A DISTRIBUTION FREE TEST FOR CHANGES IN THE TREND FUNCTION OF LOCALLY STATIONARY PROCESSES

FLORIAN HEINRICHS AND HOLGER DETTE

ABSTRACT. In the common time series model $X_{i,n} = \mu(i/n) + \varepsilon_{i,n}$ with non-stationary errors we consider the problem of detecting a significant deviation of the mean function μ from a benchmark $g(\mu)$ (such as the initial value $\mu(0)$ or the average trend $\int_0^1 \mu(t)dt$). The problem is motivated by a more realistic modelling of change point analysis, where one is interested in identifying relevant deviations in a smoothly varying sequence of means $(\mu(i/n))_{i=1,\ldots,n}$ and cannot assume that the sequence is piecewise constant. A test for this type of hypotheses is developed using an appropriate estimator for the integrated squared deviation of the mean function and the threshold. By a new concept of self-normalization adapted to non-stationary processes an asymptotically pivotal test for the hypothesis of a relevant deviation is constructed. The results are illustrated by means of a simulation study and a data example.

Key words: change point analysis, local stationary processes, nonparametric regression

1. INTRODUCTION

Within the last decades, the detection of structural breaks in time series has become a very active area of research with many applications in fields like climatology, economics, engineering, genomics, hydrology, etc. (see Aue and Horváth, 2013; Jandhyala et al., 2013; Woodall and Montgomery, 2014; Sharma et al., 2016; Chakraborti and Graham, 2019; Truong et al., 2020, among many others). In the simplest case, one is interested in detecting structural breaks in the sequence of means $(\mu_i)_{i=1,\ldots,n} = (\mu(i/n))_{n\in\mathbb{N}}$ of a time series $(X_{i,n})_{i=1,\ldots,n}$ corresponding to a location model of the form

$$X_{i,n} = \mu(i/n) + \varepsilon_{i,n}\ ,\ \ i = 1, \ldots, n\ . \tag{1.1}$$

A large part of the literature considers the problem of detecting changes in a piecewise constant mean function $\mu : [0, 1] \to \mathbb{R}$, where early references assume the existence of at most one change point (see, e. g. Priestley and Subba Rao, 1969; Wolfe and Schechtman, 1984; Horváth et al., 1999, among others) and more recent literature investigates multiple change points (see, e.g. Frick et al., 2014; Fryzlewicz, 2018; Dette et al., 2020; Baranowski et al., 2019, among many others). The errors $(\varepsilon_{i,n})_{i=1,\ldots,n}$ in model (1.1) are usually assumed to form at least a stationary process and many theoretical results for detecting multiple change points are only available for independent identically distributed error processes. These assumptions simplify the statistical analysis of structural breaks

Date: May 21, 2020.

substantially, as - after removing the piecewise constant trend - one can work under the assumption of a stationary or an independent identically distributed error process and smoothing is not necessary to estimate the trend function.

On the other hand, the assumption of a strictly piecewise constant mean function might not be realistic in many situations and it might be more reasonable to assume that μ varies smoothly rather than abrupt. A typical example is temperature data (see, e. g. Karl et al., 1995; Collins et al., 2000) where it might be of more interest to investigate whether the mean function deviates fundamentally from a given benchmark denoted by $g(\mu)$. Here g is a functional of the mean function, such as the value at the point 0, that is $g(\mu) = \mu(0)$, or an average over a certain time period, that is

$$g(\mu) = \frac{1}{t_1 - t_0} \int_{t_0}^{t_1} \mu(x)dx \quad \text{for some} \quad 0 \leq t_0 < t_1 \leq 1 \tag{1.2}$$

(see Section 2 for more details). Moreover, there also exist many time series exhibiting a non-stationary behaviour in the higher order moments and dependence structure (see Stărică and Granger, 2005; Elsner et al., 2008; Guillaumin et al., 2017, among others), and the detection of fundamental deviations from a benchmark in a sequence of gradually changing means under the assumption of a location model with a stationary error process might be misleading.

In this paper we propose a distribution free test for relevant deviations of the mean function μ from a given benchmark $g(\mu)$ in a location scale model of the form (1.1) with a non-stationary error process. More precisely, for some pre-specified threshold $\Delta > 0$ we are interested in testing the hypotheses

$$H_0 : d_0 = \Big(\int_0^1 \big(\mu(x) - g(\mu)\big)^2 \tau(\mathrm{d}x) \Big)^{1/2} \leq \Delta \quad \text{vs.} \quad H_1 : d_0 > \Delta, \tag{1.3}$$

where τ is an appropriate measure on the interval $[0, 1]$ chosen by the statistician. This means that we are looking for "substantial" deviations of the mean function from a given benchmark $g(\mu)$ in an L^2-sense. The choice of the threshold depends on the particular application and is related to a balance between bias and variance as the detection of deviations from a (constant) mean often results in an adaptation of the statistical analysis (for example in forecasting). As such an analysis is 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 adaptation might not be necessary because a potential decrease in bias might be overcompensated by an increase of variance.

In principle, a test for the hypotheses in (1.3) could be developed using a nonparametric estimate of the mean function μ to obtain an estimate, say $\hat{d}_0$, of the distance d_0. The null hypothesis in (1.3) is then rejected for large values of $\hat{d}_0$. However, the distribution of the test statistic will depend in an intricate way on the dependence structure of the non-stationary error process in model (1.1), which is difficult to estimate. To address this problem we will introduce a new concept of self-normalization and construct an (asymptotically) pivotal test statistic for the hypotheses in (1.3). The basic idea of our approach is to permute the data and consider the partial sum process of this permutation, thus, taking into account observations over the whole interval rather than only the first observations. The new concept and the asymptotic properties of the standardized statistic can be found in Section 3, while some details on the testing problem and mathematical background on locally stationary processes are introduced in Section 2. In Section 4 we investigate the finite sample properties of the proposed testing procedure by means of a simulation study and provide an application to temperature data. Finally, in Section A, the proofs of the theoretical results in Section 3 are presented.

