates given in Table 2. For the estimation of $\hat{\Sigma}_\Phi(\hat{\theta}_J)$, we chose $J_{\text{sim}} = J$ and $n_{\text{rep}} = 3000$ in each case. The intersections of the confidence regions for $\alpha \in \{0.01, 0.05, 0.1\}$ with the hyperplane can be identified from the levels; for example, the innermost yellow area corresponds to $\mathcal{C}^{(\text{dist})}_{J,0.9} \cap \{\theta_1 = \theta_1^*\}$.

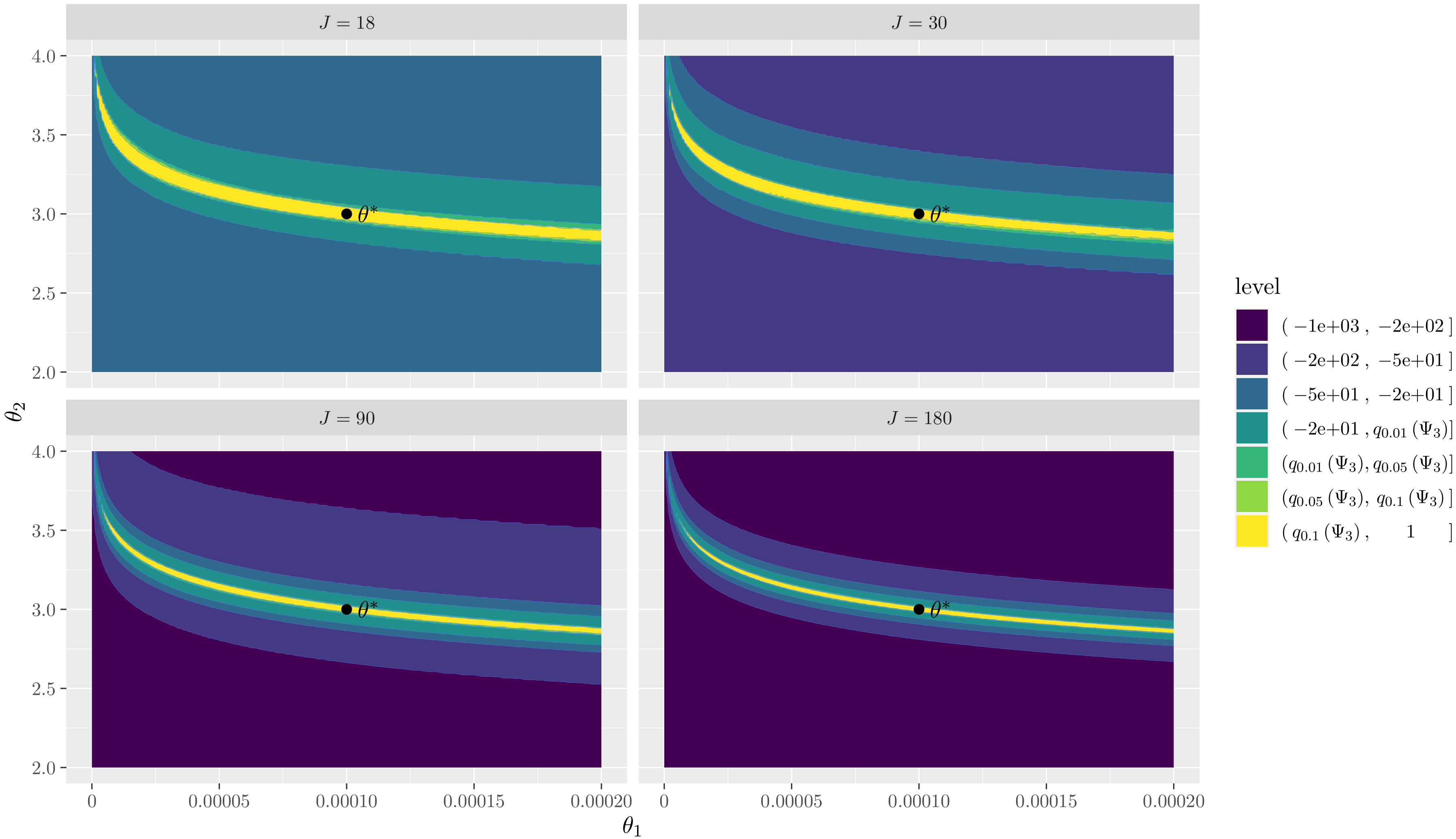


Figure 25: **3-Sign Depth, θ_1-θ_2-direction.** Visualization of $\mathcal{C}^{(\text{depth})}_{J,1-\alpha} \cap \{\theta_3 = \theta_3^*\}$, that is, the intersection of the $(1-\alpha)$-confidence region based on the 3-sign depth test with the hyperplane $\{\theta_3 = \theta_3^*\}$. For this plot, first the standardized hazard transforms of the realized point processes are calculated depending on θ and ordered with respect to $\leq_{\text{acc}}$. The plot then shows the contour lines of the normalized 3-sign depth $\Psi_3(\mathcal{W}^{\eta(J)}(\theta))$ based on these transforms as a function of θ for $J \in \{18, 30, 90, 180\}$ and the parameters & covariates given in Table 2. The intersections of the confidence regions for $\alpha \in \{0.01, 0.05, 0.1\}$ with the hyperplane can be identified from the levels; for example, the innermost yellow area corresponds to $\mathcal{C}^{(\text{depth})}_{J,0.9} \cap \{\theta_3 = \theta_3^*\}$.

