SFB
823

Smooth estimation of rating transitions by nearest neighbors
-Consistency and Application-

Rafael Weißbach, Wladyslaw Poniatowski,
Walter Krämer

Nr. 2/2011

Discussion Paper

SFB
823

Smooth Estimation of Rating Transitions by Nearest Neighbors - Consistency and Application -

Rafael Weißbach[1] *, Wladyslaw Poniatowski[2] and Walter Krämer[3]

[1] Fakultät für Wirtschafts- und Sozialwissenschaften, Universität Rostock, Germany
[2] UMA Unister Market Research & Analysis, Düsseldorf, Germany
[3] Fakultät Statistik, Technische Universität Dortmund, Germany

January 10, 2011

Abstract

Many types of duration data suffer both from left-truncation and right-censoring. We show how these deficiencies can be overcome at the same time when estimating the hazard rate nonparametrically by kernel smoothing with the nearest neighbor method. We infer the uniform consistency of the estimate from the Hoeffding inequality, applied to a generalized empirical distribution function. Finally, we apply our estimator to rating transitions of corporate loans in Germany.

*Keywords. kernel smoothing, hazard rate, left-truncation, right-censoring
address for correspondence: Prof. Dr. Rafael Weißbach, Lehrstuhl Statistik, Institut für Volkswirtschaftslehre, Fakultät für Wirtschafts- und Sozialwissenschaften, Universität Rostock, 18051 Rostock, Germany, email: rafael.weissbach@uni-rostock.de, Phone: +49/381/4984428, Fax: +49/381/4984401.
JEL numbers; C14, C13, C24, C41, G24, G12
Financial support by Deutsche Forschungsgemeinschaft is gratefully acknowledged (SFB 823)

1 Introduction and Summary

Nonparametric estimation of the distribution of left-truncated duration data has a long history, see e.g. Turnbull (1976), Woodroofe (1985), Stute (1993) or Goto (1996). Cao et al. (2005) derive implications for the smooth estimation of the hazard rate and propose a kernel estimator for the relative hazard rate. In finance, hazard rates are important for estimating rating transitions, and left-truncation is a major concern, see Weißbach et al. (2009). Allowing for right-censoring only reduces the data set to observations originating after the start of the study. Weißbach et al. (2009) loose 50% of their data that way. Explicitly allowing for left-truncation therefore retains all observations and improves the efficiency of parameter estimates. In addition, for smoothing methods, a data-adaptive bandwidth improves the bias-variance trade-off and reduces the boundary bias near the origin (Weißbach et al. 2008). This is especially important for the analysis of durations of the types which concern us here which have the time origin as left boundary.

The current standard for kernel density estimation (with independent and identically distributed data) with data-dependent - and hence stochastic - bandwidth is strong uniform consistency (Einmahl and Mason 2005; Wied and Weißbach 2010). The present paper presents a nearest-neighbor kernel estimator of the hazard function for left-truncated durations and proves its uniform consistency. In doing so, we use the Hoeffding inequality in order to study the local oscillation behavior of the empirical distribution, similar to Schäfer (1986).

In our application we restrict ourselves to rating transitions into adjacent classes. This can be justified from any continuous model for the underlying asset. Also, in a proper rating system the rating transition intensities should not be class-specific, so we employ only one rating transition hazard and estimate it on the basis of existing selectors for the nearest neighbor bandwidth

(Weißbach et al. 2008). We find that ignoring left-truncated data increases the variability of the estimated hazard rates, in particular as regards to older debt.

2 Estimating and smoothing the cumulative hazard rate

Let T_i, $i = 1, \ldots, n^*$ be independent, nonnegative survival times. The T_i are observed only when $L_i \leq T_i \leq C_i$, where L_i denotes truncation on the left and $C_i \geq 0$ denotes censoring on the right; we condition on $L_i \leq C_i$. One therefore observes only $X_i := \min(T_i, C_i)$ and $\delta_i = \mathbb{1}_{\{T_i < C_i\}}$, or nothing at all (which happens whenever $L_i \geq T_i$). Without loss of generally, we assume this to happen for observations $i = n+1, \ldots, n^*$ where $n \leq n^*$. Figure 1 illustrates all possible scenarios.

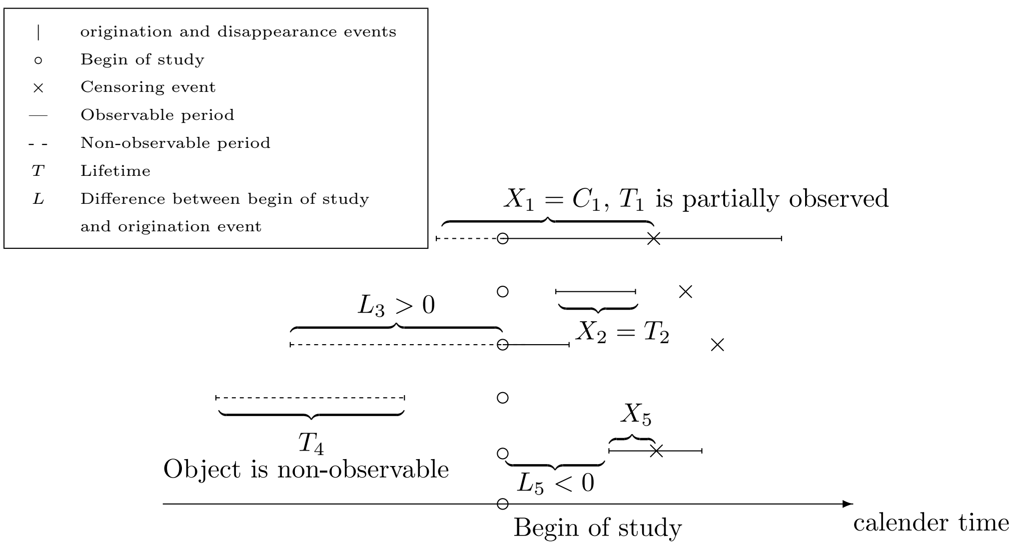

Figure 1: Possible scenarios when survival data are truncated on the left and censored on the right.

We impose the following assumptions:

(A1) $(T_i \in \mathbb{R}_0^+)_{i\in\mathbb{N}}$, $(C_i \in \mathbb{R}_0^+)_{i\in\mathbb{N}}$ and $(L_i \in \mathbb{R})_{i\in\mathbb{N}}$ are i.i.d. and independent from each other.

(A2) The respective distribution functions F, F^C, F^L und F^X are Lipschitz-continuous and strictly monotonic.

