Are deviations in a gradually varying mean relevant? A testing approach based on sup-norm estimators

Axel Bücher, Holger Dette,
Florian Heinrichs

ARE DEVIATIONS IN A GRADUALLY VARYING MEAN RELEVANT? A TESTING APPROACH BASED ON SUP-NORM ESTIMATORS

AXEL BÜCHER, HOLGER DETTE, AND FLORIAN HEINRICHS

ABSTRACT. Classical change point analysis aims at (1) detecting abrupt changes in the mean of a possibly non-stationary time series and at (2) identifying regions where the mean exhibits a piecewise constant behavior. In many applications however, it is more reasonable to assume that the mean changes gradually in a smooth way. Those gradual changes may either be non-relevant (i.e., small), or relevant for a specific problem at hand, and the present paper presents statistical methodology to detect the latter. More precisely, we consider the common nonparametric regression model $X_i = \mu(i/n) + \varepsilon_i$ with possibly non-stationary errors and propose a test for the null hypothesis that the maximum absolute deviation of the regression function μ from a functional $g(\mu)$ (such as the value $\mu(0)$ or the integral $\int_0^1 \mu(t)dt$) is smaller than a given threshold on a given interval $[x_0, x_1] \subseteq [0,1]$. A test for this type of hypotheses is developed using an appropriate estimator, say $\hat{d}_{\infty,n}$, for the maximum deviation $d_\infty = \sup_{t\in[x_0,x_1]} |\mu(t) - g(\mu)|$. We derive the limiting distribution of an appropriately standardized version of $\hat{d}_{\infty,n}$, where the standardization depends on the Lebesgue measure of the set of extremal points of the function $\mu(\cdot) - g(\mu)$. A refined procedure based on an estimate of this set is developed and its consistency is proved. The results are illustrated by means of a simulation study and a data example.

Key words: relevant change point analysis, gradual changes, maximum deviation, local-linear estimator, Gumbel distribution, Gaussian approximation.

AMS Subject classification: 62M10, 62G08.

1. INTRODUCTION

Change point analysis has found considerable interest in the last two decades because of its numerous applications in economics, climatology, engineering, hydrology, genomics, to mention just a few. Most of the recent results are well documented in the reviews by Aue and Horváth (2013), Jandhyala et al. (2013), Woodall and Montgomery (2014), Sharma et al. (2016), Chakraborti and Graham (2019), and Truong et al. (2020) among others. In the simplest case one is interested in identifying structural breaks in a sequence of means $(\mu_i)_{i=1,\dots,n}$ of a possibly non-stationary time series $(X_i)_{i=1,\dots,n}$. Often the data are modelled by a location scale model $X_i = \mu(i/n) + \varepsilon_i$ with a stationary error process $(\varepsilon_i)_{i=1,\dots,n}$ and a piecewise constant mean function $\mu : [0,1] \to \mathbb{R}$. A large amount of the literature on this problem refers to functions with exactly one change point (see, e.g. Priestley and Subba Rao, 1969; Wolfe and Schechtman, 1984; Horváth et al., 1999, among others) but more recently the problem of detecting multiple change points in a piecewise constant mean has also found

considerable attention (see, e.g. Frick et al., 2014; Fryzlewicz, 2018; Dette et al., 2018; Baranowski et al., 2019, among many others). In these cases the null hypothesis of no change point can be formulated as

$$H_0 : \text{ there exists } \mu \in \mathbb{R} \text{ such that } \mu(t) = \mu \text{ for all } t \in [0,1], \qquad (1.1)$$

while, under the alternative, the authors make the assumption that the process is stationary on time spans with constant mean. This assumption simplifies the statistical analysis of structural breaks substantially.

While the assumption of a piecewise constant mean function is well justified in some applications (see for example Aston and Kirch, 2012; Hotz et al., 2013; Cho and Fryzlewicz, 2015; Kirch et al., 2015), there are also many other situations where it is more realistic to assume that the function μ varies smoothly in the interval $[0,1]$. Typical examples include temperature data (see, e.g. Karl et al., 1995; Collins et al., 2000) or financial data (see, e.g., Vogt and Dette, 2015). In these cases, it is more reasonable to assume that the regression function changes smoothly and one might be interested in deciding whether these changes deviate in some sense "substantially" from a given benchmark. For example, if $\mu(0)$ denotes the initial mean at "time" 0, it is often of interest whether the mean stays within a certain corridor of width $\Delta > 0$, that is

$$H_0 : d_\infty = \sup_{t \in [0,1]} |\mu(t) - \mu(0)| \le \Delta \quad \text{vs.} \quad H_1 : d_\infty > \Delta. \qquad (1.2)$$

Here $\Delta \ge 0$ defines a pre-specified constant that is chosen according to the specific interests for the problem at hand. In case of rejection of H_0, a statistician may consider to subsequently analyze the data with a time dependent mean. Note that the null hypothesis in (1.1) is obtained from (1.2) for $\Delta = 0$. Other benchmarks could be used as well and it could also be reasonable to consider deviations on a sub-interval of $[0,1]$. For example, in climate research one might be interested in significant deviations of a trend from an average trend in the past, and this problem could be considered investigating the hypotheses

$$H_0 : d_\infty = \sup_{t \in [x_0,1]} \left| \mu(t) - \frac{1}{x_0} \int_0^{x_0} \mu(s) \, ds \right| \le \Delta \quad \text{vs.} \quad H_1 : d_\infty > \Delta, \qquad (1.3)$$

for some given fixed constant $x_0 \in (0,1)$. The consideration of hypotheses of the form (1.2) or (1.3) may also be motivated by the fact that the detection of structural breaks in the signal often results in an adaptation of the statistical analysis (for example in forecasting). Because such an analysis is usually performed "locally", resulting estimators will have a smaller bias but a larger variance. However, if the changes in the signal are only weak, such an adaption might not be necessary because a potential decrease in bias might be overcompensated by an increase of variance.

In this paper we develop statistical methodology to investigate hypotheses of the form (1.2) and (1.3) (and further hypotheses of similar type) in a location scale model with a stationary error process and a smooth mean function. Additionally, we also construct estimators for the first point in time where the null hypothesis is violated. Note that this problem is related to the economic design of control charts for quality control purposes which have their focus on sequentially detecting a change as quickly as possible after it occurs (see, for example Champ and Woodall, 1987; Woodall

and Montgomery, 1999). However, the focus of our approach here is on testing for the presence and identification of the time of a change in a retrospective scenario. Despite of its importance - to the best of our knowledge - not much work has been done in this direction. The paper which is most similar in spirit to the problem considered here is the work by Dette and Wu (2019) who define a change in the mean of a time series from its initial value as relevant if the amount of the change and the time period where the change is in place are reasonably large. More precisely, these authors propose to test the hypothesis that the Lebesgue measure of the set $\mathcal{M}_\Delta = \{t \in [0,1] \colon |\mu(t) - \mu(0)| > \Delta\}$ is smaller than a given value c, that is

$$\tilde{H}_0 : \lambda(\mathcal{M}_\Delta) \leq c \quad \text{vs.} \quad \tilde{H}_1 : \lambda(\mathcal{M}_\Delta) > c.$$

In contrast to this, the formulation of the hypotheses (1.2) and (1.3) defines a change as relevant whenever the difference between the mean function and the benchmark exceeds the threshold Δ. The latter hypotheses are easier to interpret for practitioners, however, various challenging mathematical problems arise from this formulation. While the Lebesgue measure of the set $\mathcal{M}_\Delta$ can be estimated by a mass excess approach and the corresponding statistic has an asymptotic normal distribution, such a simple limit distribution does not appear if one investigates the maximal deviation of the function μ from a benchmark as formulated in the hypotheses (1.2) or (1.3). More precisely, for the construction of a test for these hypotheses we propose to estimate the maximal deviation d_∞ directly and to reject the null hypothesis for large values of its estimate. In order to quantify the type I and type II error, we investigate the asymptotic properties of the estimator, for any value $d_\infty \geq 0$, which depend sensitively on specific properties of the function μ.

In Section 2, we introduce a slightly more general version of the testing problems in (1.2) and (1.3) and propose an estimator for the respective maximum deviation. Several technical assumptions required for the asymptotic analysis are collected as well. In Section 3, we show that the estimator (after appropriate standardization) converges weakly to a non-degenerate limit distribution of the Gumbel type. The limit distribution as well as the quantities used for standardization depend on the set of extremal points of the difference between the function μ and its benchmark, i.e., the set of points where the difference attains its sup-norm (see equation (2.13) for a precise definition). The results in this section may already be used to construct a simple consistent and asymptotic level α-test for the hypotheses (1.2) or (1.3), but the test tends to be conservative if the Lebesgue measure of the set of extremal points is small. In Section 4, as a circumvention for the latter, we propose suitable estimators for the set of extremal points, prove that these are consistent and use them to develop tests with a better approximation of the nominal level. In Section 5, we consider the estimation of the first point in time where the relevant change occurs. Section 6 is then devoted to the detection of relevant change points in the signal to noise ratio of a location scale model with a non-stationary error process, while the finite sample properties of the proposed methodologies are investigated by means of a Monte Carlo simulation study in Section 7. Finally, all proofs and technical details are deferred to an appendix.

Throughout this paper, the symbol $\rightsquigarrow$ denotes weak convergence, and all convergences are for $n \to \infty$ if not mentioned otherwise.

2. The general testing problem and mathematical preliminaries

We consider the common location scale model

$$X_{i,n} = \mu(i/n) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{2.1}$$

where $(\varepsilon_i)_{i \in \mathbb{Z}}$ denotes a stationary sequence of centered random variables and $\mu : [0, 1] \to \mathbb{R}$ is the unknown mean function. We are interested in detecting significant deviations of the regression function μ on an interval $[x_0, x_1] \subseteq [0, 1]$ from a given benchmark. For this purpose, we consider a real-valued functional g on the space of all functions defined on the interval $[0, 1]$ and define the distance

$$d_\infty := \sup_{t \in [x_0, x_1]} |\mu(t) - g(\mu)| \tag{2.2}$$

Note that this distance depends on the points $x_0, x_1 \in [0, 1]$ in the calculation of the supremum and on the functional g used to define the benchmark, which is not reflected by our notation (as it will always be clear from the context). We are interested in the hypotheses

$$H_0 : d_\infty \leq \Delta \quad \text{vs.} \quad H_1 : d_\infty > \Delta, \tag{2.3}$$

where $\Delta > 0$ is a given constant. In particular, we obtain
- the hypotheses (1.2) for the choice $g(\mu) = \mu(0)$, $x_0 = 0$ and $x_1 = 1$.
- the hypotheses (1.3) for the choice $g(\mu) = (1/x_0) \int_0^{x_0} \mu(t)dt$ and $x_1 = 1$.
- the hypotheses of a relevant deviation from an overall mean by choosing $g(\mu) = \bar{\mu} = \int_0^1 \mu(t)dt$, $x_0 = 0$ and $x_1 = 1$, that is

$$H_0 : \sup_{t \in [0,1]} |\mu(t) - \bar{\mu}| \leq \Delta \quad \text{vs.} \quad H_1 : \sup_{t \in [0,1]} |\mu(t) - \bar{\mu}| > \Delta. \tag{2.4}$$

In order to estimate the maximum deviation d_∞ defined in (2.2) we use local linear regression. To be precise, let K denote a kernel function (see Assumption 2.2 below for details) and define $K_h(\cdot) = K(\frac{\cdot}{h})$, for $h > 0$. The *local linear estimator* $\hat{\mu}_{h_n}$ with positive bandwidth $h_n = o(1)$ as $n \to \infty$ is defined by the first coordinate of the minimizer

$$(\hat{\mu}_{h_n}(t), \widehat{\mu'}_{h_n}(t)) = \underset{b_0, b_1 \in \mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^n \left\{ X_i - b_0 - b_1(\tfrac{i}{n} - t) \right\}^2 K_{h_n}(\tfrac{i}{n} - t), \tag{2.5}$$

see, for example, Fan and Gijbels (1996). Throughout, we will use a bias corrected version of the local linear estimator adopting the Jackknife bias reduction technique proposed by Schucany and Sommers (1977) which is defined by

$$\tilde{\mu}_{h_n}(t) = 2\hat{\mu}_{h_n/\sqrt{2}}(t) - \hat{\mu}_{h_n}(t). \tag{2.6}$$

A simple estimator for the benchmark $g(\mu)$ is now obtained by $\hat{g}_n = g(\tilde{\mu}_{h_n})$, but we note that other estimators can be used as well for this purpose. For example, if $g(\mu) = \int_0^1 \mu(t)dt$, the sample mean $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_{i,n}$ could be used as an alternative (and simpler) estimator for $g(\mu)$.

If $\hat{g}_n$ is an appropriate estimator for the benchmark $g(\mu)$, a natural estimator for the maximum deviation d_∞ is given by

$$\hat{d}_{\infty,n} = \sup_{t \in I_n} |\tilde{\mu}_{h_n}(t) - \hat{g}_n|, \tag{2.7}$$

where we restrict the supremum to the interval $I_n = [x_0 \vee h_n, x_1 \wedge (1 - h_n)]$ to account for boundary effects of $\tilde{\mu}_{h_n}$. The null hypothesis in (2.3) is rejected for large values of the statistic $\hat{d}_{\infty,n}$. In order to define suitable critical values for this test we will investigate the weak convergence of an affinely standardized version of the statistic $\hat{d}_{\infty,n}$ within the next section.

Remark 2.1. Note that the related problem of investigating the asymptotic properties of the sup-norm of the difference between a nonparametric estimator $\hat{\mu}$ and the true regression function μ (or $\mathbb{E}[\hat{\mu}]$), i.e., $\|\hat{\mu} - \mu\|_{\infty}$ (or $\|\hat{\mu} - \mathbb{E}[\hat{\mu}]\|_{\infty}$), have been investigated by several authors, mostly in the case of independent observations (see Johnston, 1982; Xia, 1998; Proksch, 2016, and the references therein), but also for stationary data (see Wu and Zhao, 2007; Zhao and Wu, 2008, among others). The most prominent statistical application of results of this type concerns the construction of simultaneous asymptotic confidence bands for the mean function. Using the duality between confidence regions and hypotheses testing, these confidence bands yield a simple test for the hypotheses in (2.3). In this remark we briefly discuss this approach and explain why it does not yield powerful tests in general.

Our starting point is a result of the above type in the case of a (possibly locally) stationary error process, which is in fact a by-product of this work derived in Section A below. In that section, we derive the asymptotic distribution of

$$\sup_{t \in I_n} |\tilde{\mu}_{h_n}(t) - \hat{g}_n - \{\mu(t) - g(\mu)\}|$$

(note that our assumptions allow for $\hat{g}_n = g = 0$), which may then be used to construct constants $c = c_{n,\alpha}$ not depending on unknown objects such that

$$I_n(t) = [\tilde{\mu}_{h_n}(t) - \hat{g}_n - c_{n,\alpha}, \tilde{\mu}_{h_n}(t) - \hat{g}_n + c_{n,\alpha}], \qquad t \in I_n, \tag{2.8}$$

defines an asymptotic simultaneous $(1 - \alpha)$ confidence band for the function $\mu - g(\mu)$, that is,

$$\lim_{n \to \infty} \mathbb{P}\big(\mu(t) - g(\mu) \in I_n(t) \ \forall \, t \in I_n\big)$$
$$= \lim_{n \to \infty} \mathbb{P}(\sup_{t \in I_n} |\tilde{\mu}_{h_n}(t) - \hat{g}_n - \{\mu(t) - g(\mu)\}| \le c_{n,\alpha}) = 1 - \alpha$$

(see Remark A.4 for more details, and (3.8) for a precise definition of $c_{n,\alpha}$).

Now using a well-known general relation between confidence intervals and statistical tests (see, e.g., Aitchison, 1964), we may use (2.8) to construct a simple asymptotic level α-test. Indeed, some thoughts reveal that accepting the null hypotheses in (2.3) whenever both the constant function Δ and the constant function $-\Delta$ is completely contained in the confidence band in (2.8) is equivalent to rejecting whenever

$$\hat{d}_{\infty,n} > \Delta + c_{n,\alpha}. \tag{2.9}$$

Next, under H_0 in (2.3), we have $d_{\infty,n} = \sup_{t \in I_n} |\mu(t) - g(\mu)| \le d_{\infty} \le \Delta$ and (2.9) implies

$$\sup_{t \in I_n} |\tilde{\mu}_{h_n}(t) - \hat{g}_n - \{\mu(t) - g(\mu)\}| \ge \hat{d}_{\infty,n} - d_{\infty,n} > c_{n,\alpha} + \Delta - \Delta = c_{n,\alpha},$$

which gives

$$\limsup_{n\to\infty} \mathbb{P}(H_0 \text{ is rejected})$$

$$\leq \limsup_{n\to\infty} \mathbb{P}\Big(\sup_{t\in I_n} |\tilde{\mu}_{h_n}(t) - \hat{g}_n - \{\mu(t) - g(\mu)\}| > c_{n,\alpha} \Big) = \alpha. \quad (2.10)$$

However, due to the first inequality in (2.10), the test in (2.9) is conservative and has very low power for testing the hypotheses in (2.3), except in the case of classical hypotheses, that is $\Delta = 0$. In fact, in the latter case, we have equality in (2.10) due to the fact the null hypothesis is rather simple, only containing the constant regression functions. If $\Delta > 0$ however, even the boundary of the null hypothesis (i.e., those regression functions μ with $d_\infty = \Delta$) is a far more complicated set, rendering the statistical problem substantially more difficult.

In light of the previous remark, it is the main goal of this work to derive alternative critical values for a test that is based on rejecting the null hypothesis for large values of $\hat{d}_{\infty,n}$. While those values will result in a substantially better approximation of the nominal level and better power properties, their derivation will also be more complicated compared to the above construction. More precisely, in Section 3 and 4, we will derive sequences $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ such that $a_n(\hat{d}_{\infty,n} - d_\infty) - b_n$ converges in distribution with a non-degenerate limit, and it will turn out that both the standardizing sequences and the limit distribution depend sensitively on the function $d = d(t) = \mu(t) - g(\mu)$, even on the boundary of the null hypothesis.

In the remaining parts of this section, we collect regularity assumptions that are sufficient to derive the intended limit results. We begin with a standard assumption regarding the kernel K used in the local linear estimator.

Assumption 2.2. The kernel K is symmetric, supported on the interval $[-1, 1]$, twice differentiable and satisfies $\int_{[-1,1]} K(x)\, dx = 1$.

Next we define the dependence structure in model (2.1). For this purpose we recall some basic definitions on physical dependence measures of stationary processes (Wu, 2005). For $q \geq 1$, let $\|X\|_{q,\Omega} = (\mathbb{E}|X|^q)^{1/q}$ denote the $\mathcal{L}_q$-norm of a random variable X defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Let $\eta = (\eta_i)_{i\in\mathbb{Z}}$ be a sequence of independent identically distributed random variables and let $(\eta') = (\eta'_i)_{i\in\mathbb{Z}}$ be an independent copy of η. Further, define $\mathcal{F}_i = (\ldots, \eta_{-2}, \eta_{-1}, \eta_0, \eta_1, \ldots, \eta_i)$ and $\mathcal{F}_i^* = (\ldots, \eta_{-2}, \eta_{-1}, \eta'_0, \eta_1, \ldots, \eta_i)$. Let $G : \mathbb{R}^\mathbb{N} \to \mathbb{R}$ denote a possibly nonlinear filter such that $\varepsilon_i = G(\mathcal{F}_i)$ and $\varepsilon_i^* = G(\mathcal{F}_i^*)$ are properly defined random variables.

The physical dependence measure of a filter G with $\|G(\mathcal{F}_0)\|_{q,\Omega} < \infty$ with respect to the norm $\|\cdot\|_{q,\Omega}$ is defined by

$$\delta_q(G, i) = \|G(\mathcal{F}_i) - G(\mathcal{F}_i^*)\|_{q,\Omega}, \qquad i \in \mathbb{N}. \quad (2.11)$$

The quantity $\delta_q(G, i)$ can be regarded as a measure for the serial dependence at lag i of $(\varepsilon_j)_{j\in\mathbb{N}}$. It plays a similar role as a mixing coefficient, yet it is easier to bound in many cases.

Assumption 2.3. The error process $(\varepsilon_i)_{i\in\mathbb{Z}}$ in model (2.1) is centered and has a representation $\varepsilon_i = G(\mathcal{F}_i)$ with a filter G such that the following conditions are satisfied:

(i) There exists $\chi \in (0, 1)$ such that $\delta_4(G, i) = \mathcal{O}(\chi^i)$, as $i \to \infty$.

(ii) The *long-run variance* of $(\varepsilon_j)_{j\in\mathbb{N}}$, defined as

$$\sigma^2 = \sum_{i=-\infty}^{\infty} \operatorname{Cov}\big(G(\mathcal{F}_i), G(\mathcal{F}_0)\big), \tag{2.12}$$

exists and is positive.