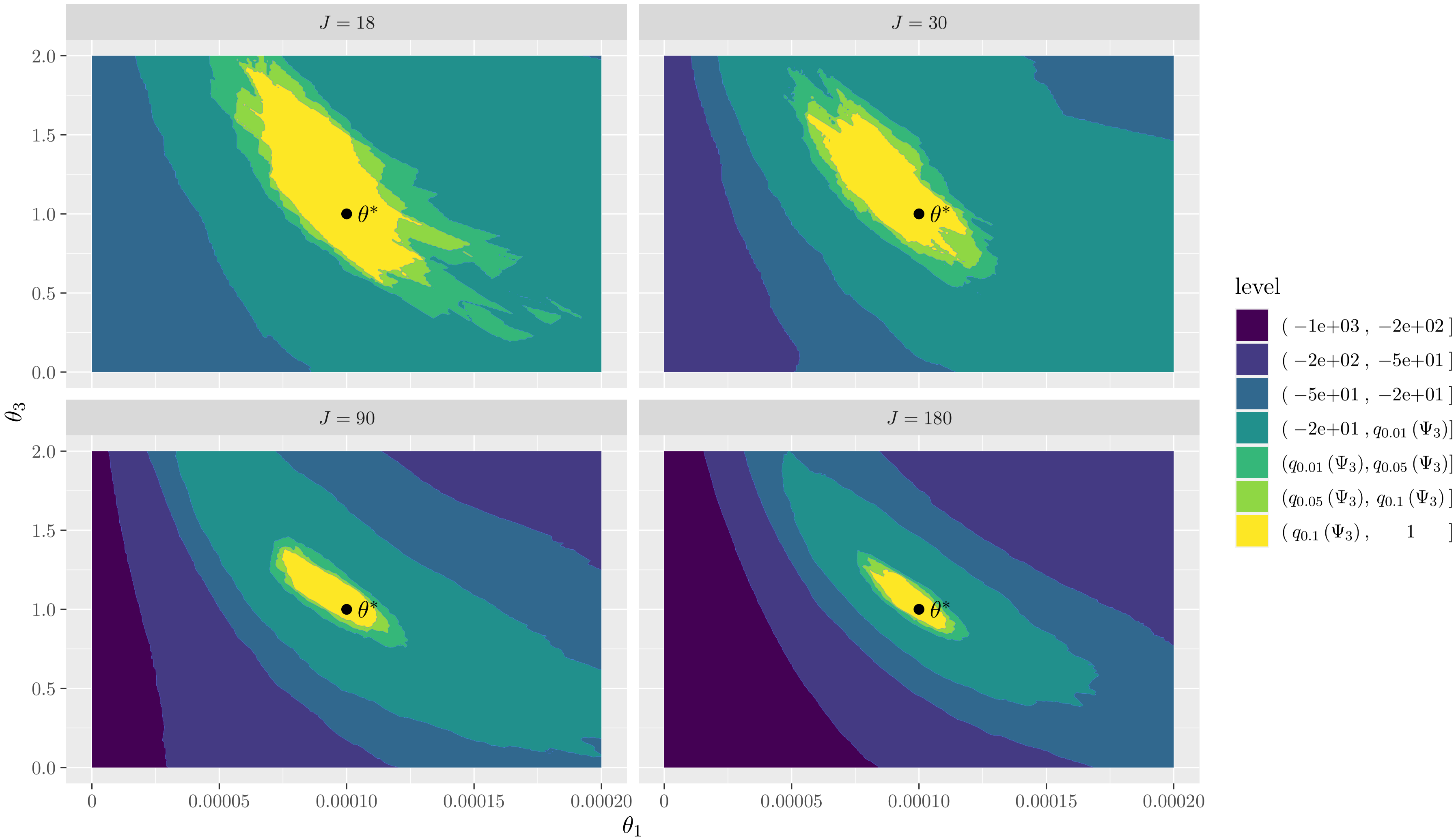


Figure 26: **3-Sign Depth, θ_1-θ_3-direction.** Visualization of $\mathcal{C}^{(\text{depth})}_{J,1-\alpha} \cap \{\theta_2 = \theta_2^*\}$, that is, the intersection of the $(1-\alpha)$-confidence region based on the 3-sign depth test with the hyperplane $\{\theta_2 = \theta_2^*\}$. For this plot, first the standardized hazard transforms of the realized point processes are calculated depending on θ and ordered with respect to $\leq_{\text{acc}}$. The plot then shows the contour lines of the normalized 3-sign depth $\Psi_3(\mathcal{W}^{\eta(J)}(\theta))$ based on these transforms as a function of θ for $J \in \{18, 30, 90, 180\}$ and the parameters & covariates given in Table 2. The intersections of the confidence regions for $\alpha \in \{0.01, 0.05, 0.1\}$ with the hyperplane can be identified from the levels; for example, the innermost yellow area corresponds to $\mathcal{C}^{(\text{depth})}_{J,0.9} \cap \{\theta_2 = \theta_2^*\}$.

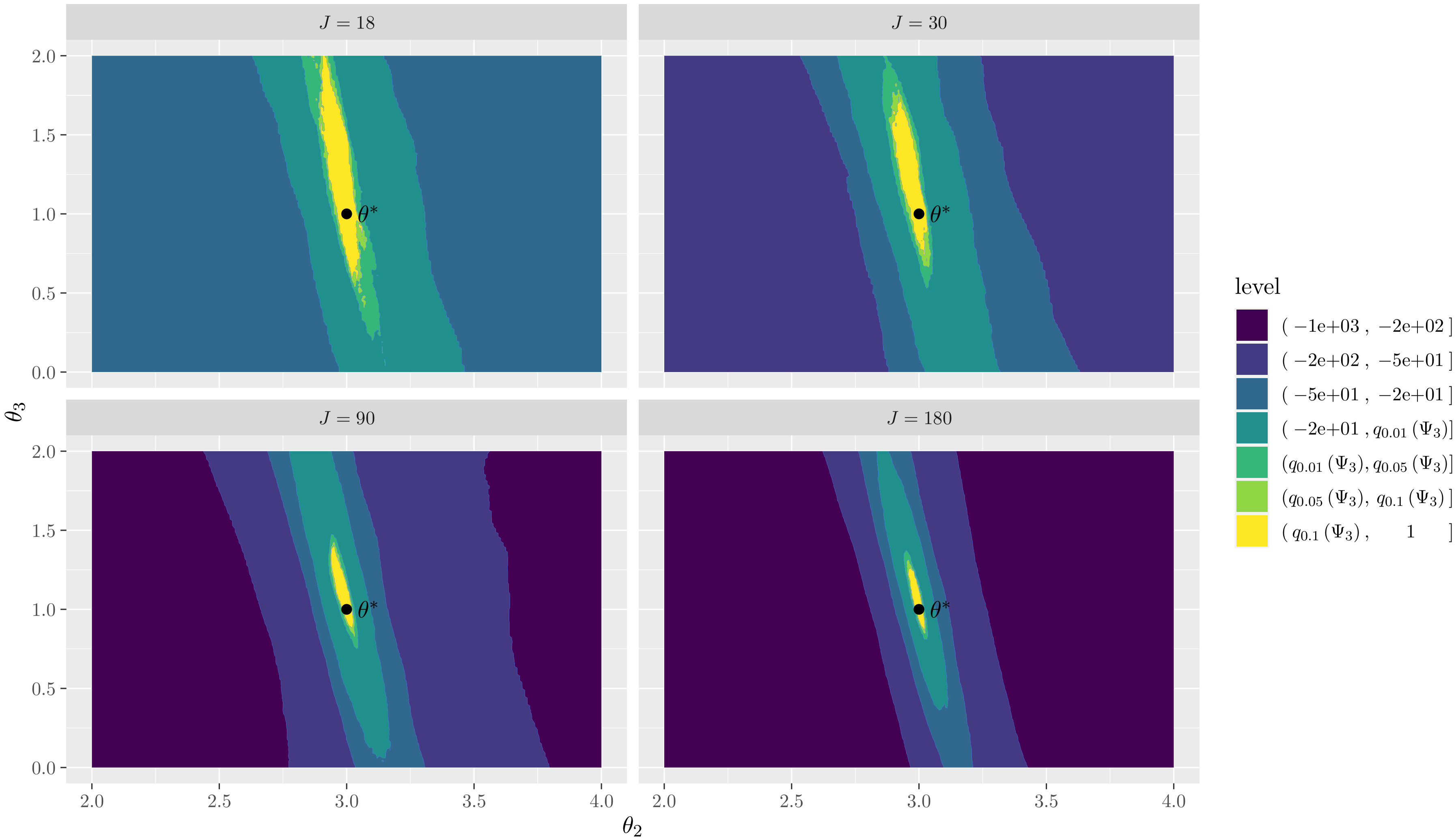


Figure 27: **3-Sign Depth, θ_2-θ_3-direction.** Visualization of $\mathcal{C}^{(\text{depth})}_{J,1-\alpha} \cap \{\theta_1 = \theta_1^*\}$, that is, the intersection of the $(1-\alpha)$-confidence region based on the 3-sign depth test with the hyperplane $\{\theta_1 = \theta_1^*\}$. For this plot, first the standardized hazard transforms of the realized point processes are calculated depending on θ and ordered with respect to $\leq_{\text{acc}}$. The plot then shows the contour lines of the normalized 3-sign depth $\Psi_3(\mathcal{W}^{\eta(J)}(\theta))$ based on these transforms as a function of θ for $J \in \{18, 30, 90, 180\}$ and the parameters & covariates given in Table 2. The intersections of the confidence regions for $\alpha \in \{0.01, 0.05, 0.1\}$ with the hyperplane can be identified from the levels; for example, the innermost yellow area corresponds to $\mathcal{C}^{(\text{depth})}_{J,0.9} \cap \{\theta_1 = \theta_1^*\}$.