(A3) There exist constants $0 < A < B$ such that $F^L(A) > 0$ and $F^X(B) < 1$.

Given an estimate $\Lambda_n(\cdot)$ of the cumulative hazard rate $\Lambda(\cdot)$, one can estimate the hazard rate $\lambda(\cdot)$ via a kernel function $K(\cdot)$ such as

$$\lambda_n(t) := \int_{\mathbb{R}_0^+} \frac{1}{R_n(s)} K\left(\frac{t-s}{R_n(s)}\right) d\Lambda_n(s). \tag{1}$$

By defining a - possibly stochastic - monotonous function $\tilde{\Psi}_n(\cdot)$ and

$$R_n(t) := \inf\left\{r > 0 : \left|\tilde{\Psi}_n\left(t - r/2\right) - \tilde{\Psi}_n\left(t + r/2\right)\right| \geq p_n\right\}$$

we allow here both for a fixed bandwidth $R_n(t) \equiv b$, but also for a variable deterministic bandwidth $R_n(t) = R(t)$. Li and Li (2010) suggest the k-nearest neighbor bandwidth in various econometric contexts, extenting Gefeller and Dette (1992). The k-nearest neighbor bandwidth is a special case of $R_n(t)$ when $\tilde{\Psi}_n(\cdot)$ estimates the cumulative distribution function and p_n is equal to k/n. Throughout we require that the bandwidth parameter obeys the restrictions $0 < p_n < 1$, $p_n \longrightarrow 0$ and $\log(n)/(np_n) \longrightarrow 0$.

Under assumptions (A1)-(A3), $\lambda_n(\cdot)$ is uniformly consistent on the closed interval [A,B]. More precisely we have:

Lemma 1. *There exists a constant $0 \leq D < \infty$ such that*

$$P\left\{\limsup_{n\longrightarrow\infty} \frac{\sup_{t\in[a,b]} |\lambda_n(t) - \lambda(t)|}{\sqrt{\log(n)/(np_n) + p_n}} = D\right\} = 1 \qquad \forall\ [a,b] \in (A,B).$$

The proof is an application of Theorem 3.1 in Weißbach (2006). It is based on integration by parts: One decomposes the error into the total variation of the kernel and the local proximity of the stochastic processes $\Lambda_n(\cdot)$ to its limit $\Lambda(\cdot)$. The total variation is calculated in an elementary fashion. The contribution of the variability of the bandwidth to the error can be taken into account by adapting the proof in Schäfer (1986).

The crucial assumption in Weißbach (2006) requires the following local asymptotic behavior of the right-continuous and monotonous cumulative hazard rate estimator $\Lambda_n(\cdot)$: For some finite $0 \leq D < \infty$,

$$P\left\{ \lim \sup_{n \longrightarrow \infty} \frac{\sup_{I \subset [A,B], \Lambda(I) \leq p_n} |\Lambda_n(I) - \Lambda(I)|}{\sqrt{\log(n) p_n / n}} = D \right\} = 1. \tag{2}$$

We now show that the Cao et al. (2005) estimator of the cumulative hazard rate under left-truncation obeys equation (2). This is done in two steps. First, we construct a general estimator and show that it converges with the rate specified in (2), and then we establish the estimator of Cao et al. (2005) as a special case.

In the classical case of right-censored durations, one starts with the bivariate sample of events and censoring times. For additional left-truncation a third dimension is needed. Let $(\mathbf{S}_i)_{i=1,\ldots,n}$ be a sample of independent identically distributed random vectors $\mathbf{S}_i : \Omega \to \mathbb{R}^3$. The hazard function can be represented by the ratio of density and survival function. We now generalize the survival function and drop the monotonicity assumption. We assume a function $G : \mathbb{R}_0^+ \to \mathbb{R}_0^+$ to be continuous, accompanied by an estimator $G_n : \mathbb{R}_0^+ \times (\mathbb{R}^3)^n \to \mathbb{R}_0^+$, $(t, s_1, \ldots, s_n) \mapsto G_n(t)(s_1, \ldots, s_n)$ being symmetric for each fixed $t \in \mathbb{R}_0^+$ and $s_1, \ldots, s_n \in \mathbb{R}^3$. In addition, we use the simplified notation $G_n(t, \omega)$ for $G_n(t)(\mathbf{S}_1(\omega), \ldots, \mathbf{S}_n(\omega))$.

With respect to the handling of censoring, it will in addition prove useful to define a weight function by the mapping $\Delta : \mathbb{R}_0^+ \times \mathbb{R}^3 \to \mathbb{R}_0^+, (t, s) \mapsto \Delta^t(s)$ with simplified notation $\Delta_i^t(\omega)$ for $\Delta^t(\mathbf{S}_i(\omega))$. All $\Delta_i^t(\omega)$ are assumed to be bounded (from above) by a Δ_{max} on the interval $[A, B]$.

Let the function $\Psi : \mathbb{R} \to \mathbb{R}_0^+$ be continuous, positive and strictly monotonically increasing on the interval $[A, B]$. We now propose to estimate $\Psi(\cdot)$ by

$$\Psi_n(t) := \frac{1}{n} \sum_{i=1}^{n} \frac{\mathbb{1}_{\{S_i^1 \leq t\}} \cdot \Delta_i^t}{G_n(S_i^1)}, \tag{3}$$

where S_i^1 is the first element of the vector $\mathbf{S}_i$.

The local consistency of the estimate (3) requires some assumptions on the target function, on the observed random variables and on the rate of convergence of $G_n(\cdot)$ to $G(\cdot)$. Our result is as follows:

Theorem 2. *Under regularity conditions (A1)-(A3) and additional conditions (B1)-(B4) specified below, there exists a constant*

$$0 \leq D \leq 2(\sqrt{2 \cdot (2\Delta_{max} M + \Psi(B))} + D_G M)$$

such that

$$P\left\{\limsup_{n\to\infty} \frac{\sup_{I \subseteq [A,B], \Psi(I) \leq p_n} |\Psi_n(I) - \Psi(I)|}{\sqrt{\log(n) p_n / n}} = D\right\} = 1.$$

The additional regularity conditions are:

(B1) There exists a finite constant $M := \sup_{t \in [A,B]} [G(t)]^{-1}$.

(B2) $[\mathbb{1}_{\{t \leq a\}} \Delta_i^a - \mathbb{1}_{\{t \leq b\}} \Delta_i^b][G(t) - G_n(t)] = 0$ for all $t \notin [a, b] \subseteq [A, B]$.