We also need to impose a certain degree of regularity on the function $d(t) = \mu(t) - g(\mu)$. For that purpose, let $\|f\|_\infty = \sup_{t\in[x_0,x_1]} |f(t)|$ denote the sup-norm of a function f on the interval $[x_0, x_1]$. Note that $d_\infty = \|d\|_\infty$ and that we again do not reflect the dependence on x_0 and x_1 in our notation. The points where $|d|$ attains its sup-norm are called extremal points and the corresponding set of extremal points is denoted by $\mathcal{E}$. Note that we have

$$\mathcal{E} = \mathcal{E}^+ \cup \mathcal{E}^-, \tag{2.13}$$

where

$$\mathcal{E}^+ = \{t \in [x_0, x_1] : d(t) = \|d\|_\infty\}, \quad \mathcal{E}^- = \{t \in [x_0, x_1] : d(t) = -\|d\|_\infty\}.$$

These sets depend on the function d and on x_0, x_1, which is not reflected in our notation as it will always be clear from the context. If the function d is continuous, the sets $\mathcal{E}, \mathcal{E}^+$ and $\mathcal{E}^-$ are compact. Moreover, unless $\|d\|_\infty = 0$, $\mathcal{E}^+$ and $\mathcal{E}^-$ are disjoint.

Assumption 2.4.
 (i) The function μ is twice differentiable with Lipschitz continuous second derivative.
 (ii) There exists a constant $\gamma > 0$ such that $d = \mu - g(\mu)$ is concave (convex) on $U_\gamma(t) := \{s \in [x_0, x_1] : |s - t| < \gamma\}$, for any $t \in \mathcal{E}^+$ ($t \in \mathcal{E}^-$).

Remark 2.5. Assumption 2.4(ii) is made to avoid functions with an irregular behaviour at the extremal points. In this remark we give some more explanation for it.
 (i) If the function d is continuously differentiable, Assumption 2.4(ii) is satisfied provided its derivative d' is decreasing on $U_\gamma(t)$, for $t \in \mathcal{E}^+$, and increasing on $U_\gamma(t)$, for $t \in \mathcal{E}^-$. In particular, if d is twice differentiable with Lipschitz continuous second derivative, and if $\varphi = \inf_{t\in\mathcal{E}} |d''(t)| > 0$, the constant γ can be chosen as $\varphi/(2L_2)$, where L_2 is the Lipschitz constant of d''. In this case it follows that $\lambda(\mathcal{E}) = 0$, where λ denotes the Lebesgue measure.
 (ii) For any interval $[t_1, t_2] \subset \mathcal{E}^+$ with $t_1, t_2 \in \partial\mathcal{E}^+$, by concavity, d is strictly increasing on the interval $(t_1 - \gamma, t_1]$ and strictly decreasing on $[t_2, t_2 + \gamma)$. Consequently, for any $s \in (t_1 - \gamma, t_1) \cup (t_2, t_2 + \gamma)$, it follows that $s \notin \mathcal{E}^+$. Analogously $s \in (t_1 - \gamma, t_1) \cup (t_2, t_2 + \gamma)$ implies that $s \notin \mathcal{E}^-$, for any interval $[t_1, t_2] \subset \mathcal{E}^-$ with $t_1, t_2 \in \partial\mathcal{E}^-$. Hence, it follows from Assumption 2.4(ii), that $\mathcal{E}$ is a finite union of at most $\lfloor (2\gamma)^{-1} \rfloor$ intervals and single points.

Assumption 2.6. The estimator $\hat{g}_n$ of the functional $g(\mu)$ satisfies

$$|\hat{g}_n - g(\mu)| = o_\mathbb{P}\Big(\frac{1}{\sqrt{nh_n |\log(h_n)|}} \Big), \qquad n \to \infty,$$

where h_n denotes the bandwidth parameter used for the estimator in (2.5).

Remark 2.7. Assumption 2.6 is not very strong and satisfied for many common estimators for $g(\mu)$. Exemplary, we consider the situations in (1.2), (1.3) and (2.4).

(i) If Assumptions 2.2, 2.3 and 2.4(i) are met, it follows from Lemma C.2 in the supplementary material of Dette and Wu (2019) that

$$|\tilde{\mu}_{\tilde{h}_n}(0) - \mu(0)| = \mathcal{O}_{\mathbb{P}}\Big(\frac{1}{\sqrt{n\tilde{h}_n}}\Big) + \mathcal{O}\Big(\tilde{h}_n^3 + \frac{1}{n\tilde{h}_n}\Big),$$

for any bandwidth $\tilde{h}_n > 0$ with $\tilde{h}_n \to 0$ and $n\tilde{h}_n \to \infty$. Thus, Assumption 2.6 holds for $g(\mu) = \mu(0)$ with $\hat{g}_n = \tilde{\mu}_{h_n|\log(h_n)|^2}(0)$, provided that $nh_n^7|\log h_n|^{12} = o(1)$.

(ii) Similarly, if $g(\mu)$ is an integral as considered in (1.3) and (2.4), define $\hat{g}_n = \bar{X}_n(x_0) = \frac{1}{\lfloor x_0 n \rfloor} \sum_{i=1}^{\lfloor x_0 n \rfloor} X_{i,n}$. Then,

$$\left| \bar{X}_n(x_0) - \frac{1}{x_0} \int_0^{x_0} \mu(t)\,\mathrm{d}t \right|$$

$$= \left| \frac{1}{\lfloor x_0 n \rfloor} \sum_{i=1}^{\lfloor x_0 n \rfloor} \varepsilon_i + \frac{1}{\lfloor x_0 n \rfloor} \sum_{i=1}^{\lfloor x_0 n \rfloor} \mu\big(\tfrac{i}{n}\big) - \frac{1}{x_0} \int_0^{x_0} \mu(t)\,\mathrm{d}t \right| = \mathcal{O}_{\mathbb{P}}(n^{-1/2})$$

and Assumption 2.6 holds for $g(\mu) = \frac{1}{x_0} \int_0^{x_0} \mu(t)\,\mathrm{d}t$, provided $h_n = o(1)$ as $n \to \infty$.

3. Weak convergence of $\hat{d}_{\infty,n}$ and a first simple test

In this section we will derive the limit distribution of the statistic $\hat{d}_{\infty,n}$ (after appropriate standardization), see Theorem 3.1. The result may be used to construct a simple test for the hypotheses in (2.3), for any fixed $\Delta \geq 0$. However, depending on the data-generating process, the test may be quite conservative whence we continue in Section 4 by proposing a less conservative test based on estimating the set of extremal points $\mathcal{E}$ defined in (2.13).

The convergence rate of the statistic $\hat{d}_{\infty,n}$ as defined in (2.7) depends crucially on the scaling sequence ℓ_n defined as

$$\ell_n = \sqrt{2\log\big(\tfrac{\Lambda_K \lambda(I_n)}{2\pi h_n}\big)} \sim \sqrt{2 \cdot |\log(h_n)|}, \tag{3.1}$$

where $I_n = [x_0 \vee h_n, x_1 \wedge (1 - h_n)]$ and the constant Λ_K is defined as

$$\Lambda_K = \frac{\|(K^*)'\|_2}{\|K^*\|_2} \tag{3.2}$$

($\|f\|_2$ denotes the L^2-norm of some function f), with the function K^* given by

$$K^*(x) = 2\sqrt{2}K(\sqrt{2}x) - K(x). \tag{3.3}$$

Next, recall the Gumbel distribution Gum_a with location parameter $a \in \mathbb{R}$, defined through its c.d.f.

$$\mathrm{Gum}_a((-\infty, x]) = \exp[-\exp\{-(x - a)\}], \quad x \in \mathbb{R}.$$

Finally, recall that $\hat{d}_{\infty,n} = \sup_{t \in I_n} |\tilde{\mu}_{h_n}(t) - \hat{g}_n|$ and let $d_{\infty,n} = \sup_{t \in I_n} |\mu(t) - g(\mu)|$.

Theorem 3.1. *Suppose that Assumptions 2.2, 2.3, 2.4(i) and 2.6 hold. Further, assume that the bandwidth satisfies*

$$h_n \to 0, \quad nh_n \to \infty, \quad nh_n^7 |\log(h_n)| \to 0, \quad \limsup_{n\to\infty} \frac{|\log(h_n)| \log^4 n}{n^{1/2} h_n} < \infty.$$

(1) If $\lambda(\mathcal{E}) > 0$ and $d_\infty > 0$, then

$$\frac{\ell_n \sqrt{nh_n}}{\sigma \|K^*\|_2} (\hat{d}_{\infty,n} - d_{\infty,n}) - \ell_n^2 \rightsquigarrow \mathrm{Gum}_{\log\{\lambda(\mathcal{E})/(x_1-x_0)\}}.$$

(2) If $d_\infty = 0$ (which implies $\lambda(\mathcal{E}) = x_1 - x_0$), then

$$\frac{\ell_n \sqrt{nh_n}}{\sigma \|K^*\|_2} (\hat{d}_{\infty,n} - d_{\infty,n}) - \ell_n^2 \rightsquigarrow \mathrm{Gum}_{\log(2)}.$$

(3) If $\lambda(\mathcal{E}) = 0$ and if additionally Assumption 2.4(ii) is met, then

$$\frac{\ell_n \sqrt{nh_n}}{\sigma \|K^*\|_2} (\hat{d}_{\infty,n} - d_{\infty,n}) - \ell_n^2 \rightsquigarrow -\infty.$$

Note that $-\infty \ll_S \mathrm{Gum}_{\log\{\lambda(\mathcal{E})/(x_1-x_0)\}} \ll_S \mathrm{Gum}_0 \ll_S \mathrm{Gum}_{\log(2)}$, where $\ll_S$ denotes the (first order) stochastic dominance ordering. Subsequently, this will be used to construct suitable critical values for a first simple test of (2.3). Further note that the results of the theorem continue to hold if ℓ_n is replaced by

$$\ell_n' = \sqrt{2 \log\left(\frac{\Lambda_K(x_1-x_0)}{2\pi h_n}\right)}. \tag{3.4}$$

In general, applications of Theorem 3.1 to the construction of inferential methodology require an estimator for the long-run variance σ^2. For this purpose, we may for instance follow Wu and Zhao (2007): define the partial sums $S_{j,k} = \sum_{i=j}^{k} X_{i,n}$ and let

$$\hat{\sigma}^2 = \frac{1}{\lfloor n/m_n \rfloor - 1} \sum_{j=1}^{\lfloor n/m_n \rfloor - 1} \frac{(S_{(j-1)m_n+1,jm_n} - S_{jm_n+1,(j+1)m_n})^2}{2m_n}, \tag{3.5}$$

where m_n denotes some integer sequence proportional to $n^{1/3}$. If Assumptions 2.2, 2.3 and 2.4(i) are satisfied, it follows from Theorem 3 in the last-named reference that

$$\hat{\sigma}^2 = \sigma^2 + \mathcal{O}_{\mathbb{P}}(n^{-1/3}). \tag{3.6}$$

Together with the latter result, Theorem 3.1 allows for the construction of a first simple test for the hypotheses in (2.3). The cases $\Delta = 0$ and $\Delta > 0$ need to be treated separately. First, in the case $\Delta = 0$, the only possible limiting distribution under the null is the Gumbel distribution $\mathrm{Gum}_{\log(2)}$, whence we propose to reject the null hypothesis

$$H_0 : d_\infty = 0 \quad \text{vs.} \quad H_1 : d_\infty > 0 \tag{3.7}$$

whenever

$$\hat{d}_{\infty,n} > (q_{\log(2),1-\alpha} + \ell_n^2) \frac{\hat{\sigma} \|K^*\|_2}{\sqrt{nh_n}\ell_n} =: c_{n,\alpha}, \tag{3.8}$$

where $q_{a,\beta}$ denotes the β-quantile of the Gumbel distribution Gum_a with location parameter a. In the case $\Delta > 0$, all possible limiting distributions on the boundary of the null hypothesis (i.e., $d_\infty = \Delta$) are stochastically dominated by the Gumbel

distribution Gum_0 (which is in fact attained for models with $\lambda(\mathcal{E}) = x_1 - x_0$). As a consequence, we propose to reject the null hypothesis

$$H_0 : d_\infty \leq \Delta \quad \text{vs.} \quad H_1 : d_\infty > \Delta, \tag{3.9}$$

whenever

$$\hat{d}_{\infty,n} > (q_{0,1-\alpha} + \ell_n^2)\frac{\hat{\sigma}\|K^*\|_2}{\sqrt{nh_n\ell_n}} + \Delta. \tag{3.10}$$

Note that this decision rule has a similar structure as the test (2.9) derived from the simultaneous confidence band for d, whose critical value are based on using $q_{\log 2, 1-\alpha}$ instead of $q_{0,1-\alpha}$; see equation (A.16) for the definition of $c_{n,\alpha}$ in (2.9). Consequently, as $q_{0,1-\alpha} < q_{\log 2, 1-\alpha}$ the test (3.10) is more powerful than the test (2.9).

Corollary 3.2. *Let Assumptions 2.2, 2.3, 2.4 and 2.6 be met.*
(1) The decision rule (3.8) defines a consistent and asymptotic level α-test for the the hypotheses in (3.7).
(2) The decision rule (3.10) defines a consistent and asymptotic level α-test for the the hypotheses in (3.9), where $\Delta > 0$.

It follows from the proof of Corollary 3.2 in Section A.2 that the test defined by (3.10) for the relevant hypotheses (3.9) is conservative for those models on the boundary of the null hypothesis for which $\lambda(\mathcal{E})$ is strictly smaller than $x_1 - x_0$. A heuristic argument for this fact is that the test does not use the quantiles of the $\text{Gum}_{\log\{\lambda(\mathcal{E})/(x_1-x_0)\}}$-distribution in Theorem 3.1(1), but only the quantiles of the (stochastically dominating) distribution Gum_0. In the following section we will address this problem by estimating the unknown set $\mathcal{E}$ and develop a test which does not have this drawback.

4. An improved test based on estimation of the set of extremal points

The test in (3.10) for the hypotheses in (3.9) suffers from the fact that the models on the boundary of the null hypothesis may have different limiting distributions, whence only the stochastically largest may be used to construct valid critical values. A workaround proposed in this section consists of estimating the Lebesgue measure $\lambda(\mathcal{E})$ of the set of extremal points first and then using a slightly different standardization for $\hat{d}_{\infty,n}$, resulting in the same limit distribution for all models with $d_\infty = \Delta > 0$ and $\lambda(\mathcal{E}) > 0$. Moreover, we will provide an approximation result in terms of i.i.d. Gaussian variables, which may be used as yet another alternative to obtain valid critical values that are non-conservative for models on the boundary of the null hypothesis.

For the estimation of $\lambda(\mathcal{E})$, recall the definition of $I_n = [x_0 \vee h_n, x_1 \wedge (1 - h_n)]$ and define, for some positive sequence $(\rho_n)_{n\in\mathbb{N}} = o(1)$, the set $\mathcal{E}_n = \mathcal{E}_n^+ \cup \mathcal{E}_n^-$, where

$$\mathcal{E}_n^\pm = \left\{t \in I_n : d_{\infty,n} \mp d(t) \leq \rho_n\right\}. \tag{4.1}$$

Replacing $d(t) = \mu(t) - g(\mu)$ by $\hat{d}_n(t) = \tilde{\mu}_{h_n}(t) - \hat{g}$, we obtain the set-valued estimator $\hat{\mathcal{E}}_n = \hat{\mathcal{E}}_n^+ \cup \hat{\mathcal{E}}_n^-$, where

$$\hat{\mathcal{E}}_n^\pm = \left\{t \in I_n : \hat{d}_{\infty,n} \mp \hat{d}_n(t) \leq \rho_n\right\}.$$

This estimator is consistent in the following sense.

Theorem 4.1. *Let $\rho_n = o(1)$ be a positive sequence such that $\rho_n^2 nh_n/|\log(h_n)| \to \infty$. Further, suppose that Assumptions 2.2, 2.3, 2.4(i) and 2.6 are satisfied. If $\lambda(\mathcal{E}) = 0$, let additionally Assumption 2.4(ii) and $\rho_n h_n^{-2} \to \infty$ be satisfied. Then,*

$$\frac{\lambda(\mathcal{E}_n)}{\lambda(\hat{\mathcal{E}}_n)} \xrightarrow{\mathbb{P}} 1.$$

Next, we present a variation of Theorem 3.1 which is based on the alternative scaling sequence $\ell_n(\mathcal{E}_n)$ (instead of $\ell_n = \ell_n([x_0, x_1])$ as used in Theorem 3.1), where

$$\ell_n(\mathcal{A}) = \sqrt{2\log\left(\frac{\Lambda_K \lambda(\mathcal{A} \cap I_n)}{2\pi h_n}\right)}, \qquad \mathcal{A} \in \mathcal{B}([x_0, x_1]) \tag{4.2}$$

and $\mathcal{B}([x_0, x_1])$ denote the Borel subsets of $[x_0, x_1]$.

Theorem 4.2. *Suppose that Assumptions 2.2, 2.3, 2.4 and 2.6 are satisfied. If $\lambda(\mathcal{E}) = 0$, let additionally $\rho_n^{1/2} h_n^{-1} \to \infty$. Let $(V_i)_{i \in \mathbb{N}}$ denote an i.i.d. sequence of standard normally distributed random variables and define*

$$G_{n,1} = \frac{\ell_n(\mathcal{E}_n)\sqrt{nh_n}}{\|K^*\|_2} \sup_{t \in \mathcal{E}_n} \left\{ \frac{1}{nh_n} \sum_{i=1}^n V_i K_{h_n}^*\left(\frac{i}{n} - t\right) \right\} - \ell_n^2(\mathcal{E}_n), \tag{4.3}$$

$$G_{n,2} = \frac{\ell_n(\mathcal{E}_n)\sqrt{nh_n}}{\|K^*\|_2} \sup_{t \in \mathcal{E}_n} \left| \frac{1}{nh_n} \sum_{i=1}^n V_i K_{h_n}^*\left(\frac{i}{n} - t\right) \right| - \ell_n^2(\mathcal{E}_n). \tag{4.4}$$

If $d_\infty > 0$, then

$$\mathbb{P}\left(\frac{\ell_n(\mathcal{E}_n)\sqrt{nh_n}}{\sigma\|K^*\|_2}(\hat{d}_{\infty,n} - d_{\infty,n}) - \ell_n^2(\mathcal{E}_n) \le x \right) \ge \mathbb{P}(G_{n,1} \le x) + o(1),$$

with equality if $\lambda(\mathcal{E}) > 0$. If $d_\infty = 0$, then

$$\mathbb{P}\left(\frac{\ell_n(\mathcal{E}_n)\sqrt{nh_n}}{\sigma\|K^*\|_2}(\hat{d}_{\infty,n} - d_{\infty,n}) - \ell_n^2(\mathcal{E}_n) \le x \right) = \mathbb{P}(G_{n,2} \le x) + o(1).$$

Moreover,

$$G_{n,1} \rightsquigarrow \mathrm{Gum}_0, \qquad G_{n,2} \rightsquigarrow \mathrm{Gum}_{\log(2)}$$

We can now use the previous two theorems to define consistent asymptotic level α tests for the hypotheses (3.7) and (3.9), based on a rescaling of the statistic $\hat{d}_{\infty,n}$ which makes use of the estimator $\hat{\mathcal{E}}_n$. To be precise, we propose to reject the null hypotheses in (3.7) and (3.9) if

$$\sup_{t \in I_n} |\hat{d}_n(t)| > \{q_{a,1-\alpha} + \ell_n^2(\hat{\mathcal{E}}_n)\} \frac{\hat{\sigma}\|K^*\|_2}{\sqrt{nh_n}\ell_n(\hat{\mathcal{E}}_n)} + \Delta, \tag{4.5}$$

where $\hat{\sigma}^2$ denotes the estimator of the long-run variance defined in (3.5) and $q_{a,1-\alpha}$ denotes the $(1-\alpha)$-quantile of the Gumbel distribution Gum_a with location parameter $a = 0$ if $\Delta > 0$ and $a = \log(2)$ if $\Delta = 0$.

Corollary 4.3. *Under the assumptions of Theorem 4.2, the test (4.5) is consistent and has asymptotic level α for the hypotheses in (3.7) and (3.9).*

From Theorem 4.1, Theorem 4.2 and (3.6), we also obtain that, if $d_\infty > 0$,

$$\mathbb{P}\left(\frac{\ell_n(\hat{\mathcal{E}}_n)\sqrt{nh_n}}{\hat{\sigma}\|K^*\|_2}(\hat{d}_{\infty,n} - d_{\infty,n}) - \ell_n^2(\hat{\mathcal{E}}_n) \leq x\right) \geq \mathbb{P}(G_{n,1} \leq x) + o(1),$$

with $G_{n,1}$ as defined in (4.3). Hence, while the test in (4.5) (for the case $\Delta > 0$) was based on using the quantiles $q_{0,1-\alpha}$ of the limiting distribution Gum_0 of $G_{n,1}$, we may alternatively use the quantiles $q_{\hat{G}_{n,1},1-\alpha}$ of the distribution of $\hat{G}_{n,1}$ (or $q_{\hat{G}_{n,2},1-\alpha}$ in case $\Delta = 0$), where $\hat{G}_{n,j}$ is defined analogously to $G_{n,j}$ in (4.3) and (4.4), but with $\mathcal{E}_n$ replaced by $\hat{\mathcal{E}}_n$. Note that these quantiles may easily be simulated up to an arbitrary precision. More precisely, we propose to reject the null hypotheses in (3.7) and (3.9) if

$$\sup_{t\in I_n} |\hat{d}_n(t)| > \{q_{\hat{G}_{n,j},1-\alpha} + \ell_n^2(\hat{\mathcal{E}}_n)\}\frac{\hat{\sigma}\|K^*\|_2}{\sqrt{nh_n}\ell_n(\hat{\mathcal{E}}_n)} + \Delta, \tag{4.6}$$

where $j = 1$ if $\Delta > 0$ and $j = 2$ if $\Delta = 0$. It can be shown by similar arguments as given in the proof of Corollary 4.3 that this test is consistent and has asymptotic level α. The numerical results in Section 7 suggest that the test in (4.6) exhibits better finite sample properties than the test in (4.5). This phenomenon may be heuristically explained by the fact that the convergence rate of a maximum of independent normal random variables to the Gumbel distribution is rather slow.