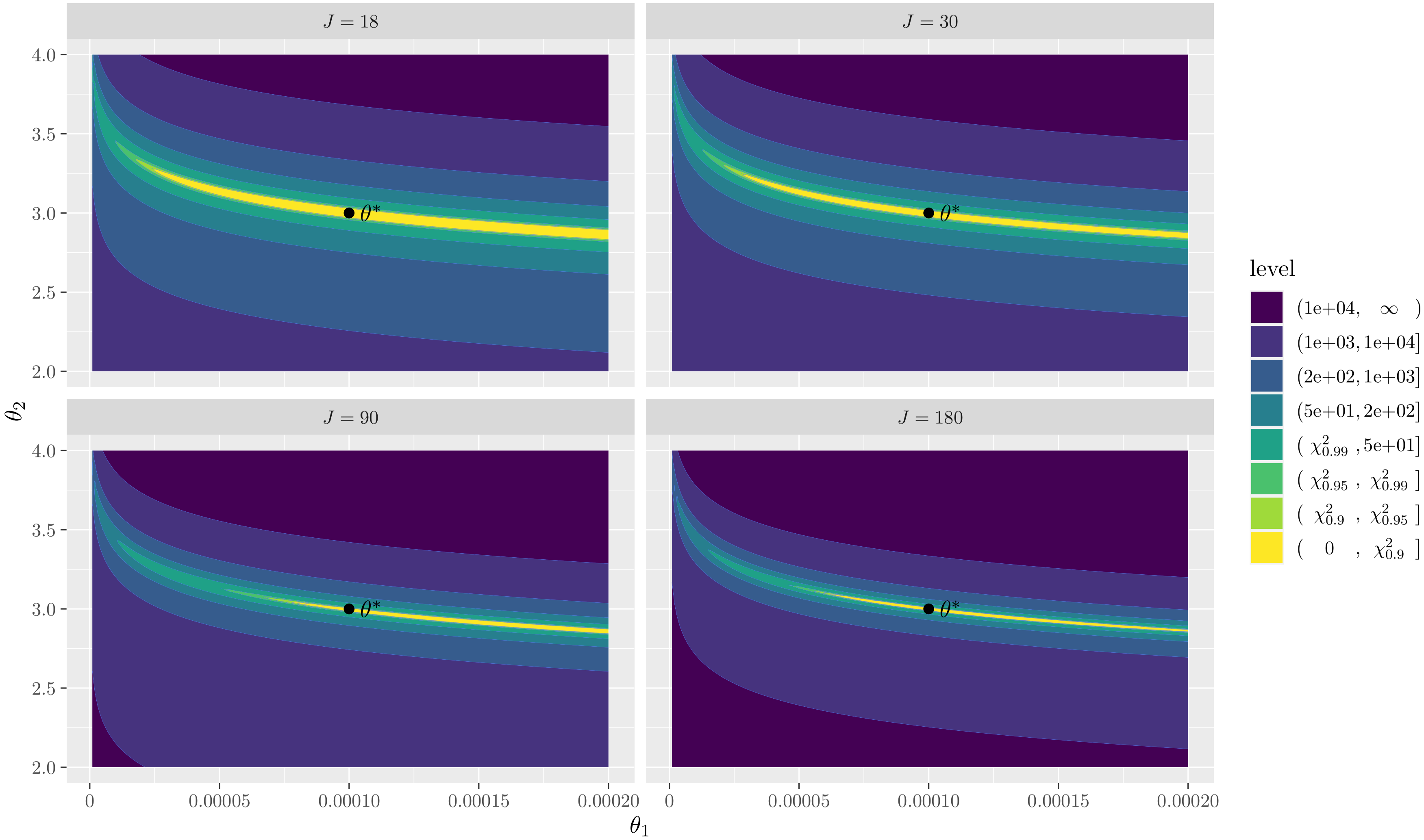


Figure 28: **Likelihood Ratio, θ_1-θ_2-direction.** Visualization of $\mathcal{C}^{(\mathrm{lr})}_{J,1-\alpha} \cap \{\theta_3 = \theta_3^*\}$, that is, the intersection of the $(1-\alpha)$-confidence region based on the likelihood ratio test with the hyperplane $\{\theta_3 = \theta_3^*\}$. The plot shows the contour lines of the likelihood ratio test statistic $\mathrm{LR}(\{\theta\}, \Theta)$ in the model $^{\times}\mathrm{D}$ as a function of θ for $J \in \{18, 30, 90, 180\}$ and the parameters & covariates given in Table 2. The intersections of the confidence regions for $\alpha \in \{0.01, 0.05, 0.1\}$ with the hyperplane can be identified from the levels; for example, the innermost yellow area corresponds to $\mathcal{C}^{(\mathrm{lr})}_{J,0.9} \cap \{\theta_3 = \theta_3^*\}$.

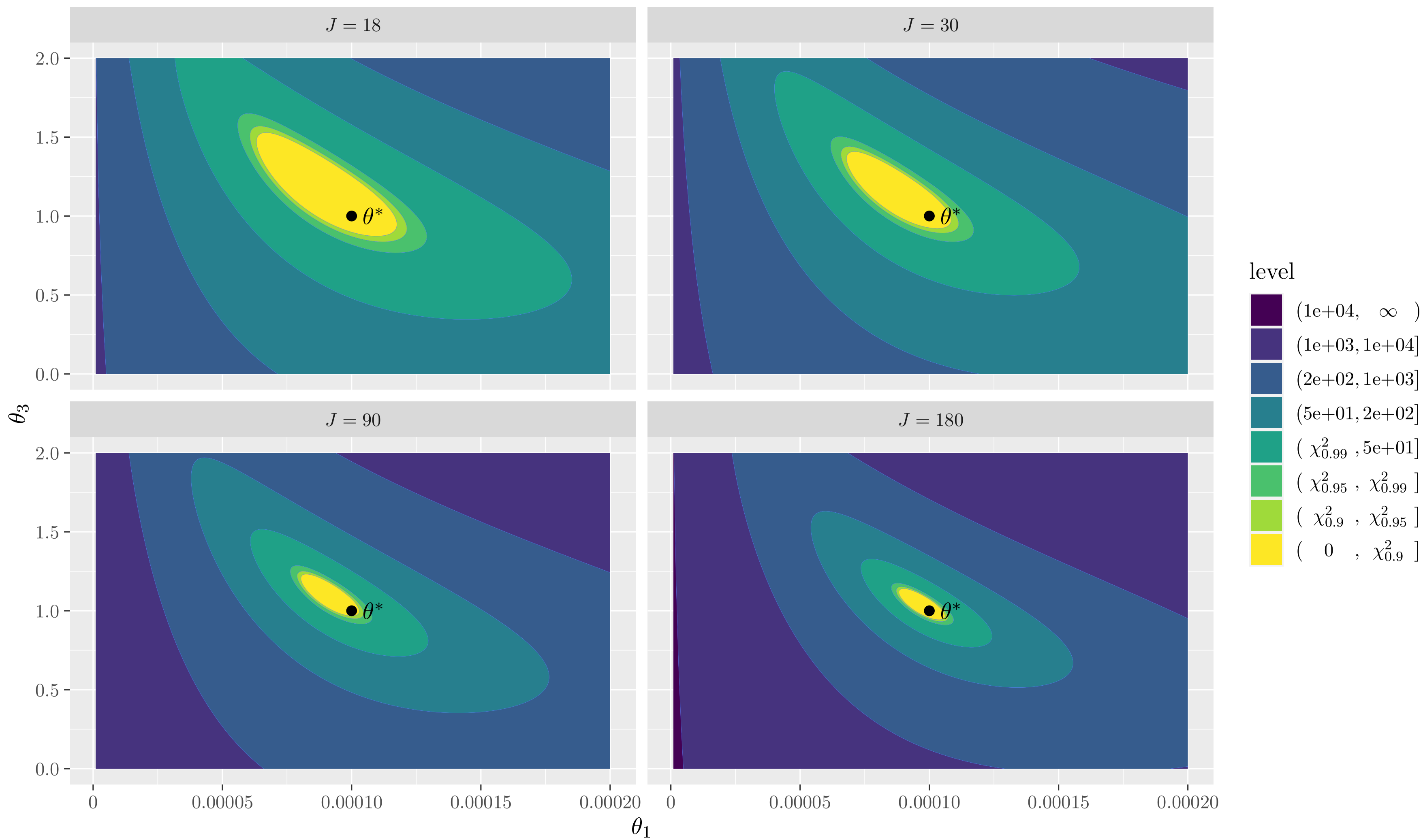


Figure 29: **Likelihood Ratio, θ_1-θ_3-direction.** Visualization of $\mathcal{C}^{(\mathrm{lr})}_{J,1-\alpha} \cap \{\theta_2 = \theta_2^*\}$, that is, the intersection of the $(1-\alpha)$-confidence region based on the likelihood ratio test with the hyperplane $\{\theta_2 = \theta_2^*\}$. The plot shows the contour lines of the likelihood ratio test statistic $\mathrm{LR}(\{\theta\}, \Theta)$ in the model $^{\times}\mathrm{D}$ as a function of θ for $J \in \{18, 30, 90, 180\}$ and the parameters & covariates given in Table 2. The intersections of the confidence regions for $\alpha \in \{0.01, 0.05, 0.1\}$ with the hyperplane can be identified from the levels; for example, the innermost yellow area corresponds to $\mathcal{C}^{(\mathrm{lr})}_{J,0.9} \cap \{\theta_2 = \theta_2^*\}$.