(B3) For each fixed $t \in [A, B]$, $\mathbb{1}_{\{S_i^1 \leq t\}} \Delta_i^t / G(S_i^1)$ is an unbiased estimator for $\Psi(t)$. In case $(\mathbf{S}_i)_{i=1,\dots,n}$ are only observable under a condition, the estimator is conditionally unbiased.

(B4) For $G(t)$ and $G_n(t)$ let a constant $0 \leq D \leq D_G < \infty$ exist, such that

$$P\left\{\limsup_{n\to\infty} \frac{\sup_{t\in[A,B]} |G_n(t) - G(t)|}{\sqrt{\log(n)/n}} = D_G\right\} = 1.$$

The proof of Theorem 2 relies on the following preliminary estimator with known $G(\cdot)$:

$$\Psi_n^*(t) := \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbb{1}_{\{S_i^1 \leq t\}} \cdot \Delta_i^t}{G(S_i^1)}.$$

The aim is to split the difference $|\Psi_n(I) - \Psi(I)|$ into two parts using the measure $\Psi_n^*(I) := \Psi_n^*(b) - \Psi_n^*(a)$ for $I = [a, b]$ and to prove the almost sure convergence for each term separately. The complete proof of Theorem 2 is in Appendix A.

Using Theorem 2 we show next that the cumulative hazard rate estimator of Cao et al. (2005) has the required local convergence rate $\sqrt{\log(n)p_n/n}$. We let $F(\cdot)$ be the distribution function and $f(\cdot)$ the density function from Assumption (A2) of T, which we assume to exist. It is easily seen that

$$\Lambda(t) := \int_0^t \lambda(s)ds = \int_0^t \frac{dF^{X*}(s)}{G(s)}, \tag{4}$$

where $F^{X*}(t) := P(X_i \leq t, \delta_i = 1 | L_i \leq X_i)$ and $G(t) = P(L_i \leq t \leq X_i | L_i \leq X_i)$.

Cao et al. (2005) propose to estimate the cumulative hazard rate by

$$\Lambda_n(t) := \sum_{i=1}^{n} \frac{\mathbb{1}_{\{X_i \leq t, \delta_i = 1\}}}{nG_n(X_i)} = \sum_{i:X_{(i)} \leq t} \frac{\delta_i}{\sharp\{j : L_j \leq X_{(i)} \leq X_j\}}, \tag{5}$$

where summation occurs only over cases where $L_i \leq X_i$, and where $G_n(t) = n^{-1}\sum_{i=1}^{n} \mathbb{1}_{\{L_i \leq t \leq X_i\}}$ is the consistent estimate of $G(\cdot)$. This is the Nelson-Aalen estimator for right-censored observations, additionally allowing for "late entry into the under-risk set".

Corollary 3. *Given (A1)-(A3) and a positive zero sequence (p_n) with $np_n/\log(n) \to \infty$, there exists a constant $D \leq 2(\sqrt{2 \cdot (2M + \Lambda(B))} + 2M)$, such that*

$$P\left\{\limsup_{n\to\infty} \frac{\sup_{I\subseteq[A,B],\Lambda(I)\leq p_n} |\Lambda_n(I) - \Lambda(I)|}{\sqrt{\log(n)p_n/n}} = D\right\} = 1$$

with finite $M := \sup_{t\in[A,B]}[P(L_i \leq t \leq X_i | L_i \leq X_i)]^{-1}$.

To prove local convergence of $\Lambda_n(\cdot)$ defined in (5) we check the conditions (B1)-(B4) and denote its components as follows:

$\Delta_i^t := \mathbb{1}_{\{\delta_i=1\}} \leq 1 =: \Delta_{max}$ for $i = 1, \ldots, n$ and for each fixed $t \in [A, B]$, $\Psi(t) := \Lambda(t)$, $G(t) := P(L_i \leq t \leq X_i \mid L_i \leq X_i) > 0$ on $[A, B]$, $G_n(t) := \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{L_i\leq t\leq X_i\}}$ and $D_G := 2$. The remainder of the proof is in Appendix B.

To estimate the hazard rate $\lambda(t)$, we apply (1) to obtain

$$\lambda_n(t) := \sum_{i=1}^n \frac{1}{R_n(X_{(i)})} K\left(\frac{t - X_{(i)}}{R_n(X_{(i)})}\right) \frac{\delta_i}{\sharp\{j : L_j \leq X_{(i)} \leq X_j\}} \tag{6}$$

from $\Lambda_n(\cdot)$, where $R_n(\cdot)$ is the nearest-neighbor bandwidth.

One of the main conditions for the kernel estimation of $\lambda_n(\cdot)$ is the Lipschitz-continuity of $\lambda(\cdot)$ and $\Lambda(\cdot)$, which follows from the Lipschitz-continuity of $G(\cdot)$ and $F^{X*}(\cdot)$.

By assumption (A2) for $F^X(\cdot)$, is $F^{X*}(\cdot)$ likewise Lipschitz-continuous. Next we rewrite $G(\cdot)$ as follows to prove its Lipschitz-continuity:

$$\begin{aligned} G(t) &= \alpha^{-1}F^L(t)(1 - F(t))(1 - F^C(t)) \\ &= \int_{-\infty}^t \alpha^{-1}(1 - F(t))(1 - F^C(t))f^L(s)ds, \end{aligned}$$

where $\alpha := P(L_i \leq X_i)$. One can see that the Lipschitz-continuity of $F^L(\cdot)$ implies the Lipschitz-continuity of $G(\cdot)$.

3 An empirical application

Next we apply the techniques described above to data from WestLB AG, Düsseldorf, which provided us with rating transitions from an internal rating system with 8 non-default classes and 1 default class observed over seven years from January 1, 1997 to December 31, 2003. The time origin is the event of entering into WestLB's credit portfolio. The history contains about 600 transitions for 359 borrowers.

A constant hazard, a common assumption in business practice (see Bluhm et al., 2002), has been questioned by Kiefer and Larson (2007) and Weißbach et al. (2009). As an alternative, Weißbach and Walter (2010) propose a parametric piecewise constant model. The asset value model of Merton (1974) allows only transitions to adjacent classes. Other events like borrowers repaying their debt without having changed rating class or the end of the study are considered as right-censoring events. There is evidence that changing rating classes is not class-specific, i.e. does not depend on the class h from where the rating change starts, neither does it depend on the target class of a transition. Hence our model for rating transitions is

$$\lambda_{hj}(t) \equiv \lambda(t) \quad \text{for} \quad h = 1, \ldots, 8,\ j = 1, \ldots, 9,\ |h - j| = 1, \tag{7}$$

where $\lambda_{hj}(t) = 0$ for $|h - j| > 1$. Next we estimate $\lambda(t)$ using (1). Good results for the bandwidth $R_n(t)$ are to be expected for the nearest-neighbor method. The kernel function is known to have little impact; we use the bi-square kernel $K(t) = 15/16(1 - t^2)^2 \mathbb{1}_{\{|t| \leq 1\}}$.