5. Estimating the time of the first relevant deviation

The aim of this section is to develop an estimator for the first relevant deviation

$$t^* = \inf\{t \in [x_0, x_1] : |d(t)| \geq \Delta\},$$

where we use the convention that $\inf(\emptyset) = +\infty$. First note that by continuity of d the point t^* can be represented as

$$t^* = x_0 + \int_{x_0}^{x_1} \mathbb{1}\left(\max_{t\in[x_0,s]}|d(t)| < \Delta\right)ds + \infty \cdot \mathbb{1}\left(d_\infty < \Delta\right).$$

Obviously, the properties of any estimator will depend on the smoothness of the function d at the point t^*. To capture the degree of smoothness, assume that there exist constants $\kappa > 0$ and $c_\kappa > 0$ such that

$$\lim_{s\uparrow t^*}\frac{|d(t^*) - d(s)|}{(t^* - s)^\kappa} = c_\kappa. \tag{5.1}$$

Note that $\kappa = 1$ if the function d is differentiable at the point t^* with non-vanishing derivative.

Theorem 5.1. *Let Assumptions 2.2, 2.3, 2.4 and (5.1) be satisfied and let δ_n denote a positive sequence such that $\delta_n \to 0$ and $\liminf_{n\to\infty}\frac{\sqrt{nh_n}\delta_n}{\sigma\|K^*\|_2} - \ell_n > 0$. If $t^* \in [x_0, x_1]$, then the estimator*

$$\hat{t}^* = (h_n \vee x_0) + \int_{I_n} \mathbb{1}\left(\max_{t\in[h_n\vee x_0,s]}|\hat{d}_n(t)| < \Delta - \delta_n\right)ds + \infty \cdot \mathbb{1}\left(\hat{d}_{\infty,n} < \Delta - \delta_n\right)$$

satisfies

$$\hat{t}^* = t^* + \mathcal{O}_\mathbb{P}\left(\left(\frac{|\log(h_n)|^{1/2}}{\sqrt{nh_n}} + \delta_n\right)^{1/\kappa} \vee h_n\right) = t^* + o_\mathbb{P}(1).$$

If $t^* = \infty$, then $\mathbb{P}(\hat{t}^* < \infty) = o_\mathbb{P}(1)$.

6. EXTENSION TO NON-STATIONARY ERROR PROCESSES

In this section, we extend the theory developed in the previous sections to the model

$$X_{i,n} = \mu(i/n) + \varepsilon_{i,n}, \qquad i = 1, \dots, n, \tag{6.1}$$

with a triangular array of centered but possibly non-stationary (e.g., heteroscedastic) errors $\{\varepsilon_{i,n}|1 \le i \le n\}_{n\in\mathbb{N}}$. For this purpose, recall the basic definitions on physical dependence measures stated before Assumption 2.3. A triangular array $\{\varepsilon_{i,n}|1 \le i \le n\}_{n\in\mathbb{N}}$ of random variables is called *locally stationary* if there exists a possibly non-linear filter $G : [0,1] \times \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ which is continuous in its first argument such that $\varepsilon_{i,n} = G(i/n, \mathcal{F}_i)$ for all $i = 1, \dots, n$ and $n \in \mathbb{N}$. The physical dependence measure defined in (2.11) may be extended to a filter $G = G(\cdot, \cdot)$ with $\sup_{t\in[0,1]} \|G(t, \mathcal{F}_0)\|_{q,\Omega} < \infty$ by

$$\delta_q(G, i) = \sup_{t\in[0,1]} \|G(t, \mathcal{F}_i) - G(t, \mathcal{F}_i^*)\|_{q,\Omega}, \quad i \in \mathbb{N}.$$

Assumption 6.1. The triangular array $\{(\varepsilon_{i,n})_{1\le i\le n}\}_{n\in\mathbb{N}}$ in model (6.1) is centered and locally stationary with a filter function G that satisfies $\sup_{t\in[0,1]} \|G(t, \mathcal{F}_0)\|_{4,\Omega} < \infty$. Moreover, the following conditions are met:
(1) There is a constant $\chi \in (0,1)$ such that $\delta_4(G, i) = \mathcal{O}(\chi^i)$, as $i \to \infty$.
(2) The filter G is Lipschitz continuous with respect to $\|\cdot\|_{4,\Omega}$, that is

$$\sup_{0\le s<t\le 1} \|G(t, \mathcal{F}_0) - G(s, \mathcal{F}_0)\|_{4,\Omega}/|t - s| < \infty.$$

(3) The *long-run variance function* of $\{(\varepsilon_{i,n})_{1\le i\le n}\}_{n\in\mathbb{N}}$, defined as

$$\sigma^2(t) = \sum_{i=-\infty}^{\infty} \mathrm{Cov}\big(G(t, \mathcal{F}_i), G(t, \mathcal{F}_0)\big), \qquad t \in [0,1],$$

is Lipschitz continuous and bounded away from zero, i.e., $\sigma_{\min}^2 := \inf_{t\in[0,1]} \sigma^2(t) > 0$.

As $\sigma^2(t)$ can be interpreted as the (asymptotic) variance of a local mean of the data in a neighbourhood of the point $t \in [0,1]$, it is reasonable to compare deviations relative to this local noise. More precisely, consider a real-valued functional $g(\mu, \sigma)$ depending on the mean μ and on the square root of the long-run variance function $\sigma = \sqrt{\sigma^2}$, and define the distance

$$d_\infty^\sigma := \sup_{t\in[x_0,x_1]} \big|\mu(t)/\sigma(t) - g(\mu, \sigma)\big|.$$

We are interested in the hypotheses

$$H_0 : d_\infty^\sigma \le \Delta \quad \text{vs.} \quad H_1 : d_\infty^\sigma > \Delta, \tag{6.2}$$

for some $\Delta \ge 0$. As in the stationary case, we need some regularity of the function

$$d^\sigma(t) = \mu(t)/\sigma(t) - g(\mu, \sigma), \qquad t \in [0,1],$$

and a suitable estimator for $g(\mu, \sigma)$. Roughly speaking, the next two assumptions correspond to the assumptions in Section 2, where the function μ is replaced by μ/σ.

Assumption 6.2. $\hat{g}_n$ is an estimator of $g(\mu, \sigma)$ such that

$$|\hat{g}_n - g(\mu, \sigma)| = o_{\mathbb{P}}\Big(\frac{1}{\sqrt{nh_n|\log(h_n)|}}\Big).$$

Assumption 6.3.
  (i) The functions μ and σ are twice differentiable with Lipschitz continuous second derivatives.
  (ii) Define $\mathcal{E}^{\sigma} = \mathcal{E}^+ \cup \mathcal{E}^-$, where

$$\mathcal{E}^{\pm} = \{t \in [x_0, x_1] : \pm d^{\sigma}(t) = \|d^{\sigma}\|_{\infty}\}.$$

There exists a constant $\gamma > 0$ such that $d^{\sigma} = \mu/\sigma - g(\mu, \sigma)$ is concave (convex) on $U_{\gamma}(t) := \{s \in [0,1] : |s - t| < \gamma\}$, for any $t \in \mathcal{E}^+$ ($t \in \mathcal{E}^-$).

Further, we will need an estimator of the (time-dependent) long-run variance. For this purpose we follow Dette and Wu (2019) and define the partial sums $S_{j,k} = \sum_{i=j}^{k} X_{i,n}$ and the weight function $\omega_{\tau_n}(t, j) = K_{\tau_n}(j/n - t)/\{\sum_{i=1}^{n} K_{\tau_n}(i/n - t)\}$, for some positive bandwidth sequence $\tau_n = o(1)$. For some integer sequence $(m_n)_{n\in\mathbb{N}}$ with $m_n \to \infty$ and $m_n \ll n$, define

$$\hat{\sigma}^2(t) = \hat{\sigma}_{\tau_n, m_n}^2(t) = \sum_{j=1}^{n} \omega_{\tau_n}(t, j) \frac{(S_{j-m_n+1, j} - S_{j+1, j+m_n})^2}{2m_n}, \qquad (6.3)$$

for $t \in [m_n/n, 1 - m_n/n]$. Extend this estimator to the whole interval $[0, 1]$ by the definition $\hat{\sigma}^2(t) = \hat{\sigma}^2(m_n/n)$, for $t \in [0, m_n/n)$, and $\hat{\sigma}^2(t) = \hat{\sigma}^2(1 - m_n/n)$, for $t \in (1 - m_n/n, 1]$. The following result specifies the convergence rate of $\hat{\sigma}^2$, and is proved in the Appendix.

Theorem 6.4. *Let Assumption 6.1 and 6.3(i) be satisfied, and assume that the smoothing parameter sequences $\tau_n > 0$ and $m_n \in \mathbb{N}$ satisfy*

$$\tau_n \to 0, \quad m_n \to \infty, \quad \frac{m_n^{1/4}}{\sqrt{n\tau_n}} \to 0, \quad \frac{m_n^{5/2}}{n} \to 0.$$

Then, with $\gamma_n = \tau_n + m_n/n$,

$$\sup_{t\in[\gamma_n, 1-\gamma_n]} |\hat{\sigma}^2(t) - \sigma^2(t)| = \mathcal{O}_{\mathbb{P}}\Big(\frac{m_n^{1/4}}{\sqrt{n\tau_n}} + \frac{1}{m_n} + \tau_n^2 + \frac{m_n^{5/2}}{n}\Big) = o_{\mathbb{P}}(1).$$

Based on this estimator, similar results as stated in Sections 3 and 4 can be derived. To be precise, define

$$d_{\infty,n}^{\sigma} = \sup_{t\in I_n} |d^{\sigma}(t)|, \qquad \hat{d}_{\infty,n}^{\sigma} = \sup_{t\in I_n} |\hat{d}_n^{\sigma}(t)|,$$

where $\hat{d}_n^{\sigma}(t) = \tilde{\mu}_{h_n}(t)/\hat{\sigma}(t) - \hat{g}_n$. Moreover, for some positive sequence $(\rho_n)_{n\in\mathbb{N}} = o(1)$, let $\mathcal{E}_n^{\sigma} = \mathcal{E}_n^+ \cup \mathcal{E}_n^-$ and $\hat{\mathcal{E}}_n^{\sigma} = \hat{\mathcal{E}}_n^+ \cup \hat{\mathcal{E}}_n^-$, where

$$\mathcal{E}_n^{\pm} = \{t \in I_n : d_{\infty,n}^{\sigma} \mp d_n^{\sigma}(t) \leq \rho_n\}, \qquad \hat{\mathcal{E}}_n^{\pm} = \{t \in I_n : \hat{d}_{\infty,n}^{\sigma} \mp \hat{d}_n^{\sigma}(t) \leq \rho_n\}.$$

Let $G_{n,j}^{\sigma}$ be defined as in (4.3) ($j = 1$) and (4.4) ($j = 2$), but with $\mathcal{E}_n$ replaced by $\mathcal{E}_n^{\sigma}$.

Theorem 6.5. *Suppose that Assumptions 2.2, 6.1, 6.2 and 6.3 hold, and assume that $h_n \to 0$, $\tau_n \to 0$ and $m_n \to \infty$, such that $nh_n \to \infty$ and*

$$|\log(h_n)|\frac{\log^4 n}{n^{1/2}h_n} \leq C < \infty,$$

$$|\log(h_n)|nh_n^7 \to 0,$$

$$|\log(h_n)|^{1/2}\Big(\frac{m_n^{1/4}\sqrt{h_n}}{\tau_n} + \frac{\sqrt{nh_n}}{m_n} + \sqrt{nh_n}\tau_n^2 + \sqrt{\frac{h_n m_n^5}{n}}\Big) \to 0,$$

Moreover, assume that $\rho_n \to 0$ and $\rho_n^2 nh_n/|\log(h_n)| \to \infty$, and if $\lambda(\mathcal{E}^\sigma) = 0$, assume additionally $\rho_n^{1/2}h_n^{-1} \to \infty$. If $d_\infty^\sigma > 0$, then

$$\mathbb{P}\Big(\frac{\ell_n(\hat{\mathcal{E}}_n^\sigma)\sqrt{nh_n}}{\|K^*\|_2}(\hat{d}_{\infty,n}^\sigma - d_{\infty,n}^\sigma) - \ell_n^2(\hat{\mathcal{E}}_n^\sigma) \leq x\Big) \geq \mathbb{P}(G_{n,1}^\sigma \leq x) + o(1),$$

with equality if $\lambda(\mathcal{E}^\sigma) > 0$. If $d_\infty^\sigma = 0$, then

$$\mathbb{P}\Big(\frac{\ell_n(\hat{\mathcal{E}}_n^\sigma)\sqrt{nh_n}}{\sigma\|K^*\|_2}(\hat{d}_{\infty,n}^\sigma - d_{\infty,n}^\sigma) - \ell_n^2(\hat{\mathcal{E}}_n^\sigma) \leq x\Big) = \mathbb{P}(G_{n,2}^\sigma \leq x) + o(1).$$

Moreover, $G_{n,1}^\sigma \rightsquigarrow \mathrm{Gum}_0$ and $G_{n,2}^\sigma \rightsquigarrow \mathrm{Gum}_{\log(2)}$.

Tests for the hypothesis (6.2) can be derived in a similar way as in Section 4. Exemplary, we consider the analogue of test (4.5) (a test based on using the representation of $G_{n,j}^\sigma$ can be derived similarly). The null hypothesis in (6.2) is rejected whenever

$$\sup_{t \in I_n}|\hat{d}_n^\sigma(t)| > \{q_{a,1-\alpha} + \ell_n^2(\hat{\mathcal{E}}_n^\sigma)\}\frac{\|K^*\|_2}{\sqrt{nh_n}\ell_n(\hat{\mathcal{E}}_n^\sigma)} + \Delta, \tag{6.4}$$

where $a = 0$ if $\Delta > 0$ and $a = \log(2)$ if $\Delta = 0$.

Corollary 6.6. *Under the assumptions of Theorem 6.5, the test defined by the decision rule (6.4) is consistent and has asymptotic level α.*

7. Finite sample results

We investigate the finite sample properties of the new methodology by means of a simulation study and illustrate its application in a data example.

7.1. **Monte Carlo simulation study**

A large scale Monte Carlo simulation study was performed to analyse the finite-sample properties of the proposed tests. Two classes of mean functions μ were considered, the first one with $\lambda(\mathcal{E}) = 0$ and the second one with $\lambda(\mathcal{E}) > 0$.

The first class of models is based on the mean function

$$\mu_a^{(1)}(x) = 10 + \tfrac{1}{2}\sin(8\pi x) + a\big(x - \tfrac{1}{4}\big)^2 \mathbb{1}\big(x > \tfrac{1}{4}\big),$$

which is plotted in Figure 1 for various choices of a. We considered the testing problem in (1.3) with $x_0 = 1/4$ and with $\Delta = 1$, that is

$$H_0 : d_\infty = \sup_{t \in [1/4,1]}\Big|\mu_a^{(1)}(t) - 4\int_0^{1/4}\mu_a^{(1)}(s)\,\mathrm{d}s\Big| \leq 1 \qquad \text{vs.} \qquad H_1 : d_\infty > 1. \tag{7.1}$$

Such a scenario might for instance be encountered and of interest in the context of climate change. Note that $\|\mu_{a^*}^{(1)} - g(\mu_{a^*}^{(1)})\| = \Delta$ for $a^* = \frac{128}{81} \approx 1.58$, whereas for $a < a^*$ and $a > a^*$ we have $\|\mu_a^{(1)} - g(\mu_a^{(1)})\|_\infty < \Delta$ and $\|\mu_a^{(1)} - g(\mu_a^{(1)})\|_\infty > \Delta$, respectively.
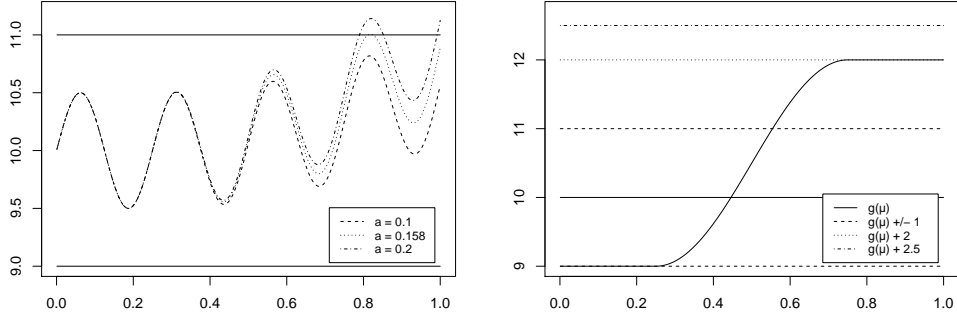
FIGURE 1. *Left: The mean function $\mu_a^{(1)}$ is plotted for three choices of a. Right: The mean function $\mu^{(2)}$ is plotted, alongside with $g + \Delta$ for five choices of the threshold Δ.*

The second class of models is based on the mean function

$$\mu^{(2)}(x) = \begin{cases} 9 & \text{for } x \leq \frac{1}{4} \\ \frac{3}{2}\sin(2\pi x) + 10.5 & \text{for } \frac{1}{4} < x \leq \frac{3}{4} \\ 12 & \text{for } \frac{3}{4} < x, \end{cases}$$

again plotted in Figure 1. For models involving $\mu^{(2)}$, we considered the testing problem in (2.3) with $x_0 = 0$, $x_1 = 1$, $g(\mu) \equiv 10$ and various choices of $\Delta > 0$, that is

$$H_0 : d_\infty = \sup_{t \in [0,1]} \left| \mu^{(2)}(t) - 10 \right| \leq \Delta \quad \text{vs.} \quad H_1 : d_\infty > \Delta. \qquad (7.2)$$

Such a setting might be encountered in quality control, where deviations from a target value might occur gradually due to wear and tear (and eventual failure) of a component of a complex system. Note that $\|\mu^{(2)} - g(\mu^{(2)})\|_\infty \leq \Delta$ for $\Delta \geq 2$, whereas $\|\mu^{(2)} - g(\mu^{(2)})\|_\infty > \Delta$ for $\Delta < 2$.

For both choices of the mean function μ we chose three different error processes $(\varepsilon_i)_{i \in \mathbb{Z}}$ in model (2.1), that is

$$\begin{aligned} \text{(IID)} \quad & \varepsilon_i = \tfrac{1}{2}\eta_i \\ \text{(MA)} \quad & \varepsilon_i = \tfrac{1}{\sqrt{5}}\big(\eta_i + \tfrac{1}{2}\eta_{i-1}\big) \\ \text{(AR)} \quad & \varepsilon_i = \tfrac{\sqrt{3}}{4}\big(\eta_i + \tfrac{1}{2}\varepsilon_{i-1}\big), \end{aligned}$$

were $(\eta_i)_{i \in \mathbb{Z}}$ is an i.i.d. sequence of standard normally distributed random variables. In particular, we have $\text{Var}(\varepsilon_i) = \frac{1}{4}$ for all error processes under consideration.

The choice of the bandwidth h_n for the estimator $\tilde{\mu}_{h_n}$ is crucial to avoid both overfitting and oversmoothing. For this purpose, we employ the following k-fold cross-validation procedure with $k = 10$ (as recommended by Hastie et al., 2009, page 242).

Algorithm 7.1 (Cross-Validation for the Choice of h_n)**.**
1. Split the observed data randomly in $k = 10$ sets $S_1, \dots, S_{10}$ of equal length.

$\mu_a^{(1)}$		test (2.9)			test (3.10)			test (4.6)		
a	$d_\infty - \Delta$	200	500	1000	200	500	1000	200	500	1000
Panel A: iid errors										
1.0	-0.18	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1.5	-0.03	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1.58	**0.00**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.2**
2.0	0.13	0.0	0.1	1.1	0.0	0.2	2.4	0.0	3.3	23.1
2.5	0.29	0.0	5.6	66.9	0.0	9.0	77.6	0.0	29.9	97.8
3.0	0.45	0.0	34.8	99.8	0.0	39.9	99.9	0.2	57.3	100.0
Panel B: MA errors										
1.0	-0.18	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1.5	-0.03	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1.58	**0.00**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.1**	**0.3**
2.0	0.13	0.0	0.3	0.9	0.0	0.5	2.1	0.0	3.7	18.7
2.5	0.29	0.1	4.6	40.2	0.1	7.4	51.7	0.2	27.0	87.9
3.0	0.45	0.1	25.4	96.1	0.1	31.0	97.9	0.5	52.8	99.7
Panel C: AR errors										
1.0	-0.18	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1.5	-0.03	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.5	1.0
1.58	**0.00**	**0.0**	**0.1**	**0.0**	**0.0**	**0.3**	**0.0**	**0.1**	**1.4**	**1.5**
2.0	0.13	0.0	0.8	1.8	0.0	1.4	3.3	0.0	7.8	23.1
2.5	0.29	0.0	4.9	29.4	0.1	8.4	40.2	0.4	27.3	77.7
3.0	0.45	0.1	21.2	86.4	0.1	27.4	90.0	1.1	53.9	98.4

TABLE 1. *Empirical rejection rates of various tests for the hypotheses (7.1), different values for the parameter a, different error processes, and sample sizes $n = 200, 500, 1000$.*

2. For $h_n = \frac{1}{n}$ and each set S_i, calculate the Jackknife estimator $\tilde{\mu}_{h_n}^{(i)}$ based on the data in the remaining sets.