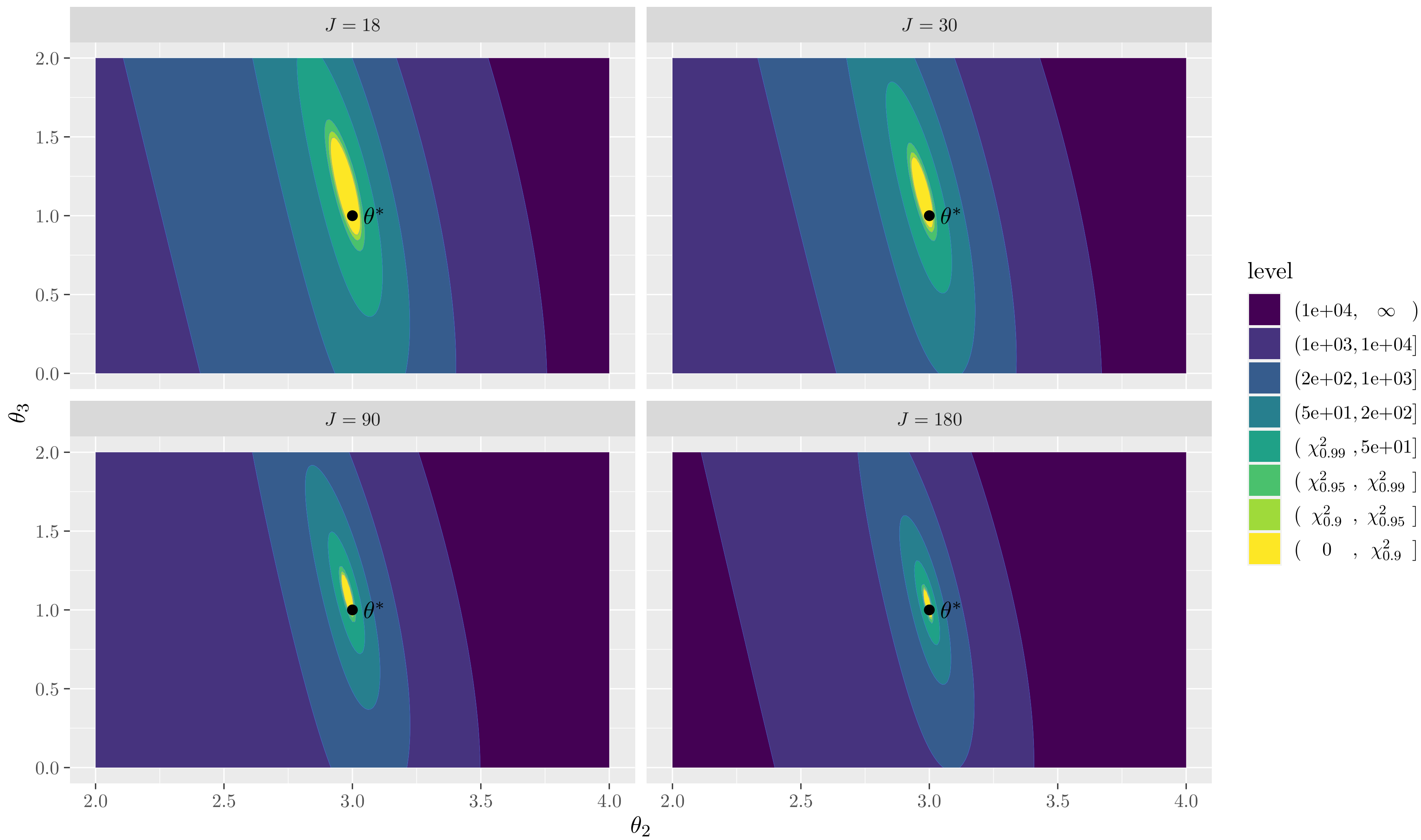


Figure 30: **Likelihood Ratio, θ_2-θ_3-direction.** Visualization of $\mathcal{C}^{(\mathrm{lr})}_{J,1-\alpha} \cap \{\theta_1 = \theta_1^*\}$, that is, the intersection of the $(1-\alpha)$-confidence region based on the likelihood ratio test with the hyperplane $\{\theta_1 = \theta_1^*\}$. The plot shows the contour lines of the likelihood ratio test statistic $\mathrm{LR}(\{\theta\}, \Theta)$ in the model $^{\times}$D as a function of θ for $J \in \{18, 30, 90, 180\}$ and the parameters & covariates given in Table 2. The intersections of the confidence regions for $\alpha \in \{0.01, 0.05, 0.1\}$ with the hyperplane can be identified from the levels; for example, the innermost yellow area corresponds to $\mathcal{C}^{(\mathrm{lr})}_{J,0.9} \cap \{\theta_1 = \theta_1^*\}$.

method	**J**	**coverage**	**# of grid points**					
		in %	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
Minimum Distance Estimator	9	97.15	0.00	30355.25	34068.00	32048.51	35459.25	36145.00
	18	97.50	0.00	19512.50	21570.00	20662.32	22803.25	23486.00
	30	93.75	0.00	9121.00	9344.00	8760.81	9416.00	9561.00
	90	88.85	106.00	1664.50	1830.00	1620.72	1856.00	1889.00
	180	86.95	32.00	409.00	578.00	487.12	614.00	647.00
3-Sign Depth Test	9	94.78	944.00	3469.00	4040.00	4086.14	4638.00	7454.00
	18	94.92	1137.00	1979.00	2271.00	2284.91	2577.00	4028.00
	30	94.88	740.00	1256.00	1416.00	1432.65	1595.00	2423.00
	90	94.33	264.00	422.00	470.00	474.85	523.00	769.00
	180	95.37	147.00	206.00	226.00	227.97	247.00	372.00
Likelihood Ratio	9	94.80	239.00	756.00	819.00	819.71	890.00	1137.00
	18	94.40	168.00	282.00	297.00	297.35	312.00	377.00
	30	95.15	109.00	132.00	138.00	137.64	144.00	172.00
	90	95.10	18.00	22.00	26.00	25.73	29.00	35.00
	180	94.70	5.00	9.00	10.00	9.79	11.00	13.00

Table 8: Comparison of 2000 simulated $(1-\alpha)$-confidence regions based on the minimum distance estimator, the 3-sign depth test and the likelihood ratio at $\alpha = 0.05$ and $J \in \{9, 18, 30, 90, 180\}$. The coverage rate in % indicates how often the true parameter θ^* was covered by the simulated confidence regions and should be equal to 95%. The table provides a five-point summary (plus the mean) of the size of the confidence regions. The size of a simulated confidence region was measured by counting the grid points of a $41 \times 41 \times 41$ grid on $[10^{-6}, 10^{-2}] \times [2, 4] \times [0, 2] \subset \Theta$ it contains.

Minimum Distance Estimator										
cont.	**J**	**rejection rate at parameter vector**								
		1	2	3	4	θ_0	6	7	8	9
no	9	0.024	0.029	0.025	0.024	0.026	0.031	0.028	0.035	0.036
	18	0.032	0.038	0.037	0.040	0.019	0.044	0.040	0.032	0.045
	30	0.037	0.035	0.048	0.046	0.038	0.045	0.048	0.045	0.053
	90	0.068	0.047	0.057	0.058	0.056	0.075	0.076	0.080	0.118
	180	0.096	0.059	0.052	0.069	0.060	0.084	0.096	0.118	0.276
d20	9	0.045	0.062	0.071	0.044	0.062	0.061	0.062	0.049	0.058
	18	0.087	0.075	0.080	0.079	0.041	0.099	0.066	0.080	0.040
	30	0.129	0.099	0.111	0.112	0.140	0.098	0.098	0.102	0.098
	90	0.286	0.212	0.176	0.192	0.268	0.159	0.170	0.198	0.175
	180	0.556	0.374	0.281	0.289	0.439	0.235	0.306	0.285	0.325
d40	9	0.089	0.101	0.091	0.074	0.096	0.098	0.089	0.080	0.095
	18	0.180	0.129	0.131	0.140	0.100	0.147	0.131	0.166	0.079
	30	0.264	0.206	0.203	0.214	0.246	0.198	0.192	0.190	0.161
	90	0.573	0.458	0.425	0.421	0.465	0.355	0.385	0.400	0.345
	180	0.881	0.745	0.695	0.645	0.721	0.571	0.613	0.579	0.569
q20	9	0.216	0.264	0.282	0.181	0.205	0.211	0.209	0.172	0.186
	18	0.385	0.359	0.354	0.323	0.256	0.330	0.291	0.304	0.236
	30	0.544	0.466	0.480	0.481	0.454	0.448	0.427	0.384	0.403
	90	0.907	0.840	0.794	0.791	0.720	0.738	0.760	0.733	0.649
	180	0.996	0.983	0.971	0.965	0.890	0.924	0.926	0.896	0.868
q40	9	0.281	0.316	0.249	0.229	0.251	0.229	0.226	0.214	0.216
	18	0.491	0.422	0.393	0.411	0.314	0.364	0.372	0.356	0.311
	30	0.640	0.568	0.541	0.562	0.480	0.516	0.510	0.459	0.431
	90	0.978	0.925	0.910	0.906	0.802	0.874	0.880	0.863	0.738
	180	0.999	0.999	1.000	0.995	0.966	0.993	0.994	0.976	0.950

Table 9: Rejection rates by data set of the level $\alpha = 0.05$ test for $\mathcal{H}_0 : \theta^* = \theta_0$ based on the minimum distance estimator at the 9 different parameter vectors of Table 6 and $J \in \{9, 18, 30, 90, 180\}$. The data sets differ in the proportion and type of contaminated data: no contamination (“no”), 20% depth-specific contamination (“d20”), 40% depth-specific contamination (“d40”), 20% quantile-based contamination (“q20”), and 40% quantile-based contamination (“q40”).