We start by considering first transitions only. The first transition of each borrower are the events, their time since start are the durations T_i. If a borrower remains in its rating class for the entire study, T_i is unobservable and the maintenance time C_i is recorded (right-censoring). The potential second transition must be ruled out at that stage because the borrower is not constantly under risk to migrate up to the transition from the origin.

We observe the transitions of our data and 359 survivals, of which 60% are right-censored. We estimate the cumulative hazard rate by the Nelson-Aalen estimator

$$\Lambda_n(t) = \sum_{i:X_{(i)}\leq t} \frac{\delta_i}{\sharp\{j : X_{(i)} \leq X_j\}}.$$

The bandwidth parameter is crucial. We use three selectors. First a fast solution, adapting the rule of thumb of Silverman (1986), from Weißbach et al. (2008). For our 359 observations, the rule-of-thumb bandwidth results in $k = 123$ nearest neighbors. Second, the idea of cross-validation for the hazard rate under right-censored data and for the nearest-neighbor bandwidth. This is described in Gefeller et al. (1996) and results in $k = 78$ nearest neighbors. Third, a plug-in rule from Weißbach (2006) which yields $k = 38$. Results are displayed in Figure 2. Unfortunately, Bayes rules as in Zhang et al. (2009) are not available for censored survival times and the nearest neighbor bandwidth.

First of all, it is reassuring that all bandwidth selectors result in similarly shaped hazard rates. On the left edge, near the origin and up to one year, the hazard rate is small for all bandwidth selectors. It appears unlikely that the well-known boundary effect is the only reason because the nearest-neighbor bandwidth reduces the boundary bias, see Weißbach et al. (2008). And it is remarkable that Weißbach and Walter (2010) find the first year's transition intensities to be too low for the stationarity assumption. Hence, the nonparametric descriptive statistics reinforces previous parametric evidence. The mode at the one year duration seems to be an artifact of an increased rating activity after one year. Note that transitions to rating classes beyond the adjacent one are censored and do not even enter this analysis. As of now we cannot explain the second mode at three-and-a-half years. The plug-in seems to be under-smoothing.

Considering only the first transition for each rating history results in a loss

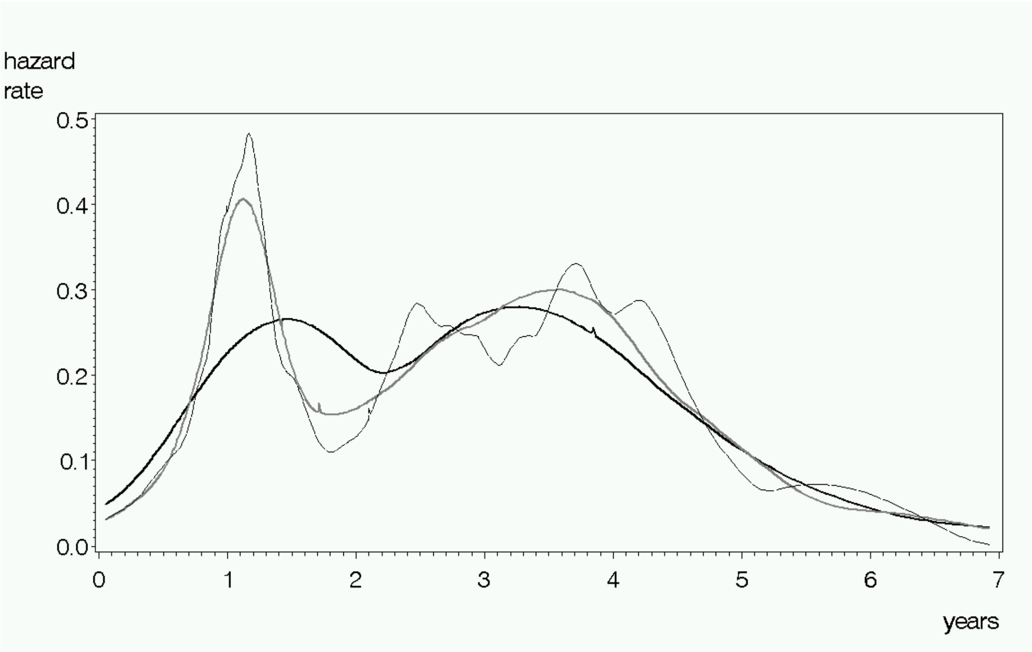

Figure 2: Estimated transition hazard $\lambda_n(t)$ from 359 right-censored rating transitions: Bandwidth selection by rule-of-thumb (thick-black), cross-validation (thick-grey), and plug-in (thin-black)

of 18% of the data (see Weißbach et al., 2009). This loss can be avoided by allowing for additional transitions later on. In particular, second transitions are now incorporated by means of left-truncation. In detail, for borrowers with more than one transition, the second transitions can be interpreted as an additional T_i subject to left-truncation L_i, where L_i is the first transition time. The object is not at risk to leave the rating under study until then. The second transition is again potentially right-censored by a C_i. There are some very rare third and further transitions which are treated similarly.

We now use estimator (6). Although the Markov property implies that the first transition and the second transition are independent, this in turn does not imply assumption (A1). Still we proceed by estimating the Markov process intensity with (7) for a sample of now 542 identically distributed univariate durations. We use the three bandwidths calculated above for the right-censored data set. We do this because we are interested in the improvement of the estimation that can be attributed to the additional observations. The implementation of the estimator is only available for the fixed bandwidth yet. However, the fixed bandwidth can be derived from the nearest neighbor bandwidth as in Weißbach et al. (2008). It is simply the number of nearest neighbors divided by the sample size (of the right-censored data) times twice the median (of the left-truncated and right-censored data set). Here, the sample size is 359 and the median is taken from the cumulative hazard rate estimate (5). A fixed bandwidth of 1.95 is optimal by the rule of thumb, in cross-validation 1.24 is optimal, and 0.60 is the optimal plug-in bandwidth. Figure 3 gives the results.