3. Based on the Jackknife estimators $\tilde{\mu}_{h_n}^{(i)}$ from Step (2), compute the mean squared prediction error

$$\text{MSE}_{h_n} = \frac{1}{1 - h_n/2} \sum_{i=1}^{10} \sum_{j \in S_i} \{X_{j,n} - \tilde{\mu}_{h_n}^{(i)}(j/n)\}^2.$$

4. Repeat Steps (2) and (3) for the bandwidths $h_n = \frac{2}{n}, \ldots, \frac{\lfloor n/2 \rfloor}{n}$
5. Choose the bandwidth h_n that minimises the mean squared prediction error MSE_{h_n}.

Throughout, we employed the quartic kernel $K(x) = \frac{15}{16}(1 - x^2)^2$ for the local linear estimator. Preliminary simulation studies showed that different choices of the kernel led to similar results. The level α was chosen as 5% and ρ_n has been set to $\rho_n = \ell_n^{1+\varepsilon}/\sqrt{nh_n}$ with $\varepsilon = 0.001$. The block length of the long-run variance estimator was chosen as

$$m_n = \max\left\{\left\lfloor \sqrt{\frac{|\hat{\gamma}_1| + \cdots + |\hat{\gamma}_4|}{|\hat{\gamma}_0| + \cdots + |\hat{\gamma}_4|}} n^{1/3} \right\rfloor, 1\right\},$$

where $\hat{\gamma}_k$ denotes the empirical autocovariance at lag k of the residuals $\hat{\varepsilon}_{i,n} = X_{i,n} - \tilde{\mu}_{h_n}(i/n)$, for $k = 0, \ldots, 4$. Note that m_n naturally adapts to the serial dependence of the residuals, with $m_n = 1$ if the absolute empirical autocorrelations are small.

$\mu^{(2)}$		test (2.9)			test (3.10)			test (4.6)		
Δ	$d_\infty - \Delta$	200	500	1000	200	500	1000	200	500	1000
Panel A: iid errors										
1.0	1.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
1.5	0.5	70.1	99.0	100.0	76.0	99.5	100.0	92.4	99.9	100.0
1.75	0.25	2.8	18.5	94.2	5.9	28.1	96.9	43.3	73.6	99.7
2.0	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.2**	**3.0**
2.25	-0.25	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Panel B: MA errors										
1.0	1.0	97.9	100.0	100.0	98.8	100.0	100.0	100.0	100.0	100.0
1.5	0.5	46.9	93.9	100.0	58.2	97.2	100.0	86.0	99.7	100.0
1.75	0.25	2.6	11.0	76.3	5.0	16.9	83.9	32.5	61.3	97.8
2.0	**0.0**	**0.0**	**0.0**	**0.1**	**0.0**	**0.0**	**0.1**	**0.6**	**0.4**	**3.7**
2.25	-0.25	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Panel C: AR errors										
1.0	1.0	96.6	100.0	100.0	98.0	100.0	100.0	99.8	100.0	100.0
1.5	0.5	42.7	84.8	99.8	51.7	90.1	100.0	80.1	99.2	100.0
1.75	0.25	5.6	11.1	55.9	8.7	17.3	67.1	33.4	56.1	91.5
2.0	**0.0**	**0.0**	**0.0**	**0.1**	**0.2**	**0.2**	**0.5**	**2.1**	**1.2**	**5.1**
2.25	-0.25	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

TABLE 2. *Empirical rejection rates of different tests for the hypotheses (7.2) for $\mu = \mu^{(2)}$, different error processes, different choices of Δ and sample sizes $n = 200, 500, 1000$.*

The quantiles of $G_{n,j}$ are calculated by 2000 simulation runs. The empirical rejection rates of the null hypothesis $H_0 : d_\infty \leq \Delta$ are based on $N = 1000$ simulation runs each and are displayed in Tables 1 and 2 for the test (2.9) (based on the confidence band), the test (3.10) (based on the estimate of the sup-norm and the bound for the quantile of the limit distribution) and for the test (4.6) (based on the estimates of the sup-norm and the Lebesgue measure of the extremal sets). Results for the test (4.5), which is the analogue of (4.6) but with quantiles depending on the Gumbel-distribution, are not presented as they were always inferior to those of (4.6). The sample size was chosen as $n = 200$, 500 and 1000. For the tests (2.9) and (3.10) we used ℓ'_n as defined in (3.4) instead of ℓ_n in (3.1) (as pointed out in Section 3 this makes asymptotically no difference). The lines marked in boldface indicate the boundary of the null hypothesis, that is, the parameter where $d_\infty = \Delta$. In order to achieve large power it is desirable that the empirical level of the test is close to the nominal level α for those models.

To interpret the empirical rejection rates, note that the null hypothesis in the models involving $\mu_a^{(1)}$ is true if and only if $a \leq a^* \approx 1.58$. Likewise, for the models involving $\mu^{(2)}$ the null hypothesis is true if and only if $\Delta \geq 2$. It can be seen that all tests under consideration are conservative, in particular for the models involving $\mu_a^{(1)}$. Recall that the theory in Sections 2–4 suggests that tests (2.9) and (3.10) should be conservative for all models under consideration, while test (4.6) should either be conservative, or yield rejection rates close to the nominal level on the boundary of the null hypothesis for models involving $\mu^{(2)}$ (for which $\mu(\mathcal{E}) > 0$). The empirical findings
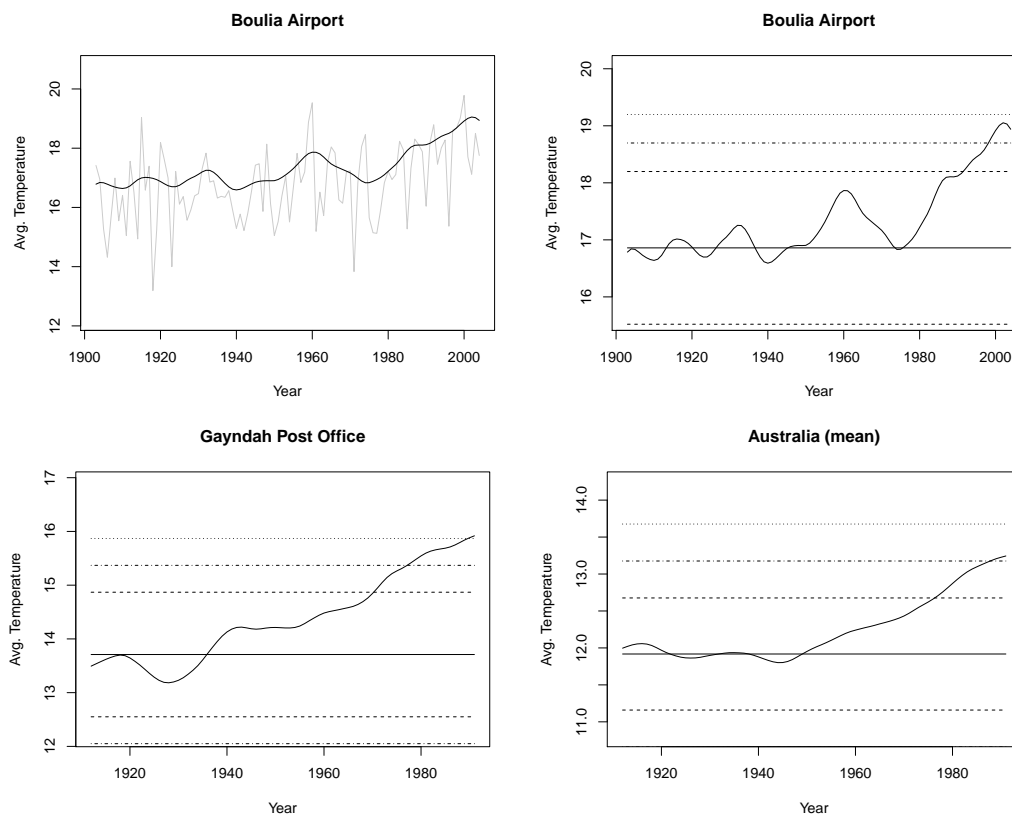
FIGURE 2. *Top left: Raw data of the temperature in Boulia. Top Right: estimate $\tilde{\mu}_h$ for the mean temperature in Boulia (solid curve); mean over the temperature from the beginning of the records until 1950 (solid straight line); boundary of the test decision for $\Delta = 0.5$, $\Delta = 1.0$ and $\Delta = 1.5$ (dashed line, dotted and dashed line, dotted line, respectively). Bottom left: Same as top right, but for Gayndah Post Office. Bottom right: Same as top right, but for the mean over the all stations under consideration.*

perfectly correspond to this theoretical prediction. In terms of power, the results are similar, with test (4.6) clearly being the most powerful. The superiority in terms of power is especially visible for alternatives for which $d_\infty - \Delta$ is small (say, around or below $1/4$, which is half the standard deviation of the noise ε_t). For $d_\infty - \Delta$ as low as 0.13 (model $\mu_a^{(1)}$ with $a = 2$), test (4.6) is the only test with a non-trivial power.

7.2. **Case Study**

Time series with possibly non-constant mean naturally arise in the field of meteorology. To illustrate the proposed methodology, we consider the mean of daily minimal temperatures (in degrees Celsius) over the month of July for a period of approximately 120 years at eight different places in Australia. Exemplary, the observed temperature curve at the weather station in Boulia is plotted in the upper left of Figure 2, alongside with its estimated smooth mean curve $\tilde{\mu}$. At each station j, we

Δ	0.5	1.0	1.5	0.5	1.0	1.5
Boulia	**0.0**	**0.6**	9.6	1957	1960	∞
Cape Otway	62.8	100.0	100.0	∞	∞	∞
Gayndah	**0.0**	**0.0**	**3.5**	1951	1969	1974
Gunnedah	**0.0**	**0.1**	**4.3**	1951	1956	1976
Hobart	**3.1**	98.0	100.0	1975	∞	∞
Melbourne	**0.0**	**0.2**	73.5	1968	1976	∞
Robe	78.3	100.0	100.0	∞	∞	∞
Sydney	**0.0**	8.8	96.0	1978	∞	∞
Australia (mean)	**0.0**	**1.4**	99.7	1970	1982	∞

TABLE 3. *p-values of test* (4.6) *for the respective null hypotheses in percent (left part) and estimated time of first relevant deviation (right part). Significant p-values (below 0.05) are in boldface.*

tested for relevant deviations of the temperature within the second half of the 20th century from g_j, the mean temperature over a historic reference period ranging from the late 19th century to 1950 at that station. As a threshold Δ, we chose $0.5, 1$ and 1.5 degrees Celsius. The estimated mean curve, alongside with a line corresponding to the overall mean g_j and three lines corresponding to the critical values for test (4.6) can be found in Figure 2 (Stations Boulia and Gayndah Post Office, as well as the mean over all weather stations).

The results for all stations under consideration can be found in Table 3, where we also provide estimates for the first point in time exhibiting a relevant deviation from the historic period. The *p*-values are highly significant at all but 2 stations for $\Delta = 0.5$ degree Celsius, and at all but 3 stations for $\Delta = 1$ degree Celsius. For $\Delta = 1.5$ degree Celsius, only two stations exhibit *p*-values slightly below 0.05. Finally, it is worthwhile to mention that tests (2.9) and (3.10) yield no significant *p*-values at all. Thus, the proposed test in (4.6) is clearly more adequate in the given context for detecting relevant changes.

APPENDIX A. PROOFS

A.1. **Preliminaries**

In this section we present some general results for Model (6.1) with a locally stationary error process. In Remark A.3 we specialize the results to the stationary case, as they are needed for the proofs of the statements in Section 3 and 4.

Recall the definition of the local long-run variance estimator $\hat{\sigma}^2(t)$ in (6.3) and let

$$Z_n(t) = \frac{\sqrt{nh_n}}{\|K^*\|_2} \left(\frac{\tilde{\mu}_{h_n}(t)}{\hat{\sigma}(t)} - \frac{\mu(t)}{\sigma(t)} \right), \qquad t \in (0, 1).$$

We are going to prove weak convergence results for the supremum of the random functions Z_n and $|Z_n|$ over sets $\mathcal{A} \subset [0, 1]$. Note that similar results are available for closely related processes in the case of independent data (see, e.g., Johnston, 1982; Xia, 1998; Proksch, 2016, among others), but the problem is less well investigated in the dependent case (see, e.g., Hansen, 2008; Li et al., 2017; Cao et al., 2018, among others).

Theorem A.1. *Let $\mathcal{A}$ denote a compact subset of the interval $[x_0, x_1]$ (or $(0, 1)$ if $x_0 = 0, x_1 = 1$) with positive Lebesgue measure $\lambda(\mathcal{A}) > 0$. Assume that $\mathcal{A}$ can be represented as a finite union of disjoint compact intervals, that is $\mathcal{A} = \bigcup_{i=1}^{m} [x_{i,1}, x_{i,2}]$, the case $x_{i,1} = x_{i,2}$ being allowed. If Assumptions 2.2, 6.1 and 6.3(i) are satisfied and if $h_n \to 0$, $\tau_n \to 0$ and $m_n \to \infty$, such that $m_n/n \to 0$, $nh_n \to \infty$ and*

$$|\log(h_n)|\frac{\log^4 n}{n^{1/2}h_n} \le C < \infty, \tag{A.1}$$

$$|\log(h_n)|nh_n^7 \to 0, \tag{A.2}$$

$$|\log(h_n)|^{1/2}\Big(\frac{m_n^{1/4}\sqrt{h_n}}{\tau_n} + \frac{\sqrt{nh_n}}{m_n} + \sqrt{nh_n}\tau_n^2 + \sqrt{\frac{h_n m_n^5}{n}}\Big) \to 0, \tag{A.3}$$

then, with $\ell_n(\mathcal{A})$ as defined in (4.2),

$$\Big\{\sup_{t \in \mathcal{A}} Z_n(t) - \ell_n(\mathcal{A})\Big\}\ell_n(\mathcal{A}) \rightsquigarrow \mathrm{Gum}_0, \quad \Big\{\sup_{t \in \mathcal{A}} |Z_n(t)| - \ell_n(\mathcal{A})\Big\}\ell_n(\mathcal{A}) \rightsquigarrow \mathrm{Gum}_{\log(2)}.$$

Proof. First observe that

$$\sup_{t \in I_n} |Z_n(t) - Z_{n,1}(t)| = \mathcal{O}_{\mathbb{P}}\Big(\frac{m_n^{1/4}\sqrt{h_n}}{\tau_n} + \frac{\sqrt{nh_n}}{m_n} + \sqrt{nh_n}\tau_n^2 + \sqrt{\frac{h_n m_n^5}{n}}\Big), \tag{A.4}$$

where

$$Z_{n,1}(t) = \frac{\sqrt{nh_n}}{\sigma(t)\|K^*\|_2}\{\tilde{\mu}_{h_n}(t) - \mu(t)\}.$$

Indeed, by Theorem 6.4 and part (3) of Assumption 6.1,

$$\sup_{t \in [\gamma_n, 1-\gamma_n]} \Big|\frac{1}{\hat{\sigma}(t)} - \frac{1}{\sigma(t)}\Big| = \sup_{t \in [\gamma_n, 1-\gamma_n]} \frac{1}{\sigma(t)\hat{\sigma}(t)\{\sigma(t) + \hat{\sigma}(t)\}}|\sigma^2(t) - \hat{\sigma}^2(t)|$$

$$= \mathcal{O}_{\mathbb{P}}\Big(\frac{m_n^{1/4}}{\sqrt{n}\tau_n} + \frac{1}{m_n} + \tau_n^2 + \frac{m_n^{5/2}}{n}\Big).$$

Recalling the definition of the Jackknife estimator $\tilde{\mu}_{h_n}$ in (2.6) and the definition of the kernel K^* in (3.3) it follows from Lemma C.2 of the Supplementary Material of Dette and Wu (2019) that

$$\sup_{t \in I_n} \Big|\tilde{\mu}_{h_n}(t) - \mu(t) - \frac{1}{nh_n}\sum_{i=1}^{n} \varepsilon_{i,n} K_{h_n}^*\big(\tfrac{i}{n} - t\big)\Big| = \mathcal{O}(h_n^3 + 1/(nh_n)). \tag{A.5}$$

By Proposition 5 of Zhou (2013), on a possibly richer probability space, there is a sequence $(V_i)_{i \in \mathbb{N}}$ of independent, standard normally distributed random variables such that

$$\max_{1 \le j \le n} \Big|\sum_{i=1}^{j} \varepsilon_{i,n} - \sum_{i=1}^{j} \sigma\big(\tfrac{i}{n}\big)V_i\Big| = o_{\mathbb{P}}(n^{1/4}\log^2 n).$$

Then, with the notation $S_j = \sum_{i=1}^{j-1}\{\varepsilon_{i,n} - \sigma\big(\tfrac{i}{n}\big)V_i\}$ for $j \ge 1$, $S_0 = 0$ and $t \in I_n = [x_0 \vee h_n, x_1 \wedge (1 - h_n)]$, summation by parts leads to

$$\Big|\sum_{i=1}^{n} \varepsilon_{i,n} K_{h_n}^*\big(\tfrac{i}{n} - t\big) - \sum_{i=1}^{n} \sigma\big(\tfrac{i}{n}\big)V_i K_{h_n}^*\big(\tfrac{i}{n} - t\big)\Big|$$

$$= \Big|\sum_{i=1}^{n} (S_{i+1} - S_i) K_{h_n}^*\big(\tfrac{i}{n} - t\big)\Big|$$

$$= \left| K^*_{h_n}(1-t)S_{n+1} - K^*_{h_n}\left(\tfrac{1}{n} - t\right)S_1 - \sum_{i=2}^{n} S_i\left(K^*_{h_n}\left(\tfrac{i}{n} - t\right) - K^*_{h_n}\left(\tfrac{i-1}{n} - t\right) \right) \right|$$

$$\leq \sum_{i=2}^{n} \left| K^*_{h_n}\left(\tfrac{i}{n} - t\right) - K^*_{h_n}\left(\tfrac{i-1}{n} - t\right) \right| o_{\mathbb{P}}(n^{1/4}\log^2 n)$$

$$= \sum_{i=\lceil nt-nh_n \rceil}^{\lceil nt+nh_n \rceil} \left| K^*_{h_n}\left(\tfrac{i}{n} - t\right) - K^*_{h_n}\left(\tfrac{i-1}{n} - t\right) \right| o_{\mathbb{P}}(n^{1/4}\log^2 n) = o_{\mathbb{P}}(n^{1/4}\log^2 n),$$

because K^* is Lipschitz continuous and $\mathrm{supp}(K^*) = [-1, 1]$. Thus, it follows from (A.5)

$$\sup_{t\in I_n} \left| \tilde{\mu}_{h_n}(t) - \mu(t) - \frac{1}{nh_n}\sum_{i=1}^{n}\sigma\left(\tfrac{i}{n}\right)V_i K^*_{h_n}\left(\tfrac{i}{n} - t\right) \right| = \mathcal{O}(h_n^3) + o_{\mathbb{P}}\left(\tfrac{\log^2 n}{n^{3/4}h_n}\right). \qquad \text{(A.6)}$$

and consequently

$$\sup_{t\in I_n} |Z_{n,1}(t) - Z_{n,2}(t)| = \mathcal{O}(\sqrt{n}h_n^{7/2}) + o_{\mathbb{P}}\left(\tfrac{\log^2 n}{n^{1/4}h_n^{1/2}}\right), \qquad \text{(A.7)}$$

where

$$Z_{n,2}(t) = \frac{1}{\sigma(t)\|K^*\|_2\sqrt{nh_n}}\sum_{i=1}^{n}\sigma\left(\tfrac{i}{n}\right)V_i K^*_{h_n}\left(\tfrac{i}{n} - t\right).$$

Further, let

$$Z_{n,3}(t) = \frac{1}{\|K^*\|_2\sqrt{nh_n}}\sum_{i=1}^{n}V_i K^*_{h_n}\left(\tfrac{i}{n} - t\right). \qquad \text{(A.8)}$$

Recalling the definition of Λ_K in (3.2) and $\ell_n(\mathcal{A})$ in (4.2) we obtain by similar arguments as in the proof of Theorem 2 in Proksch (2016) that

$$\left\{ \sup_{t\in\mathcal{A}} Z_{n,3}(t) - \ell_n(\mathcal{A}) \right\}\ell_n(\mathcal{A}) \rightsquigarrow \mathrm{Gum}_0, \quad \left\{ \sup_{t\in\mathcal{A}} |Z_{n,3}(t)| - \ell_n(\mathcal{A}) \right\}\ell_n(\mathcal{A}) \rightsquigarrow \mathrm{Gum}_{\log(2)}.$$

The assertion follows from the stochastic expansions (A.4), (A.7) and

$$\sup_{t\in I_n} |Z_{n,2}(t) - Z_{n,3}(t)| = o_{\mathbb{P}}(1), \qquad \text{(A.9)}$$

observing the assumptions (A.1), (A.2) and (A.3).