3-Sign Depth Test										
cont.	**J**	**rejection rate at parameter vector**								
		1	2	3	4	θ_0	6	7	8	9
no	9	0.774	0.264	0.108	0.061	0.055	0.067	0.101	0.292	0.858
	18	0.981	0.488	0.157	0.075	0.048	0.076	0.160	0.533	0.994
	30	0.999	0.723	0.242	0.097	0.049	0.095	0.241	0.778	1.000
	90	1.000	0.997	0.598	0.195	0.048	0.187	0.636	0.998	1.000
	180	1.000	1.000	0.900	0.330	0.051	0.347	0.917	1.000	1.000
d20	9	0.674	0.217	0.099	0.060	0.054	0.063	0.090	0.245	0.783
	18	0.946	0.399	0.134	0.072	0.048	0.071	0.135	0.452	0.980
	30	0.996	0.616	0.206	0.087	0.049	0.086	0.206	0.694	1.000
	90	1.000	0.982	0.507	0.166	0.048	0.161	0.545	0.991	1.000
	180	1.000	1.000	0.826	0.275	0.050	0.291	0.843	1.000	1.000
d40	9	0.557	0.179	0.089	0.056	0.055	0.060	0.084	0.207	0.696
	18	0.868	0.327	0.114	0.064	0.049	0.066	0.117	0.368	0.947
	30	0.983	0.512	0.172	0.079	0.049	0.078	0.172	0.589	0.997
	90	1.000	0.948	0.415	0.138	0.048	0.136	0.451	0.974	1.000
	180	1.000	1.000	0.718	0.226	0.050	0.239	0.753	1.000	1.000
q20	9	0.591	0.185	0.085	0.062	0.052	0.056	0.087	0.215	0.680
	18	0.893	0.351	0.119	0.068	0.048	0.070	0.119	0.389	0.946
	30	0.987	0.537	0.172	0.074	0.053	0.088	0.175	0.593	0.998
	90	1.000	0.964	0.433	0.140	0.047	0.141	0.464	0.980	1.000
	180	1.000	1.000	0.749	0.232	0.048	0.245	0.773	1.000	1.000
q40	9	0.372	0.124	0.064	0.061	0.059	0.058	0.074	0.152	0.430
	18	0.660	0.225	0.088	0.057	0.051	0.061	0.091	0.239	0.756
	30	0.884	0.341	0.117	0.067	0.047	0.065	0.121	0.374	0.936
	90	1.000	0.801	0.270	0.102	0.048	0.097	0.277	0.848	1.000
	180	1.000	0.984	0.488	0.154	0.048	0.155	0.522	0.990	1.000

Table 10: Rejection rates by data set of the level $\alpha = 0.05$ test for $\mathcal{H}_0 : \theta^* = \theta_0$ based on the 3-sign depth at the 9 different parameter vectors of Table 6 and $J \in \{9, 18, 30, 90, 180\}$. The data sets differ in the proportion and type of contaminated data: no contamination ("no"), 20% depth-specific contamination ("d20"), 40% depth-specific contamination ("d40"), 20% quantile-based contamination ("q20"), and 40% quantile-based contamination ("q40").

Likelihood Ratio										
cont.	**J**	**rejection rate at parameter vector**								
		1	2	3	4	θ_0	6	7	8	9
no	9	0.997	0.608	0.167	0.083	0.054	0.076	0.163	0.516	0.993
	18	1.000	0.896	0.313	0.112	0.048	0.099	0.288	0.853	1.000
	30	1.000	0.987	0.481	0.154	0.051	0.142	0.461	0.981	1.000
	90	1.000	1.000	0.944	0.385	0.043	0.346	0.943	1.000	1.000
	180	1.000	1.000	0.999	0.675	0.050	0.655	0.999	1.000	1.000
d20	9	1.000	0.838	0.483	0.318	0.200	0.150	0.172	0.364	0.945
	18	1.000	0.985	0.757	0.506	0.286	0.186	0.222	0.591	1.000
	30	1.000	0.999	0.921	0.708	0.393	0.237	0.276	0.801	1.000
	90	1.000	1.000	1.000	0.989	0.785	0.465	0.591	0.998	1.000
	180	1.000	1.000	1.000	1.000	0.971	0.708	0.862	1.000	1.000
d40	9	1.000	0.941	0.752	0.611	0.453	0.350	0.293	0.352	0.858
	18	1.000	0.998	0.947	0.855	0.675	0.496	0.404	0.504	0.990
	30	1.000	1.000	0.995	0.967	0.844	0.668	0.534	0.679	1.000
	90	1.000	1.000	1.000	1.000	0.999	0.973	0.895	0.973	1.000
	180	1.000	1.000	1.000	1.000	1.000	1.000	0.992	1.000	1.000
q20	9	1.000	0.999	0.999	0.998	0.991	0.980	0.947	0.809	0.395
	18	1.000	1.000	1.000	1.000	1.000	0.999	0.997	0.967	0.538
	30	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.701
	90	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.975
	180	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
q40	9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998	0.946
	18	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.998
	30	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	90	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	180	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 11: Rejection rates by data set of the level $\alpha = 0.05$ test for $\mathcal{H}_0 : \theta^* = \theta_0$ based on the likelihood ratio at the 9 different parameter vectors of Table 6 and $J \in \{9, 18, 30, 90, 180\}$. The data sets differ in the proportion and type of contaminated data: no contamination ("no"), 20% depth-specific contamination ("d20"), 40% depth-specific contamination ("d40"), 20% quantile-based contamination ("q20"), and 40% quantile-based contamination ("q40").

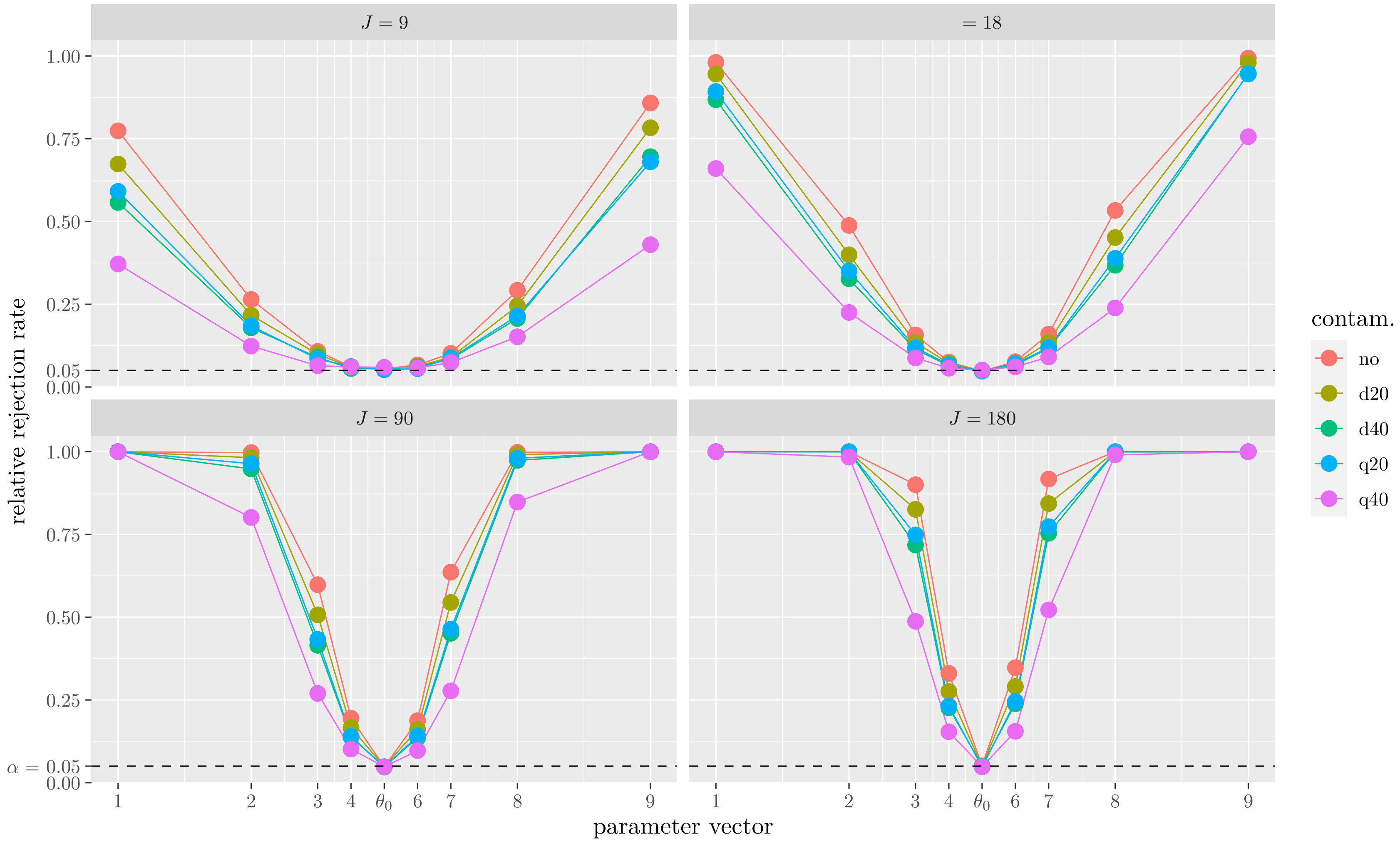


Figure 31: Rejection rates of the level $\alpha = 0.05$ 3-sign depth test with $J \in \{9, 18, 90, 180\}$ for $\mathcal{H}_0 : \theta^* = \theta_0$ at the 9 different parameter vectors of Table 6 and for the different types and proportions of contamination given in Table 5. The scaling of the x-axis was adjusted to reflect the actual distances of the parameter vectors to θ_0.