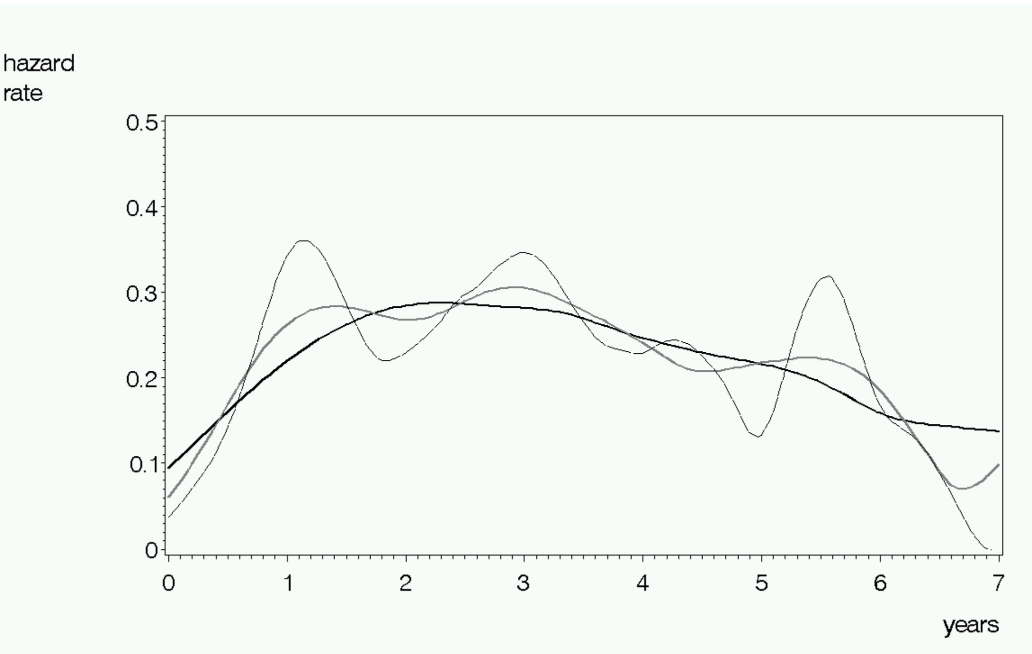

Figure 3: Estimating rating transition hazard for right-censored and left-truncated data: Bandwidth selection rule-of-thumb (thick-black), cross-validation (thick-grey), and plug-in (thin-black)

Allowing for left-truncated rating transition favors second (and third) transitions, which naturally occur later than the first. Therefore the additional 183 observations result in a more stable estimate of the hazard rate, especially from year 5 onwards. The second mode is not pronounced anymore in the rule-of-thumb smoothing. Allowing for left-truncation enables risk quantification of older debt. And overall the variability decreases, which results from using the same bandwidth as in the analysis with only right-censored data. The additional observations result in more nearest neighbors in the windows. The steep increase near the origin is confirmed, however, again only few observations are added for estimation in that region.

References

Bluhm, C., Overbeck, L., and Wagner, C. (2002). *An Introduction to Credit Risk Modeling*. Chapman & Hall.

Cao, R., Janssen, P., and Veraverbeke, N. (2005). Relative hazard rate estimation for right censored and left truncated data. *Test*, 14:257–280.

Einmahl, U. and Mason, D. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Annals of Statistics*, 33:1380–1403.

Gefeller, O. and Dette, H. (1992). Nearest neighbour kernel estimation of the hazard function from censored data. *Journal of Statistical Computation and Simulation*, 43:93–101.

Gefeller, O., Weißbach, R., and Bregenzer, T. (1996). The implementation of a data-driven selection procedure for the smoothing parameter in non-parametric hazard rate estimation using SAS/IML software. In Friedl, H., Berghold, A., and Kauermann, G., editors, *Proceedings of the 13th SAS European Users Group International Conference, Stockholm*, pages 1288–1300. SAS Institute Inc. Carry.

Goto, F. (1996). Achieving semiparametric efficiency bounds in left-censored duration models. *Econometrica*, 64(2):439–442.

Hewitt, E. and Savage, L. (1955). Symmetric measures on cartesian products. *Transactions of the American Mathematical Society*, 80:470–501.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30.

Kiefer, N. and Larson, C. (2007). A simulation estimator for testing the time homogeneity of credit rating transitions. *Journal of Empirical Finance*, 14:818–835.

Li, D. and Li, Q. (2010). Nonparametric/semiparametric estimation and testing of econometric models with data dependent smoothing parameters. *Journal of Econometrics*, 157:179 – 190.

Merton, R. (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance*, 29:449–470.

Schäfer, H. (1986). Local convergence of empirical measures in the random censorship situation with application to density and rate estimators. *Annals of Statistics*, 14:1240–1245.

Silverman, B. (1986). *Density Estimation.* Chapman & Hall, London.

Stute, W. (1993). Almost sure representations of the product-limit estimator for truncated data. *Annals of Statistics*, 21:146–156.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society, Series B*, 38:290–295.

Weißbach, R., Pfahlberg, A., and Gefeller, O. (2008). Double-smoothing in kernel hazard rate estimation. *Methods of Information in Medicine*, 47:167–173.

Weißbach, R., Tschiersch, P., and Lawrenz, C. (2009). Testing time-homogeneity of rating transitions after origination of debt. *Empirical Economics*, 36:575–596.

Weißbach, R. and Walter, R. (2010). A likelihood ratio test for stationarity of rating transitions. *Journal of Econometrics*, 155:188–194.

Weißbach, R. (2006). A general kernel functional estimator with general bandwidth - strong consistency and applications. *Journal of Nonparametric Statistics*, 18:1–12.

Wied, D. and Weißbach, R. (2010). Consistency of the kernel density estimator - a survey. *Statistical Papers*, to appear DOI: 10.1007/s00362-010-0338-1.

Woodroofe, M. (1985). Estimating a distribution function with truncated data. *Annals of Statistics*, 13:163–177.

Zhang, X., Brooks, R. D., and King, M. L. (2009). A bayesian approach to bandwidth selection for multivariate kernel regression with an application to state-price density estimation. *Journal of Econometrics*, 153:21 – 32.

A Proof of Theorem 2

The proof of Theorem 2 is in four steps. First, for an interval $I := [a, b] \subseteq [A, B]$ we establish an exponential bound for the distribution of the difference $|\Psi_n^*(I) - \Psi(I)|$:

$$P(|\Psi_n^*(I) - \Psi(I)| > \varepsilon) < 2\exp\left(\frac{-n\varepsilon^2}{2(2\Delta_{max}M + \Psi(B))(p+\varepsilon)}\right) \tag{8}$$

for all $p > 0$, $\varepsilon > 0$, $n \in \mathbb{N}_{>0}$ and for each fixed $I \subseteq [A, B]$ with $\Psi(I) \leq p$.