It remains to prove (A.9). First observe that

$$\mathbb{E}\left[\left(Z_{n,2}(t) - Z_{n,3}(t) - Z_{n,2}(s) + Z_{n,3}(s) \right)^2 \right] \qquad \text{(A.10)}$$

$$= \frac{1}{\|K^*\|_2^2 nh_n}\sum_{i=1}^{n}\left\{ \frac{\sigma\left(\tfrac{i}{n}\right) - \sigma(t)}{\sigma(t)}K^*_{h_n}\left(\tfrac{i}{n} - t\right) - \frac{\sigma\left(\tfrac{i}{n}\right) - \sigma(s)}{\sigma(s)}K^*_{h_n}\left(\tfrac{i}{n} - s\right) \right\}^2.$$

Note that the only non zero summands are those with $|i/n - t| \leq h_n$ or $|i/n - s| \leq h_n$. In the following, we only consider the case $|i/n - s| \leq h_n$ as the case $|i/n - t| \leq h_n$ can be investigated with the same arguments. By Assumption 6.3(i) and the mean value theorem it exists $\xi \in \left(\tfrac{i}{n} \wedge s, \tfrac{i}{n} \vee s \right)$ such that

$$\frac{\sigma(i/n) - \sigma(t)}{\sigma(t)}K^*_{h_n}\left(\tfrac{i}{n} - t\right) - \frac{\sigma(i/n) - \sigma(s)}{\sigma(s)}K^*_{h_n}\left(\tfrac{i}{n} - s\right)$$

$$= \frac{\sigma(i/n)[\sigma(s)-\sigma(t)]}{\sigma(s)\sigma(t)} K^*_{h_n}\left(\tfrac{i}{n}-t\right) + \frac{\sigma(i/n)-\sigma(s)}{\sigma(s)}\left[K^*_{h_n}\left(\tfrac{i}{n}-t\right) - K^*_{h_n}\left(\tfrac{i}{n}-s\right)\right]$$

$$= \frac{\sigma(i/n)[\sigma(s)-\sigma(t)]}{\sigma(s)\sigma(t)} K^*_{h_n}\left(\tfrac{i}{n}-t\right) + \frac{\sigma'(\xi)}{\sigma(s)}\left(\tfrac{i}{n}-s\right)\left[K^*_{h_n}\left(\tfrac{i}{n}-t\right) - K^*_{h_n}\left(\tfrac{i}{n}-s\right)\right].$$

Since σ is bounded away from zero and Lipschitz continuous, the first summand on the right-hand side of the previous display can be bounded by $C|s-t|$. Since $|i/n-s| \leq h_n$ and K^* is Lipschitz continuous, the second summand can be bounded by $C|s-t|$ as well. Hence, the non-zero summands in (A.10) can be bounded by $C|s-t|^2$. As there are only $\mathcal{O}(nh_n)$ non-zero summands, the right-hand side of (A.10) can be bounded by $C|s-t|^2$. With this bound, Theorem 2.2.4 of van der Vaart and Wellner (1996) leads with $\psi(x) = x^2, \delta = 2h_n$ and $\eta = \sqrt{h_n}$ to

$$\mathbb{E}\left[\sup_{|s-t|<2h_n} \left\{Z_{n,2}(t) - Z_{n,3}(t) - Z_{n,2}(s) + Z_{n,3}(s)\right\}^2\right] \leq C h_n^{1/2}.$$

With the notation

$$\mathcal{T}_n = \{2jnh_n + 1 : j \in \mathbb{N}\} \cap [nx_0 \vee nh_n, nx_1 \wedge (n-nh_n)],$$

it follows by the triangle inequality

$$\mathbb{E}\left[\sup_{t\in I_n} \left\{Z_{n,2}(t) - Z_{n,3}(t)\right\}^2\right]^{1/2} \tag{A.11}$$

$$= \mathbb{E}\left[\max_{j\in\mathcal{T}_n} \sup_{|t-j/n|\leq 2h_n} \left\{Z_{n,2}(t) - Z_{n,3}(t)\right\}^2\right]^{1/2}$$

$$\leq \mathbb{E}\left[\max_{j\in\mathcal{T}_n} \left\{Z_{n,2}\left(\tfrac{j}{n}\right) - Z_{n,3}\left(\tfrac{j}{n}\right)\right\}^2\right]^{1/2}$$

$$\quad + \mathbb{E}\left[\sup_{|s-t|<2h_n} \left\{Z_{n,2}(t) - Z_{n,3}(t) - Z_{n,2}(s) + Z_{n,3}(s)\right\}^2\right]^{1/2}$$

$$\leq \mathbb{E}\left[\max_{j\in\mathcal{T}_n} \left\{Z_{n,2}\left(\tfrac{j}{n}\right) - Z_{n,3}\left(\tfrac{j}{n}\right)\right\}^2\right]^{1/2} + C\sqrt{h_n}.$$

Observe that the indices in the maximum on the right-hand side of the previous display have a distance of $2nh_n$. Further, the summation in the definition of $Z_{n,2}(j/n) - Z_{n,3}(j/n)$ ranges from $\lfloor j - nh_n \rfloor$ to $\lfloor j + nh_n \rfloor$, thus, the random variables in the latter maximum are independent and, by definition, normally distributed. Further, observe that $|\mathcal{T}_n| \leq \frac{1}{2h_n}$. With the notation $Z_j = Z_{n,2}(j/n) - Z_{n,3}(j/n)$ and $\sigma_j^2 = \mathrm{Var}(Z_j)$, it follows that, for all $t \in (0, 1/\{2\max_{j\in\mathcal{T}_n}\sigma_j^2\})$,

$$\exp\left(t\mathbb{E}\left[\max_{j\in\mathcal{T}_n} Z_j^2\right]\right) \leq \mathbb{E}\left[\exp\left(t\max_{j\in\mathcal{T}_n} Z_j^2\right)\right] = \mathbb{E}\left[\max_{j\in\mathcal{T}_n} \exp(tZ_j^2)\right] \tag{A.12}$$

$$\leq \sum_{j\in\mathcal{T}_n} \mathbb{E}\left[\exp(tZ_j^2)\right] = \sum_{j\in\mathcal{T}_n} \mathbb{E}\left[\exp(t\sigma_j^2(Z_j/\sigma_j)^2)\right]$$

$$= \sum_{j\in\mathcal{T}_n} \frac{1}{\sqrt{1-2t\sigma_j^2}} \leq \frac{|\mathcal{T}_n|}{\sqrt{1-2t\max_{j\in\mathcal{T}_n}\sigma_j^2}},$$

where we have used the fact that the moment-generating function of the χ_1^2 distribution is $(1 - 2t)^{-1/2}$. Now observe that, by Lipschitz-continuity of σ, for all $j \in \mathcal{T}_n$,

$$
\begin{aligned}
\sigma_j^2 &= \mathbb{E}\big[\{Z_{n,2}(j/n) - Z_{n,3}(j/n)\}^2\big] \\
&= \frac{1}{\sigma^2(j/n)\|K^*\|_2^2 n h_n} \sum_{i=1}^n \Big\{\sigma\big(\tfrac{i}{n}\big) - \sigma\big(\tfrac{j}{n}\big)\Big\}^2 \Big\{K_{h_n}^*\big(\tfrac{i}{n} - \tfrac{j}{n}\big)\Big\}^2 \leq Ch_n^2.
\end{aligned}
$$

Thus, $t_n := (2Ch_n^2|\log(h_n)|)^{-1} \leq |\log h_n|^{-1} \times \{2\max_{j \in \mathcal{T}_n} \sigma_j^2\}^{-1}$ is a valid choice for t in (A.12), which leads to

$$
\begin{aligned}
\mathbb{E}\big[\max_{j=1}^{|\mathcal{T}_n|} Z_j^2\big] &\leq \frac{1}{t_n} \log\Big(\frac{|\mathcal{T}_n|}{\sqrt{1 - 2t_n Ch_n^2}}\Big) \\
&= 2Ch_n^2|\log(h_n)| \log\Big(\frac{1}{2h_n\sqrt{1 - |\log(h_n)|^{-1}}}\Big) = \mathcal{O}(h_n^2|\log(h_n)|^2).
\end{aligned}
$$

Plugging this bound into (A.11) yields $\mathbb{E}\big[\sup_{t \in I_n}\{Z_{n,2}(t) - Z_{n,3}(t)\}^2\big] = \mathcal{O}(h_n)$. In particular, $\sup_{t \in I_n}|Z_{n,2}(t) - Z_{n,3}(t)| = \mathcal{O}_{\mathbb{P}}(\sqrt{h_n})$, thus, (A.9) follows. □

Remark A.2. Conditions (A.1)–(A.3) are for instance satisfied if $h_n = n^{-c}$, for some $c \in (3/7, 1/2)$, $m_n = |\log h_n|^{1/2} \log n \sqrt{n h_n}$ and $\tau_n = m_n^{-1/2}$.

Remark A.3.
(i) The representation of $\mathcal{A}$ as a union of finitely many intervals is needed to ensure the *blowing up* property of $h_n^{-1}\mathcal{A}$ in order to apply Theorem 14.1 of Piterbarg (2012) in the proof of Theorem 2 of Proksch (2016). In fact, $\mathcal{A}$ can be replaced by any sequence $(\mathcal{A}_n)_{n \in \mathbb{N}}$ of subsets of I_n with $(h_n^{-1}\mathcal{A}_n)_{n \in \mathbb{N}}$ satisfying the blowing up property (cf. Definition 14.1 in Piterbarg, 2012).

(ii) It follows from the proof of Theorem A.1 that in the situation of a stationary error process as considered in Section 2 - 4 the weak convergence

$$
\Big\{\sup_{t \in \mathcal{A}} Z_{n,j}(t) - \ell_n(\mathcal{A})\Big\}\ell_n(\mathcal{A}) \rightsquigarrow \text{Gum}_0, \tag{A.13}
$$

$$
\Big\{\sup_{t \in \mathcal{A}} |Z_{n,j}(t)| - \ell_n(\mathcal{A})\Big\}\ell_n(\mathcal{A}) \rightsquigarrow \text{Gum}_{\log(2)}, \tag{A.14}
$$

remains valid for $Z_{n,3}$ as defined in (A.8), as well as for the processes

$$
Z_{n,1}(t) = \frac{\sqrt{n h_n}}{\sigma \|K^*\|_2}\{\tilde{\mu}_{h_n}(t) - \mu(t)\}, \quad \hat{Z}_{n,1}(t) = \frac{\sqrt{n h_n}}{\hat{\sigma}\|K^*\|_2}\{\tilde{\mu}_{h_n}(t) - \mu(t)\},
$$

where σ^2 is the long-run variance defined in (2.12), with corresponding estimator $\hat{\sigma}^2$ defined in (3.5). Moreover

$$
Z_{n,1} = Z_{n,3} + o_{\mathbb{P}}(|\log h_n|^{-1/2}). \tag{A.15}
$$

The proof does not require condition (A.3) (as there is no varying long-run variance which has to be estimated).

Remark A.4. The assertion regarding $\hat{Z}_{n,1}$ in Remark A.3(ii) allows for the construction of simultaneous $(1-\alpha)$-confidence bands for the regression function μ. More

precisely, a careful inspection reveals that we may replace $\ell_n(I_n)$ by $\ell_n = \ell_n([x_0, x_1])$ in the weak convergence result, which implies that the collection of intervals

$$\tilde{I}_n(t) = [\tilde{\mu}_{h_n}(t) - c_{n,\alpha}, \tilde{\mu}_{h_n}(t) + c_{n,\alpha}], \qquad t \in I_n,$$

with

$$c_{n,\alpha} = \left(q_{\log(2),1-\alpha} + \ell_n^2\right) \frac{\hat{\sigma}\|K^*\|_2}{\sqrt{nh_n}\ell_n}, \qquad (A.16)$$

defines an asymptotic simultaneous $(1-\alpha)$-confidence band for μ in model (2.1), i.e.,

$$\lim_{n\to\infty} \mathbb{P}(\mu(t) \in \tilde{I}_n(t) \ \forall\, t \in I_n) = \lim_{n\to\infty} \mathbb{P}(\sup_{t\in I_n} |\tilde{\mu}_{h_n}(t) - \mu(t)| \le c_{n,\alpha}) = 1 - \alpha.$$

As a further consequence, if additionally Assumption 2.6 is met, the intervals in (2.8) define an asymptotic simultaneous $(1-\alpha)$-confidence band for the function $\mu - g(\mu)$ and the decision rule (2.9) is a consistent, asymptotic level α test.

A.2. **Proofs for Section 3**

Proof of Theorem 3.1. The assertion in (3) is a consequence of Theorem 4.2 proven below: first, choose a positive sequence $\rho_n = o(1)$ such that $\rho_n h_n^{-2} \to \infty$. With $\mathcal{E}_n$ and $\ell_n(\mathcal{E}_n)$ as defined in (4.1) and (4.2) respectively, it holds $\ell_n(\mathcal{E}_n) \to \infty$ (Proposition A.8) and $\ell_n(\mathcal{E}_n) = o(\ell_n)$ since $\lambda(\mathcal{E}_n) \to 0$. Further, it holds $d_\infty > 0$. Thus, applying Theorem 4.2,

$$\tfrac{\ell_n\sqrt{nh_n}}{\sigma\|K^*\|_2}(\hat{d}_{\infty,n} - d_{\infty,n}) - \ell_n^2 \ll_S \ell_n\Big((G_{n,2} + o_{\mathbb{P}}(1))/\ell_n(\mathcal{E}_n) + \ell_n(\mathcal{E}_n) - \ell_n\Big),$$

which converges to $-\infty$ in probability as asserted.

Regarding (1) and (2), recall the definition of ℓ_n in (3.1) and let

$$D_n = \tfrac{\ell_n\sqrt{nh_n}}{\sigma\|K^*\|_2}\Big\{ \sup_{t\in I_n} |\tilde{\mu}_{h_n}(t) - g(\mu)| - \sup_{t\in I_n} |\mu(t) - g(\mu)| \Big\} - \ell_n^2. \qquad (A.17)$$

Observing that

$$\tfrac{\ell_n\sqrt{nh_n}}{\sigma\|K^*\|_2}\Big\{ \sup_{t\in I_n} |\tilde{\mu}_{h_n}(t) - \hat{g}_n| - \sup_{t\in I_n} |\tilde{\mu}_{h_n}(t) - g(\mu)| \Big\} = o_{\mathbb{P}}(1)$$

by Assumption 2.6, the assertion of Theorem 3.1 is a consequence of the weak convergence $D_n \rightsquigarrow \mathrm{Gum}_{\log\{\lambda(\mathcal{E})/(x_1-x_0)\}}$ if $d_\infty > 0$, and $D_n \rightsquigarrow \mathrm{Gum}_{\log(2)}$ if $d_\infty = 0$.

The statement in (2), i.e., for $d_\infty = 0$, now follows from Remark A.3(i) and (ii). The statement in (1), i.e., for $d_\infty > 0$, is a consequence of the following two propositions. $\square$

Proposition A.5. *If the assumptions of Theorem 3.1(1) are satisfied, then*

$$D_n(\mathcal{E}) = \tfrac{\ell_n\sqrt{nh_n}}{\sigma\|K^*\|_2}\Big\{ \sup_{t\in\mathcal{E}\cap I_n} |\tilde{\mu}_{h_n}(t) - g(\mu)| - \sup_{t\in\mathcal{E}\cap I_n} |\mu(t) - g(\mu)| \Big\} - \ell_n^2$$

converges weakly to $\mathrm{Gum}_{\log\{\lambda(\mathcal{E})/(x_1-x_0)\}}$.

Proposition A.6. *If the assumptions of Theorem 3.1(1) are satisfied, then, with D_n as defined in (A.17),*

$$\mathcal{R}_n = D_n - D_n(\mathcal{E}) = o_{\mathbb{P}}(1), \qquad n \to \infty.$$

Proof of Proposition A.5. We prove the proposition for the case that both sets $\mathcal{E}^+$ and $\mathcal{E}^-$ are non-empty. The other cases follow by the same arguments. Throughout the proof, since $\lambda(\mathcal{E}) > 0$, we may assume that n is sufficiently large such that $d_{\infty,n} = d_\infty$ and $\sup_{t \in \mathcal{E} \cap I_n} |d(t)| = \sup_{t \in \mathcal{E}} |d(t)| = d_\infty$.

Now, observe that

$$\frac{\ell_n}{\ell_n(\mathcal{E})} = \sqrt{\frac{\log(\frac{(x_1-x_0)\Lambda_K}{2\pi h_n})}{\log(\frac{\lambda(\mathcal{E})\Lambda_K}{2\pi h_n})}} = \sqrt{\frac{\log(x_1-x_0) + \log(\frac{\Lambda_K}{2\pi h_n})}{\log(\lambda(\mathcal{E})) + \log(\frac{\Lambda_K}{2\pi h_n})}} \to 1, \tag{A.18}$$

as $\lambda(\mathcal{E}) > 0$ by assumption. Recall the definition of $Z_{n,1}$ and $Z_{n,3}$ in Remark A.3(ii). Since $\mu(t) - g(\mu) = d_\infty$ for $t \in \mathcal{E}^+ \cap I_n$, we obtain from (A.15)

$$\frac{\ell_n \sqrt{n h_n}}{\sigma \|K^*\|_2} \sup_{t \in \mathcal{E}^+ \cap I_n} \left\{ -(\tilde{\mu}_{h_n}(t) - g(\mu)) \right\} \tag{A.19}$$

$$= \ell_n \sup_{t \in \mathcal{E}^+ \cap I_n} \left\{ -Z_{n,1}(t) - \frac{\sqrt{n h_n}}{\sigma \|K^*\|_2} (\mu(t) - g(\mu)) \right\}$$

$$= \ell_n \sup_{t \in \mathcal{E}^+ \cap I_n} \left\{ -Z_{n,3}(t) - \frac{\sqrt{n h_n}}{\sigma \|K^*\|_2} d_\infty \right\} + o_{\mathbb{P}}(1)$$

$$\stackrel{\mathcal{D}}{=} \ell_n \sup_{t \in \mathcal{E}^+ \cap I_n} Z_{n,3}(t) - \frac{\ell_n \sqrt{n h_n}}{\sigma \|K^*\|_2} d_\infty + o_{\mathbb{P}}(1)$$

$$\leq \ell_n \left(\sup_{t \in I_n} Z_{n,3}(t) - \ell_n \right) - \ell_n \left(\frac{\sqrt{n h_n}}{\sigma \|K^*\|_2} d_\infty - \ell_n \right) + o_{\mathbb{P}}(1).$$

Next, observe that the first term on the right-hand side of the previous display converges to a Gumbel distribution by (A.13), whereas the second term diverges to $-\infty$ since

$$\ell_n = \sqrt{n h_n} \frac{\ell_n}{n^{1/4} h_n^{1/2}} n^{-1/4} = o(\sqrt{n h_n}) \tag{A.20}$$

by the assumptions on h_n. In the same way, it may be shown that

$$\frac{\ell_n \sqrt{n h_n}}{\sigma \|K^*\|_2} \sup_{t \in \mathcal{E}^- \cap I_n} \left\{ \tilde{\mu}_{h_n}(t) - g(\mu) \right\} \to -\infty$$

in probability.