Bibliography

[1] Odd Aalen. 'Nonparametric Inference for a Family of Counting Processes'. In: *The Annals of Statistics* 6.4 (1978), pp. 701 –726. DOI: 10.1214/aos/1176344247.

[2] Odd Aalen. 'Statistical Inference for a Family of Counting Processes'. PhD thesis. University of California, Berkeley, 1975.

[3] P. K. Andersen and R. D. Gill. 'Cox's Regression Model for Counting Processes: A Large Sample Study'. In: *The Annals of Statistics* 10.4 (1982), pp. 1100–1120.

[4] Per Kragh Andersen et al. *Statistical Models Based on Counting Processes*. Springer Series in Statistics. New York et al.: Springer-Verlag, 1993. ISBN: 0-387-94519-9.

[5] Narayanaswamy Balakrishnan, Eric Beutner and Udo Kamps. 'Modeling Parameters of a Load-Sharing System Through Link Functions in Sequential Order Statistics Models and Associated Inference'. In: *IEEE Transactions on Reliability* 60.3 (2011), pp. 605–611. DOI: 10.1109/TR.2011.2161152.

[6] M. S. Bartlett. 'The Spectral Analysis of Point Processes'. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 25.2 (1963), pp. 264–296. ISSN: 00359246. URL: http://www.jstor.org/stable/2984295.

[7] Olin Hanson Basquin. 'The exponential law of endurance tests'. In: *American Society of Testing Materials* 10 (1910), pp. 625–630.

[8] Heinz Bauer. *Measure and Integration Theory*. Berlin, New York: De Gruyter, 2001. ISBN: 9783110866209. DOI: doi:10.1515/9783110866209.

[9] Heinz Bauer. *Probability Theory*. Berlin, New York: De Gruyter, 1996. ISBN: 9783110814668. DOI: doi:10.1515/9783110814668.

[10] Mathias Beiglböck, Walter Schachermayer and Bezirgen Veliyev. 'A short proof of the Doob-Meyer theorem'. In: *Stoch Process Their Appl.* 122.4 (Apr. 2012), pp. 1204–1209.

[11] P. J. Bickel and M. J. Wichura. 'Convergence Criteria for Multiparameter Stochastic Processes and Some Applications'. In: *The Annals of Mathematical Statistics* 42.5 (1971), pp. 1656 –1670. DOI: 10.1214/aoms/1177693164.

[12] Patrick Billingsley. *Convergence of Probabilty Measures*. Wiley series in probability and mathematical statistics. New York et al.: John Wiley & Sons, Inc., 1968. ISBN: 978-0-471-07242-3.

[13] René Boel, Pravin Varaiya and E. Wong. 'Martingales on Jump Processes. I: Representation Results'. In: *Siam Journal on Control* 13 (Aug. 1975), pp. 999–1021. DOI: 10.1137/0313063.

[14] Ørnulf Borgan. 'Maximum Likelihood Estimation in Parametric Counting Process Models, with Applications to Censored Failure Time Data'. In: *Scandinavian Journal of Statistics* 11.1 (1984), pp. 1–16. URL: http://www.jstor.org/stable/4615936 (visited on 10/30/2023).

[15] David Alexander Brannan. *A First Course in Mathematical Analysis*. Cambridge: Cambridge University Press, 2006. DOI: 10.1017/CBO9780511803949.

[16] Pierre Brémaud. 'A Martingale Approach to Point-Processes'. PhD thesis. EECS Department, University of California, Berkeley, Aug. 1972. URL: http://www2.eecs.berkeley.edu/Pubs/TechRpts/1972/7335.html.

[17] Pierre Brémaud. *Point Processes and Queues. Martingale Dynamics*. Springer Series in Statistics. New York, Heidelberg and Berlin: Springer-Verlag, 1981. ISBN: 0-387-90536-7.

[18] Pierre Brémaud. *Probability Theory and Stochastic Processes*. Cham: Springer-Verlag, 2020. ISBN: 978-3-030-40182-5.

[19] Ibrahim Burhan and Ho Sung Kim. 'S-N Curve Models for Composite Materials Characterisation: An Evaluative Review'. In: *Journal of Composites Science* 2.3 (2018). DOI: 10.3390/jcs2030038.

[20] D. R. Cox. 'Partial Likelihood'. In: *Biometrika* 62.2 (1975), pp. 269–276.

[21] D. R. Cox. 'Regression Models and Life-Tables'. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 34.2 (1972), pp. 187–220.

[22] David B. Dahl et al. *xtable: Export Tables to LaTeX or HTML*. R package version 1.8-4. 2019. URL: https://CRAN.R-project.org/package=xtable.

[23] D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods*. Probability and Its Applications. New York, Berlin and Heidelberg: Springer-Verlag, 2003. ISBN: 0-387-95541-0.

[24] D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes. Volume II: General Theory and Structure*. Second Edition. Probability and Its Applications. Springer-Verlag, 2008. ISBN: 978-0-387-21337-8.

[25] H. E. Daniels. 'The Statistical Theory of the Strength of Bundles of Threads. I'. In: *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* 183.995 (1945), pp. 405–435.

[26] Claude Dellacherie. *Capacités et Processus Stochastiques*. Berlin: Springer-Verlag, 1972. ISBN: 978-3-540-05676-8.

[27] Claude Dellacherie and Paul André Meyer. *Probabilities and Potential*. Paris, Amsterdam and New York: Hermann North-Holland Publisher Co. and Elsevier North-Holland, 1978. ISBN: 072040701X.

[28] Catherine Doléans-Dade. 'Existence du processus croissant naturel associé à un potentiel de la classe (D)'. In: *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* 9 (1968), pp. 309–314.

[29] François-Bernard Dolivo. *Counting processes and integrated conditional rates: A martingale approach with application to detection*. Tech. rep. College of Engineering, University of Michigan, 1974.

[30] Joseph L. Doob. 'Probability methods applied to the first boundary value problem'. In: *Proeedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*. (Statistical Laboratory of the University of California, Berkeley). Ed. by Jerzy Neyman. Vol. 2: Contributions to Probability Theory. 1956, pp. 49–80.

[31] Richard M. Dudley. *Real Analysis and Probability*. Vol. 74. Cambridge Studies in Advanced Mathematics. Revidierter Nachdruck der Originalausgabe von 1989. Cambridge: Cambridge University Press, 2002.

[32] Nelson Dunford and Jacob T. Schwartz. *Linear Operators 1: General Theory*. Vol. VII. Pure and Applied Mathematics: A Series of Texts and Monographs. New York: Interscience Publishers, Inc., 1957. ISBN: 978-0-470-22605-6.

[33] Stewart N. Ethier and Thomas G. Kurtz. *Markov Processes. Characterization and convergence*. Wiley series in probability and mathematical statistics. New York et al.: John Wiley & Sons, Inc., 1986. ISBN: 0-471-08186-8.

[34] Otto Forster. *Analysis 2. Differentialrechnung im $\mathbb{R}^n$, gewöhnliche Differentialgleichungen*. 11th ed. Grundkurs Mathematik. Wiesbaden: Springer Spektrum, 2017. ISBN: 978-3-658-19411-6.

[35] Carl Graham and Denis Talay. *Stochastic Simulation and Monte Carlo Methods. Mathematical Foundations of Stochastic Simulation*. Vol. 68. Stochastic Modelling and Applied Probability. Berlin and Heidelberg: Springer-Verlag, 2013. ISBN: 978-3-642-43840-0.