Because of definition (3) and the boundedness of $0 \leq \Delta_i^x \leq \Delta_{max} < \infty$

$$\Psi_n^*(I) - \Psi(I) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\mathbb{1}_{\{S_i^1 \leq b\}} \cdot \Delta_i^b}{G(S_i^1)} - \frac{\mathbb{1}_{\{S_i^1 \leq a\}} \cdot \Delta_i^a}{G(S_i^1)} - \Psi(I)\right) \tag{9}$$

is the arithmetic mean of the n independent and bounded random variables for each fixed $I \subseteq [A, B]$, distributed as

$$T_I := \frac{\mathbb{1}_{\{S_1^1 \leq b\}} \cdot \Delta_1^b}{G(S_1^1)} - \frac{\mathbb{1}_{\{S_1^1 \leq a\}} \cdot \Delta_1^a}{G(S_1^1)} - \Psi(I).$$

The expectation, the variance and the bound of T_I can then be calculated for fixed $I \subseteq [A, B]$ with $\Psi(I) \leq p$.

The expectation of T_I follows from assumption (B3):

$$E(T_I) = E\left(\frac{\mathbb{1}_{\{S_1^1 \le b\}} \cdot \Delta_1^b}{G(S_1^1)}\right) - E\left(\frac{\mathbb{1}_{\{S_1^1 \le a\}} \cdot \Delta_1^a}{G(S_1^1)}\right) - \Psi(b) + \Psi(a) = 0. \quad (10)$$

From assumption (B1), we get the following bound of $|T_I|$ on $[A, B]$:

$$\begin{aligned} |T_I| &= \left| \frac{\mathbb{1}_{\{S_1^1 \le b\}} \cdot \Delta_1^b}{G(S_1^1)} - \frac{\mathbb{1}_{\{S_1^1 \le a\}} \cdot \Delta_1^a}{G(S_1^1)} - \Psi(I) \right| \\ &< 2\Delta_{max} M + \Psi(B) - \Psi(A) < 2\Delta_{max} M + \Psi(B) =: g. \end{aligned} \quad (11)$$

The variance of T_I can be obtained from the expectation (10) and the bound (11) as follows:

$$\begin{aligned} \sigma_I^2 :=& Var(T_I) = E\left[\left(\frac{\mathbb{1}_{\{S_1^1 \le b\}} \cdot \Delta_1^b}{G(S_1^1)} - \frac{\mathbb{1}_{\{S_1^1 \le a\}} \cdot \Delta_1^a}{G(S_1^1)} - \Psi(I)\right)^2\right] \\ &< 2\Delta_{max} M \cdot E\left(\frac{\mathbb{1}_{\{S_1^1 \le b\}} \cdot \Delta_1^b}{G(S_1^1)} - \frac{\mathbb{1}_{\{S_1^1 \le a\}} \cdot \Delta_1^a}{G(S_1^1)}\right) \\ &= 2\Delta_{max} M \cdot \Psi(I) < g \cdot p. \end{aligned} \quad (12)$$

From equations (9), (10), (11), (12) and the inequality from Hoeffding (1963) results the following right bound:

$$P(|\Psi_n^*(I) - \Psi(I)| > \varepsilon) < 2\exp\left(\frac{-n\varepsilon^2}{2(\sigma_I^2 + g\varepsilon/3)}\right) < 2\exp\left(\frac{-n\varepsilon^2}{2g(p + \varepsilon)}\right)$$

for each fixed interval $I \subseteq [A, B]$ with $\Psi(I) \le p$.

In the second step we derive the inequality

$$\sup_{I \subseteq [A,B], \Psi(I) \le p_n} |\Psi_n^*(I) - \Psi(I)| \le C\sqrt{\log(n) p_n / n} \quad (13)$$

almost surely for a constant $C > \sqrt{2(2\Delta_{max} M + \Psi(B))}$ and large n.

On the right hand side of the inequality (8), p and ε can be substituted with p_n and $\varepsilon_n := C\sqrt{\log(n) p_n / n}$ for $C > 0$ and $n > 1$ altering the upper bound to

$$< 2 \cdot \exp\left(-\log(n)\frac{C^2}{2g}\frac{p_n}{(p_n + \varepsilon_n)}\right) = 2n^{-\frac{C^2}{2g}\frac{p_n}{(p_n + \varepsilon_n)}} =: A_n.$$

The series (A_n) is then summable starting from some large $n < \infty$ only if the exponent $\beta_n := (C^2 p_n)/(2g(p_n + \varepsilon_n)) > 1$. From $\varepsilon_n/p_n = C\sqrt{\log(n)/(np_n)}$ and the assumptions for p_n follow $\varepsilon_n/p_n \to 0$ and $p_n/(p_n + \varepsilon_n) \to 1$ for large n. The condition $\beta_n > 1$ can be then achieved with $C^2/2g > 1$ or $C > \sqrt{2g}$.

As a consequence, the series (A_n) is summable from some large $n < \infty$ and only for $C > \sqrt{2g}$. For each $I \subseteq [A, B]$ with $\Psi(I) \leq p_n$ we get then $\exists C > \sqrt{2g}\ \exists m < \infty, m \in \mathbb{N} : \sum_{n=m}^{\infty} P(|\Psi_n^*(I) - \Psi(I)| > \varepsilon_n) < \sum_{n=m}^{\infty} A_n < \infty$ and $\forall m < \infty, m \in \mathbb{N} : \sum_{n=1}^{m} P(|\Psi_n^*(I) - \Psi(I)| > \varepsilon_n) \leq m < \infty$.

Because of the summability of $P(|\Psi_n^*(I) - \Psi(I)| > \varepsilon_n)$,

$$P\left(\limsup_{n\to\infty} |\Psi_n^*(I) - \Psi(I)| > \varepsilon_n\right) = 0$$

results from the Borel-Cantelli lemma for $C > \sqrt{2g}$, i.e. $|\Psi_n^*(I) - \Psi(I)|$ does not exceed ε_n for most of the n. For large n and for all $I \subseteq [A, B]$ with $\Psi(I) \leq p_n$, we derive almost surely that $|\Psi_n^*(I) - \Psi(I))| \leq C\sqrt{\log(n)p_n/n}$.

The same inequality holds for the supremum of $|\Psi_n^*(I) - \Psi(I)|$ on $[A, B]$: $\sup_{I\subseteq[A,B],\Psi(I)\leq p_n} |\Psi_n^*(I) - \Psi(I)| \leq C\sqrt{\log(n)p_n/n}$ for $C > \sqrt{2g}$ and large n almost surely.