Now, note that $D_n(\mathcal{E}) = \max\{D_{n,1}(\mathcal{E}), D_{n,2}(\mathcal{E})\}$, where

$$D_{n,1}(\mathcal{E}) = \frac{\ell_n \sqrt{n h_n}}{\sigma \|K^*\|_2} \max \left\{ \sup_{t \in \mathcal{E}^+ \cap I_n} (\tilde{\mu}_{h_n}(t) - g(\mu) - d_\infty), \right.$$

$$\left. \sup_{t \in \mathcal{E}^- \cap I_n} (-\tilde{\mu}_{h_n}(t) + g(\mu) - d_\infty) \right\} - \ell_n^2$$

and

$$D_{n,2}(\mathcal{E}) = \frac{\ell_n \sqrt{n h_n}}{\sigma \|K^*\|_2} \max \left\{ \sup_{t \in \mathcal{E}^+ \cap I_n} (-\tilde{\mu}_{h_n}(t) + g(\mu) - d_\infty), \right.$$

$$\left. \sup_{t \in \mathcal{E}^- \cap I_n} (\tilde{\mu}_{h_n}(t) - g(\mu) - d_\infty) \right\} - \ell_n^2.$$

By the previous considerations, $D_{n,2}(\mathcal{E})$ diverges to $-\infty$, thus

$$D_n(\mathcal{E}) = D_{n,1}(\mathcal{E}) + o_{\mathbb{P}}(1). \tag{A.21}$$

Next, observe that by similar arguments as in (A.19)

$$D_{n,1}(\mathcal{E}) = \frac{\ell_n\sqrt{nh_n}}{\sigma\|K^*\|_2} \max\Big\{ \sup_{t\in\mathcal{E}^+\cap I_n} (\tilde{\mu}_{h_n}(t) - \mu(t)), \sup_{t\in\mathcal{E}^-\cap I_n} (-\tilde{\mu}_{h_n}(t) + \mu(t))\Big\} - \ell_n^2$$

$$= \ell_n \max\Big\{ \sup_{t\in\mathcal{E}^+\cap I_n} Z_{n,3}(t), \sup_{t\in\mathcal{E}^-\cap I_n} -Z_{n,3}(t)\Big\} - \ell_n^2 + o_{\mathbb{P}}(1).$$

The sets $\mathcal{E}^+$ and $\mathcal{E}^-$ are disjoint and bounded away from each other. Thus, there exists a positive integer $n_0 \in \mathbb{N}$ such that the arguments of the maximum in the previous display are stochastically independent for all $n \geq n_0$, and

$$D_{n,1}(\mathcal{E}) \stackrel{\mathcal{D}}{=} \ell_n \sup_{t\in\mathcal{E}\cap I_n} Z_{n,3}(t) - \ell_n^2 + o_{\mathbb{P}}(1)$$

$$= \frac{\ell_n}{\ell_n(\mathcal{E})}\Big(\ell_n(\mathcal{E}) \sup_{t\in\mathcal{E}\cap I_n} Z_{n,3}(t) - \ell_n^2(\mathcal{E})\Big) + \ell_n(\mathcal{E})\ell_n - \ell_n^2 + o_{\mathbb{P}}(1),$$

where

$$\ell_n(\mathcal{E})\ell_n - \ell_n^2 = \frac{\ell_n}{\ell_n + \ell_n(\mathcal{E})}(\ell_n^2(\mathcal{E}) - \ell_n^2)$$

$$= \frac{\sqrt{\log\big(\frac{(x_1-x_0)\Lambda_K}{2\pi h_n}\big)}}{\sqrt{\log\big(\frac{\lambda(\mathcal{E})\Lambda_K}{2\pi h_n}\big)} + \sqrt{\log\big(\frac{(x_1-x_0)\Lambda_K}{2\pi h_n}\big)}} 2\Big(\log\big(\tfrac{\lambda(\mathcal{E})\Lambda_K}{2\pi h_n}\big) - \log\big(\tfrac{(x_1-x_0)\Lambda_K}{2\pi h_n}\big)\Big)$$

$$\tag{A.22}$$

converges to $\log\big(\frac{\lambda(\mathcal{E})}{x_1-x_0}\big)$ since

$$\frac{\sqrt{\log\big(\frac{(x_1-x_0)\Lambda_K}{2\pi h_n}\big)}}{\sqrt{\log\big(\frac{\lambda(\mathcal{E})\Lambda_K}{2\pi h_n}\big)} + \sqrt{\log\big(\frac{(x_1-x_0)\Lambda_K}{2\pi h_n}\big)}} \to \frac{1}{2}.$$

Weak convergence of $D_{n,1}(\mathcal{E})$ to $\mathrm{Gum}_{\log\{\lambda(\mathcal{E})/(x_1-x_0)\}}$ then follows from (A.18), (A.22), and Remark A.3(ii), and this implies the assertion by (A.21). $\qquad\square$

Proof of of Proposition A.6. We only carry out the proof in the case that $\lambda(\mathcal{E}^+) > 0$ and $\lambda(\mathcal{E}^-) > 0$. For $\mathcal{A} \subset I_n$, let

$$D_n(\mathcal{A}) = \frac{\ell_n\sqrt{nh_n}}{\sigma\|K^*\|_2}\big(\sup_{t\in\mathcal{A}} |\tilde{\mu}_{h_n}(t) - g(\mu)| - d_\infty\big) - \ell_n^2 \tag{A.23}$$

and let $\mathcal{E}_n = \mathcal{E}_n^+ \cup \mathcal{E}_n^-$ as in (4.1), for some positive sequence ρ_n that converges to 0 slowly enough to guarantee that $\ell_n\sqrt{nh_n}\rho_n \to \infty$. Note that $D_n(\mathcal{A}) \leq D_n(\mathcal{B})$ for $\mathcal{A} \subset \mathcal{B}$, and that there is an $n_0 \in \mathbb{N}$ such that the sets $\mathcal{E}_n^+$ and $\mathcal{E}_n^-$ are disjoint and bounded away from each other for $n \geq n_0$. With this notation, we can rewrite

$$\mathcal{R}_n = D_n - D_n(\mathcal{E}) = \max\big\{D_n(\mathcal{E}_n) - D_n(\mathcal{E}), D_n(I_n \setminus \mathcal{E}_n) - D_n(\mathcal{E})\big\}.$$

For the second term in the maximum observe that

$$D_n(I_n \setminus \mathcal{E}_n) - D_n(\mathcal{E})$$

$$= \frac{\ell_n \sqrt{nh_n}}{\sigma\|K^*\|_2} \Big\{ \sup_{t\in I_n\backslash\mathcal{E}_n} |\tilde\mu_{h_n}(t) - g(\mu)| - \sup_{t\in\mathcal{E}} |\tilde\mu_{h_n}(t) - g(\mu)| \Big\}$$

$$\leq \frac{\ell_n \sqrt{nh_n}}{\sigma\|K^*\|_2} \Big\{ \sup_{t\in I_n\backslash\mathcal{E}_n} \Big(|\tilde\mu_{h_n}(t) - g(\mu)| - |\mu(t) - g(\mu)| \Big)$$
$$+ \sup_{t\in I_n\backslash\mathcal{E}_n} \Big(|\mu(t) - g(\mu)| - d_\infty \Big) - \sup_{t\in\mathcal{E}} \Big(|\tilde\mu_{h_n}(t) - g(\mu)| - d_\infty \Big) \Big\}$$

$$\leq \frac{\ell_n \sqrt{nh_n}}{\sigma\|K^*\|_2} \Big\{ \sup_{t\in I_n\backslash\mathcal{E}_n} |\tilde\mu_{h_n}(t) - \mu(t)| - \rho_n - \sup_{t\in\mathcal{E}} \Big(|\tilde\mu_{h_n}(t) - g(\mu)| - d_\infty \Big) \Big\}$$

$$\leq \frac{\ell_n \sqrt{nh_n}}{\sigma\|K^*\|_2} \sup_{t\in I_n} |\tilde\mu_{h_n}(t) - \mu(t)| - \ell_n^2 - \frac{\ell_n \sqrt{nh_n}}{\sigma\|K^*\|_2} \rho_n - D_n(\mathcal{E})$$

which diverges to $-\infty$ since $D_n(\mathcal{E}) \rightsquigarrow \mathrm{Gum}_{\log\{\lambda(\mathcal{E})/(x_1-x_0)\}}$ by Proposition A.5,

$$\frac{\ell_n \sqrt{nh_n}}{\sigma\|K^*\|_2} \sup_{t\in I_n} |\tilde\mu_{h_n}(t) - \mu(t)| - \ell_n^2 \rightsquigarrow \mathrm{Gum}_{\log(2)},$$

according to (A.14), and $\ell_n\sqrt{nh_n}\rho_n \to \infty$ by assumption.

As a consequence, it is sufficient to prove that $D_n(\mathcal{E}_n) - D_n(\mathcal{E}) = o_{\mathbb{P}}(1)$. Suppose we have shown that

$$\mathcal{R}_n^\pm = D_n(\mathcal{E}_n^\pm) - D_n(\mathcal{E}^\pm) = o_{\mathbb{P}}(1). \tag{A.24}$$

Then, since $D_n(\mathcal{E}_n) - D_n(\mathcal{E}) = \max\{D_n(\mathcal{E}_n^+) - D_n(\mathcal{E}), D_n(\mathcal{E}_n^-) - D_n(\mathcal{E})\} \leq \max\{\mathcal{R}_n^+, \mathcal{R}_n^-\}$, the proof of the proposition is finished.

For the proof of (A.24), we only consider $\mathcal{R}_n^+$, as $\mathcal{R}_n^-$ can be treated similarly. Similar to (A.19), it holds

$$\frac{\ell_n \sqrt{nh_n}}{\sigma\|K^*\|_2} \sup_{t\in\mathcal{E}_n^+} -\big(\tilde\mu_{h_n}(t) - g(\mu)\big) = \ell_n \sup_{t\in\mathcal{E}_n^+} \Big\{ -Z_{n,3}(t) - \frac{\sqrt{nh_n}}{\sigma\|K^*\|_2}\big(\mu(t) - g(\mu)\big) \Big\} + o_{\mathbb{P}}(1)$$

$$\leq \ell_n \sup_{t\in\mathcal{E}_n^+} \{-Z_{n,3}(t)\} - \frac{\ell_n \sqrt{nh_n}}{\sigma\|K^*\|_2}(d_\infty - \rho_n) + o_{\mathbb{P}}(1)$$

$$\overset{\mathcal{D}}{=} \ell_n \sup_{t\in\mathcal{E}_n^+} Z_{n,3}(t) - \frac{\ell_n \sqrt{nh_n}}{\sigma\|K^*\|_2}(d_\infty - \rho_n) + o_{\mathbb{P}}(1)$$

$$\leq \ell_n \sup_{t\in I_n} Z_{n,3}(t) - \ell_n^2 - \ell_n\Big\{ \frac{\sqrt{nh_n}}{\sigma\|K^*\|_2}(d_\infty - \rho_n) - \ell_n \Big\} + o_{\mathbb{P}}(1). \tag{A.25}$$

The first term on the right-hand side converges by Remark A.3(ii) to a Gumbel distribution, whereas the second term diverges to $-\infty$ since $\ell_n/\sqrt{nh_n} \to 0$ by (A.20). Hence,

$$\frac{\ell_n \sqrt{nh_n}}{\sigma\|K^*\|_2} \sup_{t\in\mathcal{E}_n^+} -\big(\tilde\mu_{h_n}(t) - g(\mu)\big) \to -\infty$$

in probability, and by monotonicity the same is true if $\mathcal{E}_n^+$ is replaced by $\mathcal{E}^+$. Thus, by the definition of $D_n(\mathcal{E}_n^+)$ in (A.23), we have

$$\mathcal{R}_n^+ = \frac{\ell_n \sqrt{nh_n}}{\sigma\|K^*\|_2} \Big\{ \sup_{t\in\mathcal{E}_n^+} |\tilde\mu_{h_n}(t) - g(\mu)| - \sup_{t\in\mathcal{E}^+} |\tilde\mu_{h_n}(t) - g(\mu)| \Big\}$$

$$= \frac{\ell_n \sqrt{nh_n}}{\sigma \|K^*\|_2} \Big\{ \sup_{t \in \mathcal{E}_n^+} \tilde{\mu}_{h_n}(t) - \sup_{t \in \mathcal{E}^+} \tilde{\mu}_{h_n}(t) \Big\} + o_{\mathbb{P}}(1).$$

Next, for $\theta \geq 0$, let $U_\theta(\mathcal{A})$ denote the θ-neighbourhood of $\mathcal{A} \subset I_n$ in I_n. Define

$$\delta_n = 2 \inf\{\theta \geq 0 \mid \mathcal{E}_n^+ \subset U_\theta(\mathcal{E}^+)\}.$$

We proceed by showing that $\delta_n = o(1)$ for $n \to \infty$. As $\mathcal{E}_n^+$ is a descending sequence of sets, the nonnegative sequence δ_n decreases and therefore converges. Suppose $\delta_n > 0$ for all n. Then, by definition, $\mathcal{E}_n^+ \subset U_{\delta_n}(\mathcal{E}^+)$ but $\mathcal{E}_n^+ \not\subset U_{\delta_n/4}(\mathcal{E}^+)$. Thus, there exists a sequence $t_n \in \mathcal{E}_n^+ \subset [0,1]$ such that $|t_n - t| \geq \delta_n/4$ for all $t \in \mathcal{E}^+$ and all $n \in \mathbb{N}$. By compactness of $[0,1]$ and continuity of μ, there is a convergent subsequence $(t_{n_k})_{k \in \mathbb{N}}$ with $\lim_{k \to \infty} t_{n_k} = t^* \in [0,1]$ and $\mu(t^*) - g(\mu) = \lim_{k \to \infty} \mu(t_{n_k}) - g(\mu) = d_\infty$, thus $t^* \in \mathcal{E}^+$. Hence, $|t_{n_k} - t^*| \geq \delta_{n_k}/4$, which implies $\delta_n = o(1)$ as asserted.

Since $\mathcal{E}_n^+ \subset U_{\delta_n}(\mathcal{E}^+) \cap I_n$ and since $\mathcal{E}^+ \cap I_n \subset \mathcal{E}^+$, the assertion $\mathcal{R}_n^+ = o_{\mathbb{P}}(1)$ and hence the proposition follows from

$$\frac{\ell_n \sqrt{nh_n}}{\sigma \|K^*\|_2} \Big\{ \sup_{t \in U_{\delta_n}(\mathcal{E}^+) \cap I_n} \tilde{\mu}_{h_n}(t) - \sup_{t \in \mathcal{E}^+ \cap I_n} \tilde{\mu}_{h_n}(t) \Big\} = o_{\mathbb{P}}(1). \tag{A.26}$$

For the proof of (A.26) observe that,

$$\frac{\sqrt{nh_n}}{\sigma \|K^*\|_2} \Big(\sup_{t \in U_{\delta_n}(\mathcal{E}^+) \cap I_n} \tilde{\mu}_{h_n}(t) - \sup_{t \in \mathcal{E}^+ \cap I_n} \tilde{\mu}_{h_n}(t) \Big) \tag{A.27}$$

$$= \max \Big\{ 0, \frac{\sqrt{nh_n}}{\sigma \|K^*\|_2} \Big(\sup_{t \in (U_{\delta_n}(\mathcal{E}^+) \setminus \mathcal{E}^+) \cap I_n} \tilde{\mu}_{h_n}(t) - \sup_{t \in \mathcal{E}^+ \cap I_n} \tilde{\mu}_{h_n}(t) \Big) \Big\}$$

$$\leq \max \Big\{ 0, \frac{\sqrt{nh_n}}{\sigma \|K^*\|_2} \Big(\sup_{t \in (U_{\delta_n}(\mathcal{E}^+) \setminus \mathcal{E}^+) \cap I_n} \{\tilde{\mu}_{h_n}(t) - \mu(t)\} - \sup_{t \in \mathcal{E}^+ \cap I_n} \{\tilde{\mu}_{h_n}(t) - \mu(t)\} \Big) \Big\}$$

$$= \max \Big\{ 0, \sup_{t \in (U_{\delta_n}(\mathcal{E}^+) \setminus \mathcal{E}^+) \cap I_n} Z_{n,3}(t) - \sup_{t \in \mathcal{E}^+ \cap I_n} Z_{n,3}(t) \Big\} + o_{\mathbb{P}}(|\log h_n|^{-1/2}),$$

by (A.15), where the process $Z_{n,3}$ is defined in (A.8).

By Remark 2.5 and since $\lambda(\mathcal{E}^+) > 0$, $\mathcal{E}^+$ can be rewritten as $\bigcup_{j=1}^m [x_{j,1}, x_{j,2}]$, for some $m \in \mathbb{N}$, where at least one interval has positive length. Without loss of generality, let $x_{j,1} < x_{j,2}$, for $1 \leq j \leq m_1$, and $x_{j,1} = x_{j,2}$, for $m_1 + 1 \leq j \leq m$. There exists an integer $n_0 \in \mathbb{N}$ such that for any $n \geq n_0$, the intervals $I_{j,1} = (x_{j,1} - \delta_n - h_n, x_{j,1} + \delta_n + h_n)$ and $I_{j,2} = (x_{j,2} - \delta_n - h_n, x_{j,2} + \delta_n + h_n)$, for $1 \leq j \leq m_1$, and $I_{j,3} = (x_{j,1} - \delta_n - h_n, x_{j,2} + \delta_n + h_n)$, for $m_1 + 1 \leq j \leq m$, are disjoint with pairwise distance greater than $2h_n$. By definition, the kernel K^* vanishes outside of the interval $[-1,1]$, hence, observing the definition of $Z_{n,3}$ in (A.8), the supremum over the set $(U_{\delta_n}(\mathcal{E}^+) \setminus \mathcal{E}^+) \cap I_n$ in (A.27) can be bounded by

$$Y_n = \max \big\{ \max_{j=1}^{m_1} R_{j,n,1}, \max_{j=1}^{m_1} R_{j,n,2}, \max_{j=m_1+1}^m R_{j,n,3} \big\}, \tag{A.28}$$

where the random variables

$$R_{j,n,k} = \sup_{t \in I_{j,k}} \frac{1}{\|K^*\|_2 \sqrt{nh_n}} \sum_{i \in \mathbb{Z}} V_i K_{h_n}^* \big(\tfrac{i}{n} - t \big)$$

are all independent (note that in each term the range of summation is given by $i \in \{\lceil nt - nh_n \rceil, \ldots, \lfloor nt + nh_n \rfloor\}$, where $t \in I_{j,k}$).

Next, consider the second supremum on the right-hand side of (A.27), which can be bounded below by restricting the supremum to $[x_{1,1}, x_{1,2}]$, where, by assumption $x_{1,1} < x_{1,2}$. Next, one may choose $\lfloor \frac{x_{1,2}-x_{1,1}}{2\delta_n+4h_n} \rfloor - 2$ disjoint intervals of length $2\delta_n + 2h_n$ with distance $2h_n$ contained in the interval $[x_{1,1}, x_{1,2}]$, that have a distance of at least $2h_n$ to the intervals $I_{1,1}$ and $I_{1,2}$. We may hence bound the second supremum in (A.27) from below by

$$Y_n' = \max_{j=1}^{\lfloor \frac{x_{1,2}-x_{1,1}}{2\delta_n+4h_n} \rfloor - 2} R_{j,n}',$$

where $R_{1,n}', R_{2,n}', \ldots$ are independent with

$$R_{j,n}' \overset{\mathcal{D}}{=} \sup_{t \in (-\delta_n - h_n, \delta_n + h_n)} \frac{1}{\|K^*\|_2 \sqrt{nh_n}} \sum_{i \in \mathbb{Z}} V_i K_{h_n}^*(\tfrac{i}{n} - t) \overset{\mathcal{D}}{=} R_{j,n,k},$$

and also independent from the random variables $R_{j,n,k}$. Using the notation $Y_{j,n}' = \max_{k=1}^{m+m_1} R_{(j-1)(m+m_1)+k,n}'$ for $j = 1, \ldots, \nu_n$, where

$$\nu_n = \lfloor \tfrac{(x_{1,2}-x_{1,1})-2(2\delta_n+4h_n)}{(m+m_1)(2\delta_n+4h_n)} \rfloor,$$

we may write $Y_n' = \max_{j=1}^{\nu_n} Y_{j,n}'$. Note that $Y_{1,n}', Y_{2,n}', \ldots$ are independent. Then, using the notation (A.28)

$$\mathbb{P}\Big(\sup_{t \in (U_{\delta_n}(\mathcal{E}^+) \backslash \mathcal{E}^+) \cap I_n} Z_{n,3}(t) > \sup_{t \in \mathcal{E}^+ \cap I_n} Z_{n,3}(t) \Big)$$

$$\leq \mathbb{P}\Big(\max \Big\{ \max_{j=1}^{m_1} R_{j,n,1}, \max_{j=1}^{m_1} R_{j,n,2}, \max_{j=m_1+1}^{m} R_{j,n,3} \Big\} > \max_{j=1}^{\lfloor \frac{x_{1,2}-x_{1,1}}{2\delta_n+4h_n} \rfloor - 2} R_{j,n}' \Big)$$

$$\leq \mathbb{P}\Big(Y_n > \max_{j=1}^{\nu_n} Y_{j,n}' \Big)$$

$$= \int_{-\infty}^{\infty} f_{Y_n}(y) \prod_{j=1}^{\nu_n} \mathbb{P}(Y_{j,n}' < y) \, dy = \mathbb{E}\Big[F_{Y_n}^{\nu_n}(Y_n) \Big] = \frac{1}{\nu_n + 1} = \mathcal{O}(\delta_n + h_n),$$

where F_{Y_n} and f_{Y_n} denote the cumulative distribution and the density function of Y_n, respectively. Here the last estimate follows since $F_{Y_n}(Y_n)$ is uniformly distributed on $[0, 1]$. Thus, the right-hand side of (A.27) converges to 0 in probability, which implies the proposition. $\square$

Proof of Corollary 3.2. We only prove the case (2) - the first case follows by similar arguments. Observe that

$$\mathbb{P}\Big(\hat{d}_{\infty,n} > (q_{0,1-\alpha} + \ell_n^2) \frac{\hat{\sigma} \|K^*\|_2}{\sqrt{nh_n}\ell_n} + \Delta \Big)$$

$$= \mathbb{P}\Big(\frac{\sqrt{nh_n}\ell_n}{\hat{\sigma}\|K^*\|_2}(\hat{d}_{\infty,n} - d_{\infty,n}) - \ell_n^2 + \frac{\sqrt{nh_n}\ell_n}{\hat{\sigma}\|K^*\|_2}(d_{\infty,n} - d_\infty + d_\infty - \Delta) > q_{0,1-\alpha} \Big).$$

In case $\lambda(\mathcal{E}) > 0$, the right-hand side converges to 0 for $d_\infty < \Delta$, to 1 for $d_\infty > \Delta$, and its limit can be bounded from above by

$$\alpha^* = \mathrm{Gum}_{\log\{\lambda(\mathcal{E})/(x_1-x_0)\}}((q_{0,1-\alpha,\infty})) \leq \mathrm{Gum}_0((q_{0,1-\alpha,\infty})) = \alpha$$

for $d_\infty = \Delta$, by Theorem 3.1 and (3.6).