[36] E. F. Harding and D. G. Kendall, eds. *Stochastic Geometry. A Tribute to the Memory of Rollo Davidson*. Wiley series in probability and mathematical statistics. John Wiley & Sons, Ltd., 1974. ISBN: 0-471-35141-5.

[37] D. Gary Harlow and S. Leigh Phoenix. 'Probability Distributions for the Strength of Fibrous Materials under Local Load Sharing I: Two-Level Failure and Edge Effects'. In: *Advances in Applied Probability* 14.1 (1982), pp. 68–94.

[38] Alan G. Hawkes. 'Point Spectra of Some Mutually Exciting Point Processes'. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 33.3 (1971), pp. 438–443. ISSN: 00359246. URL: http://www.jstor.org/stable/2984686.

[39] Alan G. Hawkes. 'Spectra of Some Self-Exciting and Mutually Exciting Point Processes'. In: *Biometrika* 58.1 (1971), pp. 83–90. ISSN: 00063444. URL: http://www.jstor.org/stable/2334319.

[40] Wassily Hoeffding. 'A class of statistics with asymptotically normal distributions'. In: *Annals of Statistics* 19 (1948), pp. 293–325.

[41] Melanie Horn and Christine H. Müller. 'Sign depth tests in multiple regression'. In: *Journal of Statistical Computation and Simulation* 93.7 (2023), pp. 1169–1191. DOI: 10.1080/00949655.2022.2130922. URL: https://doi.org/10.1080/00949655.2022.2130922.

[42] Martin Jacobsen. *Point Process Theory and Applications. Marked Point and Piecewise Deterministic Processes*. Probability and its Applications. Boston: Birkhäuser, 2006. ISBN: 0-8176-4215-3.

[43] Jean Jacod. *On the stochastic intensity of a random point process over the half-line*. Tech. rep. Princeton, NJ: Department of Statistics, Princeton University, 1973.

[44] Mirko Alexander Jakubzik. 'Asymptotische Eigenschaften eines Minimum-Distanz-Schätzers für selbstanregende Punktprozesse'. MA thesis. TU Dortmund University, Feb. 2017. URL: https://stabi.statistik.tu-dortmund.de/storages/stabi-statistik/r/Masterarbeiten/Mirko_Jakubzik.pdf.

[45] John D. Kalbfleisch and Ross L. Prentice. *The Statistical Analysis of Failure Time Data*. Second Edition. Wiley series in probability and statistics. John Wiley & Sons, Inc., 2002. ISBN: 0-471-36357-X.

[46] Gopinath Kallianpur. *Stochastic Filtering Theory*. Vol. 13. Applications of Mathematics. New York, Heidelberg and Berlin: Springer-Verlag, 1980. ISBN: 0-387-90445-X.

[47] Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*. Vol. 113. Graduate texts in mathematics. New York: Springer-Verlag, 1988. ISBN: 0-387-96535-1.

[48] Alan F. Karr. *Point Processes and Their Statistical Inference*. 2. ed., rev. and expanded. Vol. 7. Probability: pure and applied; a series of textbooks and reference books. New York, Basel and Hong Kong: Dekker, 1991. ISBN: 0824785320.

[49] Alexander S. Kechris. *Classical Descriptive Set Theory*. Vol. 156. Graduate Texts in Mathematics. New York, Berlin and Heidelberg: Springer-Verlag, 1995. ISBN: 0-387-94374-9.

[50] H. Kim and P. H. Kvam. 'Reliability Estimation Based on System Data with an Unknown Load Share Rule'. In: *Lifetime Data Analysis* 10.1 (2004), pp. 83–94.

[51] John P. Klein and Melvin L. Moeschberger. *Survival Analysis. Techniques for Censored and Truncated Data*. Statistics for Biology and Health. New York: Springer-Verlag, 2003. ISBN: 978-0-387-21645-4. DOI: https://doi.org/10.1007/b97377.

[52] Achim Klenke. *Probability Theory. A Comprehensive Course*. Third Edition. Cham: Springer, 2020. ISBN: 978-3-030-56402-5.

[53] J. Kohout and S. Věchet. 'A new function for fatigue curves characterization and its multiple merits'. In: *International Journal of Fatigue* 23.2 (2001), pp. 175–183. DOI: https://doi.org/10.1016/S0142-1123(00)00082-7.

[54] Andrei N. Kolmogorov. 'Sulla Determinazione Empirica di Una Legge di Distribuzione'. In: *Giornale dell'Istituto Italiano degli Attuari* 4 (1933), pp. 83–91.

[55] Kai Kopperschmidt. 'Statistische Analyse von Kaufverhaltensmodellen'. PhD thesis. Justus-Liebig-Universität Gießen, Apr. 2005. URL: http://geb.uni-giessen.de/geb/volltexte/2005/2205/pdf/KopperschmidtKai-2005-05-27.pdf.

[56] Kai Kopperschmidt and Winfried Stute. 'The Statistical Analysis of Self-Exciting Point Processes'. In: *Statistica Sinica* 23 (2013), pp. 1273–1298.

[57] Victor M. Kruglov. 'On Natural and Predictable Processes'. In: *Sankhya A* 78 (2016), pp. 43–51.

[58] Hiroshi Kunita. *Stochastic flows and stochastic differential equations*. Vol. 24. Cambridge studies in advanced mathematics. Cambridge University Press, 1990. ISBN: 0-521-59925-3.

[59] Hiroshi Kunita. 'Tightness of Probability measures in $D([0,T];C)$ and $D([0,T];D)$'. In: *Journal of the Mathematical Society of Japan* 38.2 (1986), pp. 309 –334. DOI: 10.2969/jmsj/03820309. URL: https://doi.org/10.2969/jmsj/03820309.

[60] Hui-Hsiung Kuo. *Introduction to Stochastic Integration*. Universitext. Springer-Verlag, 2006. ISBN: 0-387-28720-5.

[61] Christoph P. Kustosz, Anne Leucht and Christine H. Müller. 'Tests Based on Simplicial Depth for AR(1) Models With Explosion'. In: *Journal of Time Series Analysis* 37 (2016), pp. 763–784.

[62] Christoph P. Kustosz, Martin Wendler and Christine H. Müller. 'Simplified Simplicial Depth for Regression and Autoregressive Growth Processes'. In: *Journal of Statistical Planning and Inference* 173 (2016), pp. 125–146. DOI: 10.1016/j.jspi.2016.01.005.

[63] Paul H. Kvam and Edsel A. Peña. 'Estimating Load-Sharing Properties in a Dynamic Reliability System'. In: *Journal of the American Statistical Association* 100.469 (2005), pp. 262–272.

[64] Petr Lachout. 'Billingsley-type tightness criteria for multiparameter stochastic processes'. In: *Kybernetika* 24.5 (1988), pp. 363–371. URL: http://www.kybernetika.cz/content/1988/5/363.

[65] Kevin Leckey, Mirko A. Jakubzik and Christine H. Müller. 'On the consistency of K-sign depth tests'. In: *Econometrics and Statistics* (2023). DOI: https://doi.org/10.1016/j.ecosta.2023.10.002.

[66] Kevin Leckey et al. 'Prediction intervals for load-sharing systems in accelerated life testing'. In: *Quality and Reliability Engineering International* 36.6 (2020), pp. 1895–1915. DOI: https://doi.org/10.1002/qre.2664.

[67] Kevin Leckey et al. 'Simple powerful robust tests based on sign depth'. In: *Statistical Papers* 64 (2023), pp. 857–882.

[68] A. J. Lee. *U-Statistics. Theory and Practice*. Vol. 110. Statistics: textbooks and monographs. New York: Marcel Dekker, Inc., 1990. ISBN: 0-8247-8253-4.

[69] Shiowjen Lee, S. Durham and J. Lynch. 'On the Calculation of the Reliability of General Load Sharing Systems'. In: *Journal of Applied Probability* 32.3 (1995), pp. 777–792.

[70] Shuang Li and James Lynch. 'On a threshold representation for complex load-sharing systems'. In: *Journal of Statistical Planning and Inference* 141.8 (2011), pp. 2811–2823. ISSN: 0378-3758. DOI: https://doi.org/10.1016/j.jspi.2011.03.006.

[71] Robert S. Liptser and Albert N. Shiryaev. *Statistics of Random Processes. I. General Theory*. Second, Revised and Expanded Edition. Vol. 5. Application of Mathematics: Stochastic Modelling and Applied Probability. Berlin, Heidelberg and New York: Springer-Verlag, 2001. ISBN: 978-3-540-63929-9.

[72] Regina Liu. 'On a Notion of Data Depth Based on Random Simplices'. In: *Ann. Stat* 18 (Mar. 1990). DOI: 10.1214/aos/1176347507.