Using the results above we prove the following inequality in a third step:

$$\sup_{I\subseteq[A,B],\Psi(I)\leq p_n} |\Psi_n^*(I) - \Psi(I)| \leq C \cdot p_n\sqrt{\log(n)/n}$$

almost surely for some $C > D_G \cdot M$ and large n.

From the assumption (B4) and the limes superior formulation of Hewitt and Savage (1955) we get the right bound $G_n(x) - G(x) \leq |G_n(x) - G(x)| \leq \sup_{x\in[A,B]} |G_n(x) - G(x)| \leq C_1' \cdot \sqrt{\log(n)/n}$ almost surely for $C_1' > D_G$, large n and all $x \in [A, B]$. These bounds can be rewritten for $G_n(x)$ as follows: $G_n(x) \geq G(x) - C_1'\sqrt{\log(n)/n} \geq \inf_{t\in[A,B]} G(t) - C_1' \cdot \sqrt{\log(n)/n}$.

From assumption (B4) we have $\inf_{t\in[A,B]} G(t) > 0$. Because of

$\sqrt{\log(n)/n} \to 0$, the following inequalities hold for $x \in [A, B]$ and large n:

$$\inf_{t\in[A,B]} G(t) - C_1' \cdot \sqrt{\log(n)/n} > 0,$$

$$\frac{1}{G_n(x)} \leq \frac{1}{\inf\limits_{t\in[A,B]} G(t) - C_1' \cdot \sqrt{\log(n)/n}}$$

and

$$\frac{|G_n(x) - G(x)|}{G_n(x)} \leq \frac{C_1' \cdot \sqrt{\log(n)/n}}{\inf\limits_{t\in[A,B]} G(t) - C_1' \cdot \sqrt{\log(n)/n}}.$$

The following bounds for $\Psi_n^*(I) - \Psi(I)$ and $\Psi_n^*(I)$ result from the equation (13) almost surely for $I \subseteq [A, B]$ with $\Psi(I) \leq p_n$, large n and $C_2' > \sqrt{2 \cdot (2\Delta_{max} M + \Psi(B))}$:

$$\begin{aligned} \Psi_n^*(I) - \Psi(I) &\leq |\Psi_n^*(I) - \Psi(I)| \\ &\leq \sup_{I\subseteq[A,B],\Psi(I)\leq p_n} |\Psi_n^*(I) - \Psi(I)| \leq C_2' \sqrt{\log(n) p_n/n} \end{aligned}$$

and consequently $\Psi_n^*(I) \leq \Psi(I) + C_2'\sqrt{\log(n)p_n/n} \leq p_n + C_2'\sqrt{\log(n)p_n/n}$.

We then obtain the following equation from assumption (B2) almost surely for each $I \subseteq [A, B]$ with $\Psi(I) \leq p_n$ and large n:

$$\begin{aligned} &|\Psi_n^*(I) - \Psi(I)| = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{1}{G_n(S_i^1)} - \frac{1}{G(S_i^1)}\right|(\mathbb{1}_{\{S_i^1\leq b\}} \cdot \Delta_i^b - \mathbb{1}_{\{S_i^1\leq a\}} \cdot \Delta_i^a) \\ &\leq \frac{1}{n}\sum_{i=1}^{n}\left|\frac{G_n(S_i^1) - G(S_i^1)}{G_n(S_i^1)}\right|\frac{\mathbb{1}_{\{S_i^1\leq b\}} \cdot \Delta_i^b - \mathbb{1}_{\{S_i^1\leq a\}} \cdot \Delta_i^a}{G(S_i^1)} \\ &\leq \frac{C_1'\sqrt{\log(n)/n} \cdot \Psi_n^*(I)}{\inf\limits_{t\in[A,B]} G(t) - C_1'\sqrt{\log(n)/n}} \leq \frac{C_1'\sqrt{\log(n)/n} \cdot \left(p_n + C_2'\sqrt{\log(n)p_n/n}\right)}{\inf\limits_{t\in[A,B]} G(t) - C_1'\sqrt{\log(n)/n}}. \end{aligned}$$

By $p_n + C_2'\sqrt{\log(n)p_n/n} = p_n[1 + C_2'\sqrt{\log(n)/(p_n n)}]$ it is evident, that $C_2'\sqrt{\log(n)/(p_n n)}$ can be neglected for large n because of the assumptions for p_n. For large n, we can also neglect the term $\sqrt{\log(n)/n}$ in the numerator. For all $I \subseteq [A, B]$ with $\Psi(I) \leq p_n$ and for large n, we derive the inequality

$|\Psi_n^*(I) - \Psi(I)| \leq \frac{C_1'}{\inf_{t\in[A,B]} G(t)} p_n \sqrt{\log{(n)}/n} = C_1' \cdot M \cdot p_n \sqrt{\log{(n)}/n}$ almost surely.

The right bound $\sup_{I\subseteq[A,B],\Psi(I)\leq p_n} |\Psi_n^*(I) - \Psi(I)| \leq C \cdot p_n \sqrt{\log{(n)}/n}$ results for some $C > D_G \cdot M$ and large n almost surely.

In a final step we examine the expression $\sup_{I\subseteq[A,B],\Psi(I)\leq p_n} |\Psi_n(I) - \Psi(I)|$. This overall difference can be represented by the sum of the deviations of the empirical and theoretical measures $\Psi_n(I)$ and $\Psi(I)$ from the preliminary measure $\Psi_n^*(I)$ as follows: $\sup_{I\subseteq[A,B],\Psi(I)\leq p_n} |\Psi_n(I) - \Psi(I)|$
$\leq \sup_{I\subseteq[A,B],\Psi(I)\leq p_n} |\Psi_n(I) - \Psi_n^*(I)| + \sup_{I\subseteq[A,B],\Psi(I)\leq p_n} |\Psi_n^*(I) - \Psi(I)|$.

Because $p_n\sqrt{\log(n)/n}/\sqrt{\log(n)p_n/n} = \sqrt{p_n}$ approaches zero, i.e. $p_n\sqrt{\log(n)/n} \leq \sqrt{\log(n)p_n/n}$ holds for large n.