For $\lambda(\mathcal{E}) = 0$, observing Assumption 2.6, (A.6) and the inverse triangle inequality, we have

$$
|\hat{d}_{\infty,n} - d_{\infty,n}| = \left| \sup_{t \in I_n} |\tilde{\mu}_{h_n}(t) - \hat{g}_n| - \sup_{t \in I_n} |\mu(t) - g(\mu)| \right|
$$

$$
\leq \sup_{t \in I_n} |\tilde{\mu}_{h_n}(t) - \mu(t)| + o_\mathbb{P}(1) = \sup_{t \in I_n} \left| \frac{1}{nh_n} \sum_{i=1}^n \sigma\left(\tfrac{i}{n}\right) V_i K_{h_n}^*\left(\tfrac{i}{n} - t\right) \right| + o_\mathbb{P}(1).
$$

(A.29)

By the same arguments that led to (A.9), the right-hand side of the previous display is of order $o_\mathbb{P}(1)$, thus, $|\hat{d}_{\infty,n} - d_{\infty,n}| = o_\mathbb{P}(1)$. Now, under the alternative,

$$
p_n \equiv \mathbb{P}\left(\hat{d}_{\infty,n} > (q_{0,1-\alpha} + \ell_n^2) \frac{\hat{\sigma}\|K^*\|_2}{\sqrt{nh_n}\ell_n} + \Delta \right)
$$

$$
= \mathbb{P}\left(\frac{\sqrt{nh_n}\ell_n}{\hat{\sigma}\|K^*\|_2} \left\{ \hat{d}_{\infty,n} - d_{\infty,n} + d_{\infty,n} - d_\infty + d_\infty - \Delta - \frac{\ell_n \hat{\sigma}\|K^*\|_2}{\sqrt{nh_n}} \right\} > q_{0,1-\alpha} \right),
$$

converges to 1 since $(\hat{d}_{\infty,n} - d_{\infty,n}), (d_{\infty,n} - d_\infty)$ and $\ell_n(nh_n)^{-1/2}$ vanish as n tends to infinity, while $d_\infty - \Delta > 0$. The assertion regarding the null hypothesis follows from Theorem 3.1(3). $\qquad \square$

A.3. **Proofs for Section 4**

We first prove an auxiliary result, which will be used in the proof of Theorem 4.1.

Proposition A.7. *Let the assumptions of Theorem 4.1 and Assumption 2.4 be satisfied. Then,*

$$
\frac{\lambda(\mathcal{E}_{n,1}^\pm)}{\lambda(\mathcal{E}_{n,2}^\pm)} \to 1 \qquad \text{if } \mathcal{E}^\pm \neq \emptyset,
$$

where $\mathcal{E}_{n,1}^\pm = \mathcal{E}_n^\pm$ are defined in (4.1) and where

$$
\mathcal{E}_{n,2}^\pm = \left\{ t \in I_n : d_{\infty,n} \mp d(t) \leq (1 + 2e_n)\rho_n \right\},
$$

(A.30)

with e_n an arbitrary positive sequence converging to 0.

Proof. We only consider the assertion for $\mathcal{E}^+ \neq \emptyset$. As both the numerator and the denominator converge from above to $\lambda(\mathcal{E}^+)$, the convergence is trivial for $\lambda(\mathcal{E}^+) > 0$. In the case $\lambda(\mathcal{E}^+) = 0$, by Assumption 2.4, there exists a positive constant $\gamma > 0$ such that the function $d = \mu - g(\mu)$ is concave on $U_\gamma(\mathcal{E}^+)$. Let us show that $\mathcal{E}_{n,2}^+ \subset U_\gamma(\mathcal{E}^+)$ for all sufficiently large n. First, by continuity, d attains its maximum on the compact set $[0,1] \setminus U_\gamma(\mathcal{E}^+)$, say at a point t'. As $t' \notin \mathcal{E}^+$, $d(t') < d_\infty$. Next, $\sup_{t \in \mathcal{E}_{n,2}^+} d_\infty - d(t) = (1 + 2e_n)\rho_n$ by definition of $\mathcal{E}_{n,2}^+$ and continuity of d, which converges to 0 for $n \to \infty$. Thus, there is a natural number $n_0 \in \mathbb{N}$ such that for any $n \geq n_0$ and any $t \in \mathcal{E}_{n,2}^+$, it holds $d(t) > d(t')$, and hence $t \notin [0,1] \setminus U_\gamma(\mathcal{E}^+)$. Therefore, it follows $\mathcal{E}_{n,2}^+ \subset U_\gamma(\mathcal{E}^+)$.

Since $\lambda(\mathcal{E}^+) = 0$, we can use part (ii) of Remark 2.5 and rewrite $\mathcal{E}^+$ as $\{t_1, \ldots, t_m\}$, for some $m \in \mathbb{N}$. By Assumption 2.4, $\mu - g(\mu)$ is strictly increasing on $U_\gamma^<(t_i) := (t_i - \gamma, t_i]$ and strictly decreasing on $U_\gamma^>(t_i) := [t_i, t_i + \gamma)$, for any $t_i \in \mathcal{E}^+$. Thus, for

sufficiently large n, there are points $x_{i,\sim,1} > x_{i,\sim,2} \in U_\gamma^\sim(t_i)$ (depending on n) with $\mu(x_{i,\sim,1}) - g(\mu) = d_\infty - \rho_n$ and $\mu(x_{i,\sim,2}) - g(\mu) = d_\infty - (1+2e_n)\rho_n$, where the symbol $\sim$ denotes either '<' or '>'. Moreover, these numbers are unique. As the function μ is concave on the interval $[x_{i,<,2}, t_i]$, we have

$$\frac{\rho_n}{t_i - x_{i,<,1}} = \frac{\mu(t_i) - \mu(x_{i,<,1})}{t_i - x_{i,<,1}} \leq \frac{\mu(x_{i,<,1}) - \mu(x_{i,<,2})}{x_{i,<,1} - x_{i,<,2}} = \frac{2e_n\rho_n}{x_{i,<,1} - x_{i,<,2}}.$$

Thus, $x_{i,<,1} - x_{i,<,2} \leq 2e_n(t_i - x_{i,<,1})$ and analogously $x_{i,>,2} - x_{i,>,1} \leq 2e_n(x_{i,>,1} - t_i)$. Since $\mathcal{E}_{n,k}^+ = \bigcup_{i=1}^m (x_{i,<,k}, t_i] \cup (t_i, x_{i,>,k})$, for $k = 1, 2$, it follows that

$$\frac{\lambda(\mathcal{E}_{n,2}^+ \setminus \mathcal{E}_{n,1}^+)}{\lambda(\mathcal{E}_{n,1}^+)} = \sum_{i=1}^m \frac{x_{i,<,1} - x_{i,<,2} + x_{i,>,2} - x_{i,>,1}}{\lambda(\mathcal{E}_{n,1}^+)}$$

$$\leq \sum_{i=1}^m 2e_n \frac{t_i - x_{i,<,1} + x_{i,>,1} - t_i}{\lambda(\mathcal{E}_{n,1}^+)} = 2e_n = o(1),$$

which implies also $\lambda(\mathcal{E}_{n,2}^+)/\lambda(\mathcal{E}_{n,1}^+) \to 1$. □

Proof of Theorem 4.1. First, consider the case $d_\infty > 0$. Choose a positive sequence e_n converging to 0 such that $\liminf_{n\to\infty} e_n\rho_n\sqrt{nh_n}/2 - \ell_n > 0$. We prove that

$$\frac{\lambda(\mathcal{E}_n)}{\lambda(\hat{\mathcal{E}}_{n,2})} = 1 + o_\mathbb{P}(1), \tag{A.31}$$

where $\hat{\mathcal{E}}_{n,2} = \hat{\mathcal{E}}_{n,2}^+ \cup \hat{\mathcal{E}}_{n,2}^-$ and

$$\hat{\mathcal{E}}_{n,2}^\pm = \{t \in I_n : \hat{d}_{\infty,n} \mp \hat{d}_n(t) \leq (1 + e_n)\rho_n\}.$$

By the same arguments one may show that $\lambda(\mathcal{E}_n)/\lambda(\hat{\mathcal{E}}_{n,3}) = 1 + o_\mathbb{P}(1)$, where $\hat{\mathcal{E}}_{n,3}$ is defined analogously to $\hat{\mathcal{E}}_{n,2}$ with $(1+e_n)\rho_n$ replaced by $(1-e_n)\rho_n$. As $\hat{\mathcal{E}}_{n,3} \subset \hat{\mathcal{E}}_n \subset \hat{\mathcal{E}}_{n,2}$ this leads to $\lambda(\mathcal{E}_n)/\lambda(\hat{\mathcal{E}}_n) = 1 + o_\mathbb{P}(1)$ by the sandwich theorem and proves the assertion of the theorem.

Without loss of generality, we assume that $\mathcal{E}^+ \neq \emptyset$ (which implies $\lambda(\mathcal{E}_n^+) > 0$) and observe that

$$2\sup_{t \in I_n} |\hat{d}_n(t) - d(t)| \geq |\hat{d}_{\infty,n} - d_{\infty,n}| + \sup_{t \in I_n} |\hat{d}_n(t) - d(t)|$$

$$\geq |\hat{d}_{\infty,n} - d_{\infty,n}| + \sup_{t \in \mathcal{E}_n^+} |\hat{d}_n(t) - d(t)|$$

$$\geq |\hat{d}_{\infty,n} - d_{\infty,n}| + \sup_{t \in \mathcal{E}_n^+} |\hat{d}_n(t) - d_{\infty,n}| - \rho_n$$

$$\geq \sup_{t \in \mathcal{E}_n^+} |\hat{d}_{\infty,n} - \hat{d}_n(t)| - \rho_n. \tag{A.32}$$

Thus

$$\mathbb{P}(\mathcal{E}_n^+ \subset \hat{\mathcal{E}}_{n,2}^+) = \mathbb{P}\Big(\sup_{t \in \mathcal{E}_n^+} \{\hat{d}_{\infty,n} - \hat{d}_n(t)\} \leq (1 + e_n)\rho_n \Big)$$

$$= \mathbb{P}\Big(\sup_{t \in \mathcal{E}_n^+} |\hat{d}_{\infty,n} - \hat{d}_n(t)| - \rho_n \leq e_n\rho_n \Big)$$

$$\geq \mathbb{P}\Big(2 \sup_{t \in I_n} |\hat{d}_n(t) - d(t)| \leq e_n \rho_n\Big)$$

$$= \mathbb{P}\Big(\sqrt{nh_n}\ell_n \sup_{t \in I_n} |\hat{d}_n(t) - d(t)| - \ell_n^2 \leq e_n \rho_n \ell_n \sqrt{nh_n}/2 - \ell_n^2\Big) \quad \text{(A.33)}$$

which converges to unity by Remark A.3(ii), by Assumption 2.6 and since $e_n \rho_n \ell_n \sqrt{nh_n}/2 - \ell_n^2 \to \infty$ by the choice of e_n. Note that

$$\sup_{t \in \hat{\mathcal{E}}_{n,2}^+} |\hat{d}_{\infty,n} - \hat{d}_n(t)| = (1 + e_n)\rho_n,$$

thus, analogously to (A.32),

$$\sup_{t \in \hat{\mathcal{E}}_{n,2}^+} |d_{\infty,n} - d(t)| - (1 + e_n)\rho_n$$

$$\leq |d_{\infty,n} - \hat{d}_{\infty,n}| + \sup_{t \in \hat{\mathcal{E}}_{n,2}^+} |\hat{d}_{\infty,n} - \hat{d}_n(t)| + \sup_{t \in \hat{\mathcal{E}}_{n,2}^+} |\hat{d}_n(t) - d(t)| - (1 + e_n)\rho_n$$

$$\leq 2 \sup_{t \in I_n} |\hat{d}_n(t) - d(t)|.$$

Therefore, as in (A.33), $\mathbb{P}(\hat{\mathcal{E}}_{n,2}^+ \subset \mathcal{E}_{n,2}^+) \to 1$, where the set $\mathcal{E}_{n,2}^+$ is defined in (A.30). In particular,

$$\mathcal{E}_n^+ \subset \hat{\mathcal{E}}_{n,2}^+ \subset \mathcal{E}_{n,2}^+,$$

with probability converging to one. Hence,

$$0 < \lambda(\mathcal{E}_n^+) \leq \lambda(\hat{\mathcal{E}}_{n,2}^+) \leq \lambda(\mathcal{E}_{n,2}^+), \quad \text{(A.34)}$$

with probability converging to one and

$$\mathbb{P}\Big(1 \leq \frac{\lambda(\hat{\mathcal{E}}_{n,2}^+)}{\lambda(\mathcal{E}_n^+)} \leq z_n\Big) \xrightarrow{n \to \infty} 1,$$

where the deterministic sequence $z_n = \lambda(\mathcal{E}_{n,2}^+)/\lambda(\mathcal{E}_n^+)$ converges to 1, by Proposition A.7. Let A_n denote the event that (A.34) holds. Then, $\mathbb{P}(A_n^C) \to 0$ and, for any $\varepsilon > 0$,

$$\mathbb{P}\Big(\Big|\frac{\lambda(\hat{\mathcal{E}}_{n,2}^+)}{\lambda(\mathcal{E}_n^+)} - 1\Big| \geq \varepsilon\Big) = \mathbb{P}\Big(\Big|\frac{\lambda(\hat{\mathcal{E}}_{n,2}^+)}{\lambda(\mathcal{E}_n^+)} - 1\Big| \geq \varepsilon \cap A_n\Big) + \mathbb{P}\Big(\Big|\frac{\lambda(\hat{\mathcal{E}}_{n,2}^+)}{\lambda(\mathcal{E}_n^+)} - 1\Big| \geq \varepsilon \cap A_n^C\Big)$$

$$\leq \mathbb{P}(|z_n - 1| \geq \varepsilon) + \mathbb{P}(A_n^C) \to 0.$$

Therefore, it follows

$$\frac{\lambda(\mathcal{E}_n^+)}{\lambda(\hat{\mathcal{E}}_{n,2}^+)} = 1 + o_{\mathbb{P}}(1).$$

If $\mathcal{E}^- \neq \emptyset$, it holds analogously

$$\frac{\lambda(\mathcal{E}_n^-)}{\lambda(\hat{\mathcal{E}}_{n,2}^-)} = 1 + o_{\mathbb{P}}(1).$$

Note that, since $d_\infty > 0$, $\hat{\mathcal{E}}_{n,2}^+$ and $\hat{\mathcal{E}}_{n,2}^-$ with probability converging to one. Thus, the previous displays imply

$$\frac{\lambda(\mathcal{E}_n)}{\lambda(\hat{\mathcal{E}}_{n,2})} = \frac{\lambda(\mathcal{E}_n^+)}{\lambda(\hat{\mathcal{E}}_{n,2}^+)} \frac{\lambda(\hat{\mathcal{E}}_{n,2}^+)}{\lambda(\hat{\mathcal{E}}_{n,2})} + \frac{\lambda(\mathcal{E}_n^-)}{\lambda(\hat{\mathcal{E}}_{n,2}^-)} \frac{\lambda(\hat{\mathcal{E}}_{n,2}^-)}{\lambda(\hat{\mathcal{E}}_{n,2})}$$

$$= \{1 + o_{\mathbb{P}}(1)\}\Big\{\frac{\lambda(\hat{\mathcal{E}}_{n,2}^+)}{\lambda(\hat{\mathcal{E}}_{n,2})} + \frac{\lambda(\hat{\mathcal{E}}_{n,2}^-)}{\lambda(\hat{\mathcal{E}}_{n,2})}\Big\} = 1 + o_{\mathbb{P}}(1).$$

Conversely, if $\mathcal{E}^- = \emptyset$, both $\mathcal{E}_n^-$ and $\mathcal{E}_{n,2}^-$ are empty for almost every $n \in \mathbb{N}$. Further, $\mathbb{P}(\hat{\mathcal{E}}_{n,2}^- \subset \mathcal{E}_{n,2}^-) \to 1$, which implies

$$\frac{\lambda(\hat{\mathcal{E}}_{n,2})}{\lambda(\mathcal{E}_n)} = \frac{\lambda(\hat{\mathcal{E}}_{n,2}^+)}{\lambda(\mathcal{E}_n^+)} + \frac{\lambda(\hat{\mathcal{E}}_{n,2}^-)}{\lambda(\mathcal{E}_n^+)} - \frac{\lambda(\hat{\mathcal{E}}_{n,2}^+ \cap \hat{\mathcal{E}}_{n,2}^-)}{\lambda(\mathcal{E}_n^+)} = 1 + o_{\mathbb{P}}(1),$$

and this is equivalent to (A.31).

Finally, if $d_\infty = 0$, observe that $\mathcal{E}_n = I_n$ and

$$\hat{\mathcal{E}}_{n,2} = \Big\{t \in I_n : \Big|\sup_{s \in I_n} |\hat{d}_{\infty,n}(s) - d(s)| - (\hat{d}_n(t) - d(t))\Big| \leq (1 + e_n)\rho_n\Big\}.$$

By definition $\hat{\mathcal{E}}_{n,2} \subset I_n$, and further

$$\mathbb{P}(I_n \subset \hat{\mathcal{E}}_{n,2}) \geq \mathbb{P}(2\sup_{t \in I_n} |\hat{d}_n(t) - d(t)| \leq (1 + e_n)\rho_n),$$

hence, the theorem's assertion follows by similar arguments as in the case $d_\infty > 0$. $\square$

The proof of Theorem 4.2 will be based on the following auxiliary result.

Proposition A.8. *Let Assumptions 2.2 and 2.4 be satisfied, $\rho_n \to 0$ and $h_n \to 0$. Additionally, if $\lambda(\mathcal{E}) = 0$ assume that $\rho_n h_n^{-2} \to \infty$. Then, $(h_n^{-1}\mathcal{E}_n)_n$ satisfies the blowing up property from Piterbarg (2012):*
 (i) $h_n^{-1}\lambda(\mathcal{E}_n) = \lambda(h_n^{-1}\mathcal{E}_n) \to \infty$.
 (ii) For any $R > 1$,

$$\lambda(U_R(h_n^{-1}\mathcal{E}_n) \setminus (h_n^{-1}\mathcal{E}_n)) \leq 2R|\partial(h_n^{-1}\mathcal{E}_n)|,$$

 where $|\mathcal{A}|$ and $\partial\mathcal{A}$ denote the cardinality and the boundary of a set $\mathcal{A}$.
 (iii) For any $\alpha \in (0,1)$, there exists a constant $L = L_\alpha > 0$ such that $|\partial(h_n^{-1}\mathcal{E}_n)| \leq L(\lambda(h_n^{-1}\mathcal{E}_n))^\alpha$.

Proof. (i) If $\lambda(\mathcal{E}) > 0$, then $\lambda(h_n^{-1}\mathcal{E}_n) \geq h_n^{-1}\lambda(\mathcal{E}) \to \infty$. If $\lambda(\mathcal{E}) = 0$, we assume without loss of generality that $\mathcal{E}^+ \neq \emptyset$ and define $c_n = (\frac{\rho_n}{\|\mu''\|_\infty})^{1/2}$. Then for $t^* \in \mathcal{E}^+$ and $t \in U_{c_n}(t^*)$, it follows

$$d_\infty - d(t) = \mu(t^*) - \mu(t) \leq \|\mu''\|_\infty|t^* - t|^2 \leq \rho_n,$$

by Taylor's theorem and $\mu'(t^*) = 0$. Thus, $U_{c_n}(\mathcal{E}^+) \cap I_n \subset \mathcal{E}_n^+$. In particular, $\lambda(\mathcal{E}_n^+) \geq \lambda(U_{c_n}(\mathcal{E}^+) \cap I_n) \geq 2(c_n - h_n)$. The assertion follows from $\sqrt{\rho_n}h_n^{-1} \to \infty$.

(ii) For $t \in U_R(h_n^{-1}\mathcal{E}_n) \setminus (h_n^{-1}\mathcal{E}_n)$, there is an $s \in h_n^{-1}\mathcal{E}_n$ with distance $|s - t| < R$. Therefore, there is an intermediate value $u \in [s \wedge t, s \vee t]$ in $\partial(h_n^{-1}\mathcal{E}_n)$ such that $t \in U_R(u)$. Thus,

$$\lambda\Big(U_R(h_n^{-1}\mathcal{E}_n) \setminus h_n^{-1}\mathcal{E}_n\Big) \leq \lambda\Big(\bigcup_{u \in \partial(h_n^{-1}\mathcal{E}_n)} U_R(u)\Big) \leq 2R|\partial(h_n^{-1}\mathcal{E}_n)|,$$

and (ii) follows.

(iii) By Remark 2.5 (ii), the set of extremal points can be represented as $\mathcal{E} = \bigcup_{i=1}^m [x_{i,1}, x_{i,2}]$, for some $m \in \mathbb{N}$. Further, there exists an integer $n_0 \in \mathbb{N}$, such that for any $n \geq n_0$, $\mathcal{E}_n \subset U_\gamma(\mathcal{E})$, with γ from Assumption 2.4. By concavity of μ on $U_\gamma([x_{i,1}, x_{i,2}])$, for $[x_{i,1}, x_{i,2}] \subset \mathcal{E}^+$ and convexity on $U_\gamma([x_{i,1}, x_{i,2}])$, for $[x_{i,1}, x_{i,2}] \subset \mathcal{E}^-$, there are real numbers $x_{i,1,n} \leq x_{i,1} \leq x_{i,2} \leq x_{i,2,n}$ such that $\mathcal{E}_n = \bigcup_{i=1}^m [x_{i,1,n}, x_{i,2,n}]$. Hence, $|\partial(h_n^{-1}\mathcal{E}_n)| \leq 2m$, which implies the assertion by (i). $\square$

Proof of Theorem 4.2. If $d_\infty = 0$, it follows that $\mathcal{E}_n = I_n$ and further $\ell_n(\mathcal{E}_n) = \ell_n$. Thus, the theorem's statement follows from the beginning of the proof of Theorem 3.1(2).