[73] Dennis Malcherczyk. 'K-sign depth: Asymptotic distribution, efficient computation and applications'. PhD thesis. TU Dortmund University, 2022. URL: https://eldorado.tu-dortmund.de/handle/2003/40787.

[74] Dennis Malcherczyk, Kevin Leckey and Christine H. Müller. 'K-sign depth: From asymptotics to efficient implementation'. In: *Journal of Statistical Planning and Inference* 215 (2021), pp. 344–355.

[75] R. Maurer, Guido Heeke and G. Marzahn. 'Fatigue strength of prestressing steel tendons embedded in concrete of an aged highway bridge'. In: *Bauingenieur* 87 (2012), pp. 226–236.

[76] Paul-André Meyer. 'A decomposition theorem for supermartingales'. In: *Illinois Journal of Mathematics* 6.2 (1962), pp. 193 –205. DOI: 10.1215/ijm/1255632318.

[77] Paul-André Meyer. 'Decomposition of supermartingales: The uniqueness theorem'. In: *Illinois Journal of Mathematics* 7.1 (1963), pp. 1 –17. DOI: 10.1215/ijm/1255637477.

[78] Ivan Mizera. 'On Depth and Deep Points: A Calculus'. In: *The Annals of Statistics* 30.6 (2002), pp. 1681–1736. ISSN: 00905364. (Visited on 09/26/2023).

[79] Christine H. Müller. 'Depth estimators and tests based on the likelihood principle with application to regression'. In: *Journal of Multivariate Analysis* 95.1 (2005), pp. 153–181. DOI: https://doi.org/10.1016/j.jmva.2004.06.006.

[80] Christine H. Müller and Renate Meyer. 'Inference of Intensity-Based Models for Load-Sharing Systems With Damage Accumulation'. In: *IEEE Transactions on Reliability* 71.2 (2022), pp. 539–554. DOI: 10.1109/TR.2022.3140483.

[81] Christine Müller, Sebastian Szugat and Reinhard Maurer. 'Simulation free prediction intervals for a state dependent failure process using accellerated lifetime experiments'. In: *SFB Discussion Paper 45/16, TU Dortmund* (Jan. 2016). DOI: 10.17877/DE290R-17240.

[82] Whitney K. Newey. 'Uniform Convergence in Probability and Stochastic Equicontinuity'. In: *Econometrica* 59.4 (1991), pp. 1161–1167.

[83] Bernt Øksendal. *Stochastic Differential Equations. An Introduction with Applications*. 6th corrected printing. Universitext. Berlin and Heidelberg: Springer-Verlag, 2013. ISBN: 978-3-540-04758-2.

[84] Guodong Pang, Rishi Talreja and Ward Whitt. 'Martingale proofs of many-server heavy-traffic limits for Markovian queues'. In: *Probability Surveys* 4 (2007), pp. 193–267.

[85] Fredos Papangelou. 'Stochastic Geometry. A Tribute to the Memory of Rollo Davidson'. In: ed. by E. F. Harding and D. G. Kendall. Wiley series in probability and mathematical statistics. John Wiley & Sons, Ltd., 1974. Chap. On the Palm Probabilities of Processes of Points and Processes of Lines, pp. 114–147. ISBN: 0-471-35141-5.

[86] Kalyanapuram R. Parthasarathy. 'I - The Borel Subsets of a Metric Space'. In: *Probability Measures on Metric Spaces*. Ed. by Kalyanapuram R. Parthasarathy. Probability and Mathematical Statistics: A Series of Monographs and Textbooks. Academic Press, 1967, pp. 1–25. DOI: https://doi.org/10.1016/B978-1-4832-0022-4.50005-3.

[87] Ross L. Prentice and Steven G. Self. 'Asymptotic Distribution Theory for Cox-Type Regression Models with General Relative Risk Form'. In: *The Annals of Statistics* 11.3 (1983), pp. 804–813.

[88] Philip E. Protter. *Stochastic Integration and Differential Equations. A New Approach*. Second Edition, Version 2.1. Vol. 21. Stochastic Modelling and Applied Probability. Berlin, Heidelberg and New York: Springer-Verlag, 2005. ISBN: 3-540-00313-4.

[89] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2023. URL: https://www.R-project.org/.

[90] K. Murali Rao. 'On Decomposition Theorems of Meyer'. In: *Mathematica Scandinavica* 24 (Dec. 1969), pp. 66–78. DOI: 10.7146/math.scand.a-10920.

[91] B. Walter Rosen. 'Tensile failure of fibrous composites'. In: *AIAA Journal* 2.11 (1964), pp. 1985–1991. DOI: 10.2514/3.2699.

[92] Peter J. Rousseeuw and Mia Hubert. 'Regression Depth'. In: *Journal of the American Statistical Association* 94.446 (1999), pp. 388–402.

[93] Walter Rudin. *Real and complex analysis*. Third Edition. New York et al.: McGraw-Hill Book Company, 1987. ISBN: 0-07-054234-1.

[94] Robert J. Serfling. *Approximation Theorems of Mathematical Statistics*. Wiley series in probability and statistics. John Wiley & Sons, Inc., 1980. ISBN: 0-471-21927-4.

[95] Charlie Sharpsteen and Cameron Bracken. *tikzDevice: R Graphics Output in LaTeX Format*. R package version 0.12.5. 2023. URL: https://CRAN.R-project.org/package=tikzDevice.

[96] Albert N. Shiryaev. *Probability-1*. Third Edition. Graduate Texts in Mathematics. New York et al.: Springer-Verlag, 2016. ISBN: 978-0-387-72205-4.

[97] Anatoliy Skorokhod. 'Limit Theorems for Stochastic Processes'. In: *Theor. Probability Appl.* 3 (1956), pp. 261–290.

[98] Nikolai V. Smirnov. 'On the estimation of the discrepancy between empirical curves of distribution for two independent samples'. In: *Bulletin Mathématique de l'Université de Moscou* 2.2 (1939).

[99] Donald L. Snyder and Michael I. Miller. *Random Point Processes in Time and Space*. Second Edition. New York: Springer-Verlag, 1991. ISBN: 978-1-4612-7821-4.

[100] Fabio L. Spizzichino. 'Reliability, signature, and relative quality functions of systems under time-homogeneous load-sharing models'. In: *Applied Stochastic Models in Business and Industry* 35.2 (2019), pp. 158–176. DOI: https://doi.org/10.1002/asmb.2397.

[101] Winfried Stute. 'Parameter estimation in smooth empirical processes'. In: *Stochastic Processes and their Applications* 22.2 (1986), pp. 223–244. ISSN: 0304-4149. DOI: https://doi.org/10.1016/0304-4149(86)90003-7.

[102] Béla Szőkefalvi-Nagy. *Introduction to Real Functions and Orthogonal Expansions*. University Texts in the Mathematical Sciences. New York: Oxford University Press, 1965.

[103] Sebastian Szugat et al. 'Prediction Intervals for the Failure Time of Prestressed Concrete Beams'. In: *Advances in Materials Science and Engineering* 2016 (2016). Article ID 9605450.

[104] Haruo Totoki. 'A Method of Construction of Measures on Function Spaces and its Applications to Stochastic Processes'. In: *Memoirs of the Faculty of Science, Kyushu University. Series A, Mathematics* 15.2 (1962), pp. 178–190. DOI: 10.2206/kyushumfs.15.178.

[105] J. W. Tukey. 'Mathematics and the Picturing of Data'. In: *Proceedings of the International Congress of Mathematicians, Vancouver* 2 (1975), pp. 523–531.

[106] D. Wang, C. Jiang and C. Park. 'Reliability analysis of load-sharing systems with memory'. In: *Lifetime Data Anal* 25 (2019), pp. 341–360.

[107] Shinzo Watanabe. 'On discontinuous additive functionals and Lévy measures of a Markov process'. In: *Japanese journal of mathematics* 34 (1964), pp. 53–70.

[108] Waloddi Weibull. *Fatigue testing and analysis of results*. Oxford et al.: Pergamon Press, 1961.

[109] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN: 978-3-319-24277-4. URL: https://ggplot2.tidyverse.org.

[110] Kong Yaonan and Ye Zhisheng. 'Interval estimation for failure data from a load-sharing system'. In: *2015 First International Conference on Reliability Systems Engineering (ICRSE)*. 2015, pp. 1–7. DOI: 10.1109/ICRSE.2015.7366445.

[111] Jianchun Zhang, Yu Zhao and Xiaobing Ma. 'Reliability modeling methods for load-sharing k-out-of-n system subject to discrete external load'. In: *Reliability Engineering & System Safety* 193 (2020), p. 106603. ISSN: 0951-8320. DOI: https://doi.org/10.1016/j.ress.2019.106603.