The previously mentioned right bounds of $|\Psi_n(I) - \Psi_n^*(I)|$ and $|\Psi_n^*(I) - \Psi(I)|$ imply the existence of a constant $C > \sqrt{2 \cdot (2\Delta_{max} M + \Psi(B))} + D_G \cdot M$, such that almost surely for large n

$$\begin{aligned}\sup_{I\subseteq[A,B],\Psi(I)\leq p_n} |\Psi_n(I) - \Psi(I)| &\leq C(\sqrt{\log{(n)}p_n/n} + p_n\sqrt{\log{(n)}/n}) \\ &\leq 2C\sqrt{\log{(n)}p_n/n}.\end{aligned}$$

Due to the symmetry of $\Psi_n(I)$ the limes superior formulation of the convergence follows from Hewitt and Savage (1955). □

B Proof of Corollary 3

The boundedness of the Δ_i^x for each $x \in [A, B]$ and conditions (B1) and (B2) follow from the definition of Δ_i^x. This is so because the variables Δ_i^x do not change over the x for each $i = 1, \ldots, n$.

From condition (B4), the consistency of the estimator $G_n(\cdot)$

$$P\left\{\limsup_{n\to\infty} \frac{\sup_{x\in[A,B]} |G_n(x) - G(x)|}{\sqrt{\log{(n)}/n}} = D\right\} = 1,$$

with a constant $0 \leq D \leq D_G$ can be easily shown.

The assumption (A2) for $F(\cdot)$ implies that the cumulative hazard rate $\Lambda(\cdot)$ grows strictly monotonously and the hazard rate $\lambda(\cdot)$ is obviously strictly positive on $[A, B]$.

Now only the condition (B3) needs to be shown. We note that the vectors $(L_i, X_i, \delta_i)_{i=1,\dots,n}$ are observable under $L_i \leq X_i$. Hence, we derive the following conditional expectation from (L6):

$$\begin{aligned} &E\left(\frac{\mathbb{1}_{\{X_i \leq x\}} \cdot \Delta_i^x}{G(X_i)} \mid L_i \leq X_i\right) = E\left(\frac{\mathbb{1}_{\{X_i \leq x, \delta_i = 1\}}}{G(X_i)} \mid L_i \leq X_i\right) \\ =&\sum_{\delta_1=0}^{1} \int_{-\infty}^{\infty} \frac{\mathbb{1}_{\{x_1 \leq x, \delta_1 = 1\}}}{G(x_1)} dF^{X,\delta}(x_1, \delta_1) = \int_{-\infty}^{x} \frac{dF^{X,\delta}(x_1, 1)}{G(x_1)}, \end{aligned} \tag{14}$$

where $F^{X,\delta}(x, y) = P(X \leq x, \delta \leq y \mid L \leq X)$ is the conditional distribution function of (X, δ).

The integral $\int_{x_1 \in I} dF^{X,\delta}(x_1, 1)$ for the intervals $I := [a, b] \subseteq [A, B]$ can now be calculated. First we express the probability $P(X_i \in I, \delta_i = 1 \mid L_i \leq X_i)$ in the terms of the non-observable vector (T_i, L_i, C_i) as follows:

$$\begin{aligned} &P(X_i \in I, \delta_i = 1 \mid L_i \leq X_i) = \alpha^{-1} P(X_i \in I, \delta_i = 1, L_i \leq X_i) \\ =&\alpha^{-1}[P(T_i \in I, T_i \leq C_i, L_i \leq T_i, T_i \leq C_i) \\ &+P(C_i \in I, T_i \leq C_i, L_i \leq C_i, C_i < T_i)] = \alpha^{-1} P(T_i \in I, L_i \leq T_i \leq C_i), \end{aligned} \tag{15}$$

where $\alpha = P(L_i \leq X_i)$. Hence, we write the probabilities $P(X_i \in I, \delta_i = 1 \mid L_i \leq X_i)$ and $P(T_i \in I, L_i \leq T_i \leq C_i)$ as the following expectations of the Bernoulli-variables:

$$\begin{aligned} &P(X_i \in I, \delta_i = 1 \mid L_i \leq X_i) = E(\mathbb{1}_{\{X_i \in I, \delta_i = 1\}} \mid L_i \leq X_i) \\ =&\sum_{\delta_1=0}^{1} \int_{-\infty}^{\infty} \mathbb{1}_{\{x_1 \in I, \delta_1 = 1\}} dF^{X,\delta}(x_1, \delta_1) = \int_{x_1 \in I} dF^{X,\delta}(x_1, 1) \end{aligned} \tag{16}$$

and

$$\begin{aligned} &\alpha^{-1} P(T_i \in I, L_i \leq T_i \leq C_i) = \alpha^{-1} E(\mathbb{1}_{\{T_i \in I, L_i \leq T_i \leq C_i\}}) \\ =&\int_{t \in \mathbb{R}} \int_{c \in \mathbb{R}} \int_{l \in \mathbb{R}} \alpha^{-1} \mathbb{1}_{\{t \in I\}} \mathbb{1}_{\{l \leq t\}} \mathbb{1}_{\{t \leq c\}} dF(t) dF^C(c) dF^L(l) \\ =&\int_{t \in I} \alpha^{-1} F^L(t)(1 - F^C(t)) dF(t). \end{aligned} \tag{17}$$

One can see that $dF^{X,\delta}(x,1) = \alpha^{-1}F^L(j)(1-F^C(j))dF(x)$ follows from the expressions (15), (16) and (17). Consequently the expectation (14) can be written as follows:

$$E\left(\frac{\mathbb{1}_{\{X_i \le x\}} \cdot \Delta_i^x}{G(X_i)} \mid L_i \le X_i\right) = \int_{-\infty}^{x} \frac{dF^{X,\delta}(x_1,1)}{G(x_1)}$$
$$= \int_{-\infty}^{x} \frac{\alpha^{-1}F^L(x_1)(1-F^C(x_1))dF(x_1)}{G(x_1)}$$
$$= \int_{-\infty}^{x} \frac{\alpha^{-1}F^L(x_1)(1-F^C(x_1))dF(x_1)}{\alpha^{-1}F^L(x_1)(1-F^C(x_1))(1-F(x_1))} = \int_{-\infty}^{x} \frac{dF(x_1)}{1-F(x_1)} = \Lambda(x) = \Psi(x).$$

Obviously the conditions (B1)-(B4) are fulfilled and the local convergence

$$P\left\{\limsup_{n\to\infty} \frac{\sup_{I\subseteq[A,B],\Lambda(I)\le p_n} |\Lambda_n(I)-\Lambda(I)|}{\sqrt{\log(n)p_n/n}} = D\right\} = 1$$

follows for a constant $D \le 2(\sqrt{2\cdot(2M+\Lambda(B))} + 2M)$. □