Now, let $d_\infty > 0$. First observe that with the notation

$$D_n^{\mathcal{E}}(\mathcal{A}) = \ell_n(\mathcal{E}_n)\frac{\sqrt{nh_n}}{\sigma\|K^*\|_2}\Big\{\sup_{t \in \mathcal{A}} |\hat{d}_n(t)| - d_{\infty,n}\Big\} - \ell_n^2(\mathcal{E}_n),$$

for $\mathcal{A} \subset I_n$, we have

$$D_n^{\mathcal{E}}(I_n) = \tfrac{\ell_n(\mathcal{E}_n)\sqrt{nh_n}}{\sigma\|K^*\|_2}(\hat{d}_{\infty,n} - d_{\infty,n}) - \ell_n^2(\mathcal{E}_n) = D_n^{\mathcal{E}}(I_n \setminus \mathcal{E}_n) \vee D_n^{\mathcal{E}}(\mathcal{E}_n).$$

Let us first consider the case $\mathcal{E}_n \neq I_n$. Then, by the definition of $\mathcal{E}_n = \mathcal{E}_n^+ \cup \mathcal{E}_n^-$ in (4.1),

$$D_n^{\mathcal{E}}\big(I_n \setminus \mathcal{E}_n\big)$$

$$\leq \ell_n(\mathcal{E}_n)\frac{\sqrt{nh_n}}{\sigma\|K^*\|_2}\Big\{\sup_{t \in I_n \setminus \mathcal{E}_n} \{|\hat{d}_n(t) - d(t)| + |d(t)|\} - d_{\infty,n}\Big\} - \ell_n^2(\mathcal{E}_n)$$

$$\leq \ell_n(\mathcal{E}_n)\frac{\sqrt{nh_n}}{\sigma\|K^*\|_2}\Big\{\sup_{t \in I_n} |\hat{d}_n(t) - d(t)| - \rho_n\Big\} - \ell_n^2(\mathcal{E}_n)$$

$$= \frac{\ell_n(\mathcal{E}_n)}{\ell_n}\Big(\ell_n\frac{\sqrt{nh_n}}{\sigma\|K^*\|_2}\sup_{t \in I_n} |\hat{d}_n(t) - d(t)| - \ell_n^2\Big) - \ell_n(\mathcal{E}_n)\Big(\frac{\sqrt{nh_n}}{\sigma\|K^*\|_2}\rho_n - \ell_n + \ell_n(\mathcal{E}_n)\Big),$$

which diverges to $-\infty$ by Assumption 2.6 and Remark A.3(ii), and since $\sqrt{nh_n}\rho_n/\ell_n \to \infty$ by assumption and $\ell_n(\mathcal{E}_n) \to \infty$ by Proposition A.8(i). Thus,

$$D_n^{\mathcal{E}}(I_n) = D_n^{\mathcal{E}}(\mathcal{E}_n) + o_{\mathbb{P}}(1),$$

and the same assertion is obviously true if $\mathcal{E}_n = I_n$.

If $\mathcal{E}^+ \neq \emptyset$, recall the definition of $Z_{n,3}$ in (A.8) and observe that analogously to (A.25), and by (A.15) and Assumption 2.6,

$$D_n^{\mathcal{E}}(\mathcal{E}_n^+) = \ell_n(\mathcal{E}_n)\frac{\sqrt{nh_n}}{\sigma\|K^*\|_2}\Big\{\sup_{t \in \mathcal{E}_n^+} |\hat{d}_n(t)| - d_{\infty,n}\Big\} - \ell_n^2(\mathcal{E}_n)$$

$$= \ell_n(\mathcal{E}_n) \frac{\sqrt{nh_n}}{\sigma\|K^*\|_2} \Big\{ \sup_{t\in\mathcal{E}_n^+} \hat{d}_n(t) - d_{\infty,n} \Big\} - \ell_n^2(\mathcal{E}_n) + o_{\mathbb{P}}(1)$$

$$\leq \ell_n(\mathcal{E}_n) \frac{\sqrt{nh_n}}{\sigma\|K^*\|_2} \sup_{t\in\mathcal{E}_n^+} \{\hat{d}_n(t) - d(t)\} - \ell_n^2(\mathcal{E}_n) + o_{\mathbb{P}}(1)$$

$$= \ell_n(\mathcal{E}_n) \sup_{t\in\mathcal{E}_n^+} Z_{n,3}(t) - \ell_n^2(\mathcal{E}_n) + o_{\mathbb{P}}(1).$$

Analogously, if $\mathcal{E}^- \neq \emptyset$, we have

$$D_n^{\mathcal{E}}(\mathcal{E}_n^-) \leq \ell_n(\mathcal{E}_n) \sup_{t\in\mathcal{E}_n^-} -Z_{n,3}(t) - \ell_n^2(\mathcal{E}_n) + o_{\mathbb{P}}(1).$$

Thus, since

$$(\sup_{t\in\mathcal{E}_n^+} Z_{n,3}(t), \sup_{t\in\mathcal{E}_n^-} -Z_{n,3}(t)) \stackrel{\mathcal{D}}{=} (\sup_{t\in\mathcal{E}_n^+} Z_{n,3}(t), \sup_{t\in\mathcal{E}_n^-} Z_{n,3}(t))$$

as $\mathcal{E}_n^+$ and $\mathcal{E}_n^-$ are disjoint with distance larger than h_n, we have

$$D_n^{\mathcal{E}}(I_n) = D_n^{\mathcal{E}}(\mathcal{E}_n) + o_{\mathbb{P}}(1) = D_n^{\mathcal{E}}(\mathcal{E}_n^-) \vee D_n^{\mathcal{E}}(\mathcal{E}_n^+) + o_{\mathbb{P}}(1)$$

$$\ll_S \ell_n(\mathcal{E}_n) \sup_{t\in\mathcal{E}_n} Z_{n,3}(t) - \ell_n^2(\mathcal{E}_n) + o_{\mathbb{P}}(1) = G_{n,1} + o_{\mathbb{P}}(1).$$

By Proposition A.8, $h_n^{-1}\mathcal{E}_n$ satisfies the blowing up property. Thus, by Remark A.3(i) and (ii), we obtain that $G_{n,1} \rightsquigarrow \mathrm{Gum}_0$ as asserted. The fact that we may replace $\ll_S$ by $\stackrel{\mathcal{D}}{=}$ in the previous display if $\lambda(\mathcal{E}) > 0$ follows by the same arguments as given in the proof of Theorem 3.1. □

Proof of Corollary 4.3. We only consider the case $\Delta > 0$ (the proof $\Delta = 0$ follows by similar arguments). Recall that $\hat{\sigma}^2 - \sigma^2 = \mathcal{O}_{\mathbb{P}}(n^{-1/3})$ by (3.6). By Theorem 4.1 and similar calculations as in (A.22) we obtain

$$\frac{\ell_n(\hat{\mathcal{E}}_n)}{\ell_n(\mathcal{E}_n)} \stackrel{\mathbb{P}}{\longrightarrow} 1 \quad \text{and} \quad \ell_n(\hat{\mathcal{E}}_n)\ell_n(\mathcal{E}_n) - \ell_n^2(\hat{\mathcal{E}}_n) \stackrel{\mathbb{P}}{\longrightarrow} 0.$$

Together with Theorem 4.2 this yields

$$\liminf_{n\to\infty} \mathbb{P}\Big(\frac{\ell_n(\hat{\mathcal{E}}_n)\sqrt{nh_n}}{\hat{\sigma}\|K^*\|_2}(\hat{d}_{\infty,n} - d_{\infty,n}) - \ell_n^2(\hat{\mathcal{E}}_n) \leq x \Big) \geq \begin{cases} \mathrm{Gum}_0((-\infty,x]) & , d_\infty > 0, \\ \mathrm{Gum}_{\log 2}((-\infty,x]) & , d_\infty = 0. \end{cases}$$

Moreover, we have $|\hat{d}_{\infty,n} - d_{\infty,n}| = o_{\mathbb{P}}(1)$, which follows from Theorem 3.1 for $\lambda(\mathcal{E}) > 0$ and from (A.29) for $\lambda(\mathcal{E}) = 0$. The derived convergences imply the corollary by similar arguments as in the proof of Corollary 3.2. □

A.4. **Proofs for Section 5**

Proof of Theorem 5.1. Define $\gamma_n = \big(2\sigma\|K^*\|_2 \frac{\ell_n}{\sqrt{nh_n}} + \delta_n\big)^{1/\kappa}$ and let $c_1 = (2/c_\kappa)^{1/\kappa}$, where κ and c_κ are as in (5.1). First, consider the case $t^* = \infty$. Then $\Delta - d_{\infty,n} \geq \Delta - d_\infty > 0$ for all $n \in \mathbb{N}$. Hence, since $|\hat{d}_{\infty,n} - d_{\infty,n}| = o_{\mathbb{P}}(1)$ (this follows from Theorem 3.1 for $\lambda(\mathcal{E}) > 0$ and from (A.29) for $\lambda(\mathcal{E}) = 0$), we obtain

$$\mathbb{P}(\hat{t}^* < t^*) = \mathbb{P}(\hat{d}_{\infty,n} \geq \Delta - \delta_n) = \mathbb{P}(\hat{d}_{\infty,n} - d_{\infty,n} + \delta_n \geq \Delta - d_{\infty,n}) = o_{\mathbb{P}}(1).$$

Next, consider the case $t^* < \infty$. The assertion follows from

$$\mathbb{P}(\hat{t}^* - t^* < -c_1 \gamma_n) = o(1), \tag{A.35}$$

$$\mathbb{P}(\hat{t}^* - t^* > h_n) = o(1). \tag{A.36}$$

For the proof of (A.35) note that it follows from (5.1) that $|d(t^*) - d(s)| - c_\kappa(t^* - s)^\kappa = o(|t^* - s|^\kappa)$. A careful case-by-case study of the absolute value $|d(t^*) - d(t^* - c_1 \gamma_n)|$, depending on whether $d(t^*) = \Delta$ or $d(t^*) = -\Delta$, then implies

$$|d(t^* - c_1 \gamma_n)| = \Delta - c_\kappa(c_1 \gamma_n)^\kappa + o(\gamma_n^\kappa) = \Delta - 2\gamma_n^\kappa + o(\gamma_n^\kappa).,$$

as $\gamma_n \to 0$ by assumption. Thus, by continuity of d, compactness of $[h_n \vee x_0, t^* - c_1 \gamma_n]$ and Remark 2.5, it follows for all sufficiently large n that

$$\max_{s \in [h_n \vee x_0, t^* - c_1 \gamma_n]} |d(s)| = |d(t^* - c_1 \gamma_n)| \leq \Delta - \gamma_n^\kappa.$$

This implies, by the definition of $\hat{t}^*$, Assumption 2.6 and Remark A.3 (ii),

$$\mathbb{P}(\hat{t}^* < t^* - c_1 \gamma_n)$$

$$\leq \mathbb{P}(|\hat{d}_n(s)| \geq \Delta - \delta_n \text{ for some } s \in [h_n \vee x_0, t^* - c_1 \gamma_n))$$

$$\leq \mathbb{P}(|d(s)| + |\hat{d}_n(s) - d(s)| \geq \Delta - \delta_n \text{ for some } s \in [h_n \vee x_0, t^* - c_1 \gamma_n))$$

$$\leq \mathbb{P}\Big(|\hat{d}_n(s) - d(s)| - \gamma_n^\kappa \geq -\delta_n \text{ for some } s \in [h_n \vee x_0, t^* - c_1 \gamma_n)\Big)$$

$$\leq \mathbb{P}\Big(\frac{\ell_n \sqrt{nh_n}}{\sigma \|K^*\|_2} \sup_{s \in I_n} |\hat{d}_n(s) - d(s)| - \ell_n^2 \geq \ell_n\Big(\frac{\sqrt{nh_n}(\gamma_n^\kappa - \delta_n)}{\sigma \|K^*\|_2} - \ell_n\Big)\Big) \to 0.$$

For (A.36), observe that similarly,

$$\mathbb{P}(\hat{t}^* > t^* + h_n)$$

$$\leq \mathbb{P}\Big(\max_{t \in [h_n \vee x_0, s]} |\hat{d}_n(t)| < \Delta - \delta_n \text{ for some } s \geq t^*\Big)$$

$$\leq \mathbb{P}\Big(\max_{t \in [h_n \vee x_0, s]} |d(t)| - \max_{t \in [h_n \vee x_0, s]} |\hat{d}_n(t) - d(t)| < \Delta - \delta_n \text{ for some } s \geq t^*\Big)$$

$$\leq \mathbb{P}\Big(\ell_n\big(\frac{\sqrt{nh_n}\delta_n}{\sigma \|K^*\|_2} - \ell_n\big) < \frac{\ell_n \sqrt{nh_n}}{\sigma \|K^*\|_2} \sup_{t \in I_n} |\hat{d}_n(t) - d(t)| - \ell_n^2\Big),$$

which vanishes as the left-hand side in the latter probability diverges to infinity by assumption and the right-hand side is bounded in probability by Remark A.3(ii) and Assumption 2.6. □

A.5. **Proofs for Section 6**

Proof of Theorem 6.4. The convergence rate in the theorem is an improved version of Theorem 4.4 of Dette and Wu (2019) and for the sake of brevity we only state the main idea of the proof. Observe that equation (1.1) in Fan (2004) generalises Burkholder's inequality from quadratic variation to the general case of p-variation. Using this generalised version of Burkholder's inequality, the inequalities in Theorem 1 of Wu (2007) are valid not only for $q' = \min(2, q)$, but for q itself. Under Assumption 6.1, Theorem 1 of Wu (2007) can now be applied with $q = 4$. Thus, the convergence rate in Lemma 3 of Zhou and Wu (2010) improves to $\mathcal{O}(m^{1/4}(n\tau_n)^{-1/2})$, which finally

allows us to derive the improved convergence of Theorem 4.4 of Dette and Wu (2019) as stated in the theorem. □

Proof of Theorem 6.5. The proof follows by same arguments as given in Section A.3, where one uses Theorem 6.4 instead of (3.6) and Theorem A.1 instead of Remark A.3. The details are omitted for the sake of brevity. □

Acknowledgements

References

Aitchison, J. (1964). Confidence-region tests. *Journal of the Royal Statistical Society, Ser. B 26*, 462–476.

Aston, J. A. D. and C. Kirch (2012). Evaluating stationarity via change-point alternatives with applications to fmri data. *Ann. Appl. Stat. 6*(4), 1906–1948.

Aue, A. and L. Horváth (2013). Structural breaks in time series. *Journal of Time Series Analysis 34*(1), 1–16.

Baranowski, R., Y. Chen, and P. Fryzlewicz (2019). Narrowest-over-threshold detection of multiple change-points and change-point-like features. *Journal of the Royal Statistical Society, Ser. B 81*, 649–672.

Cao, H., W. Liu, and Z. Zhou (2018). Simultaneous nonparametric regression analysis of sparse longitudinal data. *Bernoulli 24*(4A), 3013–3038.

Chakraborti, S. and M. A. Graham (2019). Nonparametric (distribution-free) control charts: An updated overview and some results. *Quality Engineering 31*(4), 523–544.

Champ, C. W. and W. H. Woodall (1987). Exact results for Shewhart control charts with supplementary runs rules. *Technometrics 29*(4), 393–399.

Cho, H. and P. Fryzlewicz (2015). Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 77*(2), 475–507.

Collins, D., P. Della-Marta, N. Plummer, and B. Trewin (2000). Trends in annual frequencies of extreme temperature events in australia. *Australian Meteorological Magazine 49*(4), 277–292.

Dette, H., T. Schüler, and M. Vetter (2018). Multiscale change point detection for dependent data. *arxiv:1811.05956*.

Dette, H. and W. Wu (2019). Detecting relevant changes in the mean of nonstationary processes - a mass excess approach. *Ann. Statist. 47*(6), 3578–3608.

Fan, J. and I. Gijbels (1996). Local polynomial modelling and its applications. *Monographs on Statistics and Applied Probability. Chapman & Hall/CRC*.

Fan, M. (2004). Some inequalities for p-variations of martingales. *Stochastic Processes and their Applications 109*(2), 189 – 201.

Frick, K., A. Munk, and H. Sieling (2014). Multiscale change point inference. *Journal of the Royal Statistical Society, Ser. B 76*(3), 495–580.

Fryzlewicz, P. (2018). Tail-greedy bottom-up data decompositions and fast multiple change-point detection. *Ann. Statist. 46*(6B), 3390–3421.

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory 24*(3), 726–748.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The elements of statistical learning* (Second ed.). Springer Series in Statistics. Springer, New York. Data mining, inference, and prediction.

Horváth, L., P. Kokoszka, and J. Steinebach (1999). Testing for changes in multivariate dependent observations with an application to temperature changes. *Journal of Multivariate Analysis 68*(1), 96 – 119.

Hotz, T., O. Schütte, H. Sieling, T. Polupanow, U. Diederichsen, C. Steinem, and A. Munk (2013). Idealizing ion channel recordings by jump segmentation and statistical multiresolution analysis. *IEEE Trans. Nanobiosci. 12*, 376–386.

Jandhyala, V., S. Fotopoulos, I. MacNeill, and P. Liu (2013). Inference for single and multiple change-points in time series. *Journal of Time Series Analysis 34*(4), 423–446.

Johnston, G. J. (1982). Probabilities of maximal deviations for nonparametric regression function estimates. *Journal of Multivariate Analysis 12*(3), 402 – 414.

Karl, T. R., R. W. Knight, and N. Plummer (1995). Trends in high-frequency climate variability in the twentieth century. *Nature 377*(6546), 217.

Kirch, C., B. Muhsal, and H. Ombao (2015). Detection of changes in multivariate time series with application to eeg data. *Journal of the American Statistical Association 110*(511), 1197–1216.

Li, D., W. Liu, Q. Wang, and W. B. Wu (2017). Simultaneous confidence bands in nonlinear regression models with nonstationarity. *Statistica Sinica 27*(3), 1385–1400.

Piterbarg, V. (2012). *Asymptotic Methods in the Theory of Gaussian Processes and Fields*. Translations of Mathematical Monographs. American Mathematical Society.

Priestley, M. B. and T. Subba Rao (1969). A test for non-stationarity of time series. *Journal of the Royal Statistical Society 31*(1), 140–149.

Proksch, K. (2016). On confidence bands for multivariate nonparametric regression. *Annals of the Institute of Statistical Mathematics 68*(1), 209–236.

Schucany, W. R. and J. P. Sommers (1977). Improvement of kernel type density estimators. *Journal of the American Statistical Association 72*(358), 420–423.

Sharma, S., D. A. Swayne, and C. Obimbo (2016). Trend analysis and change point techniques: a survey. *Energy, Ecology and Environment 1*(3), 123–130.

Truong, C., L. Oudre, and N. Vayatis (2020). Selective review of offline change point detection methods. *Signal Processing 167*, 107299.

van der Vaart, A. and J. Wellner (1996). *Weak Convergence and Empirical Processes*, Volume 1 of *Springer series in statistics*. Springer Science+Business Media New York.

Vogt, M. and H. Dette (2015). Detecting gradual changes in locally stationary processes. *The Annals of Statistics 43*, 713–740.

Wolfe, D. A. and E. Schechtman (1984). Nonparametric statistical procedures for the changepoint problem. *Journal of Statistical Planning and Inference 9*(3), 389

– 396.

Woodall, W. H. and D. C. Montgomery (1999). Research issues and ideas in statistical process control. *Journal of Quality Technology 31*(4), 376–386.

Woodall, W. H. and D. C. Montgomery (2014). Some current directions in the theory and application of statistical process monitoring. *Journal of Quality Technology 46*(1), 78–94.

Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences 102*(40), 14150–14154.

Wu, W. B. (2007). Strong invariance principles for dependent random variables. *Ann. Probab. 35*(6), 2294–2320.

Wu, W. B. and Z. Zhao (2007). Inference of trends in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69*(3), 391–410.

Xia, Y. (1998). Bias-corrected confidence bands in nonparametric regression. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 60*(4), 797–811.

Zhao, Z. and W. B. Wu (2008, 08). Confidence bands in nonparametric time series regression. *Ann. Statist. 36*(4), 1854–1878.

Zhou, Z. (2013). Heteroscedasticity and autocorrelation robust structural change detection. *Journal of the American Statistical Association 108*, 726–740.

Zhou, Z. and W. B. Wu (2010). Simultaneous inference of linear models with time varying coefficients. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72*(4), 513–531.

HEINRICH-HEINE-UNIVERSITÄT DÜSSELDORF, MATHEMATISCHES INSTITUT, UNIVERSITÄTSSTR. 1, 40225 DÜSSELDORF, GERMANY.
Email address: `axel.buecher@hhu.de`

RUHR-UNIVERSITÄT BOCHUM, FAKULTÄT FÜR MATHEMATIK, UNIVERSITÄTSSTR. 150, 44780 BOCHUM, GERMANY.
Email address: `holger.dette@rub.de`
Email address: `florian.heinrichs@rub.de`