1.1. **Related literature.** Despite of its importance the problem of detecting relevant deviations in a sequence of gradually changing means has only been considered by a few authors. Dette and Wu (2019) investigate a mass excess approach for this problem. More precisely, these authors measure deviations from the benchmark by the Lebesgue measure of the set $\{t \in [0,1] : |\mu(t) - g(\mu)| > \Delta\}$ and test whether this quantity exceeds a certain threshold $c > 0$. Their approach requires estimation of the local long-run variance and multiplier bootstrap. More recently, Bücher et al. (2020) propose the maximal distance to measure relevant deviations from the benchmark and consider the null hypothesis $H_0 : \sup_{t\in[0,1]} |\mu(t) - g(\mu)| \leq \Delta$. While the maximum deviation might be easy to interpret for practitioners, the asymptotic analysis of a corresponding estimate is challenging. In particular it requires an estimation of the long-run variance and additionally the estimation of the sets, where the absolute difference $|\mu(t) - g(\mu)|$ attains its sup-norm. The methodology proposed here avoids the problem of estimating tuning parameters of this type using an L^2-norm in combination with a new concept of self-normalization.

Ratio statistics or self-normalization have been introduced by Horváth et al. (2008) and Shao (2010) in the context of change point detection in stationary processes and avoid a direct estimation of the long-run variance through a convenient rescaling of the test statistic. The currently available self-normalization procedures are based on partial sum processes (see Shao, 2015, for a recent review), which usually (under the assumption of stationarity) have a limiting process of the form $\{\sigma W(\lambda)\}_{\lambda\in[0,1]}$, where $\{W(\lambda)\}_{\lambda\in[0,1]}$ is a known stochastic process and σ an unknown factor encapsulating the dependency structure of the underlying process. In this case the factorisation of the limit into the long-run variance and a probabilistic term is used to construct a pivotal test statistic by forming a ratio such that the factor σ in the numerator and denominator cancels. However, in the case of non-stationarity, the situation is more complicated, because the limiting process is of the form $\{\int_0^\lambda \sigma(u) \mathrm{d}W(u)\}_{\lambda\in[0,1]}$ such that the probabilistic and the part representing the dependence structure cannot be separated. Zhao and Li (2013) and Rho and Shao (2015) discuss in fact these problems in the context of locally stationary time series, but the proposed self-normalizations need to be combined with a wild bootstrap. In this paper, we present a full self-normalization procedure for non-stationary time series, which might be also useful for testing classical hypotheses.

2. The testing problems and mathematical Preliminaries

Throughout this paper $\mathcal{L}^2([0,1])$ denotes the space of real-valued square-integrable functions on $[0,1]$ and $L^2([0,1])$ the corresponding normed vector space of equivalence classes. Let $\langle f,g\rangle = \int_0^1 f(x)g(x)\mathrm{d}x$ denote the scalar product in $L^2([0,1])$ and $\|f\|_2 = \langle f,f\rangle^{1/2}$ the corresponding norm, for $f,g \in L^2([0,1])$. Further, $\langle f,g\rangle_\tau = \int_0^1 f(x)g(x)\tau(\mathrm{d}x)$ and $\|f\|_{2,\tau} = \langle f,f\rangle_\tau^{1/2}$, for $f,g\in L^2([0,1],\tau)$. Finally, for the sake of readability, for functions in $\mathcal{L}^2([0,1])$, we denote the integral $\int_0^1 f(x)g(x)\mathrm{d}x$ by $\langle f,g\rangle$. Finally, if X is a real-valued random variable we use the notation (in the case of existence) $\|X\|_{q,\Omega} = (\mathbb{E}[|X|^q])^{1/q}$, for $q\geq 1$.

2.1. **Relevant deviations in a sequence of gradually changing means.** Recall the definition of model (1.1) and the hypotheses (1.3). Different benchmarks may be of interest in applications. For example, if one is interested in deviations from the value of the mean function at a given time, say $t\in[0,1]$, one could choose $g(\mu)=\mu(t)$, while relevant deviations from an average over a certain time period are obtained for the choice (1.2). In particular if t_0, t_1 and τ are chosen 0, 1 and the Lebesgue measure, respectively, one compares the local mean $\mu(x)$ with the overall mean $g(\mu)=\bar\mu=\int_0^1\mu(y)\mathrm{d}y$ and the hypotheses in (1.3) read as follows

$$H_0: \Big(\int_0^1 (\mu(x)-\bar\mu)^2\mathrm{d}x\Big)^{1/2} \leq \Delta \quad \text{vs.} \quad \Big(\int_0^1 (\mu(x)-\bar\mu)^2\mathrm{d}x\Big)^{1/2} > \Delta.$$

The tests which will be developed in this paper are based on an appropriate estimate of the quantity

$$d_0 = \Big(\int_0^1 (\mu(x)-g(\mu))^2\tau(\mathrm{d}x)\Big)^{1/2} \tag{2.1}$$

for which we require precise estimates of the mean function μ and the threshold $g(\mu)$. Note that the measure τ in (2.1) is chosen by the statistician and therefore known.

Throughout this paper, we assume that τ is absolutely continuous with respect to the Lebesgue measure and has a piecewise continuous density, say f_τ. Further, we assume that the mean function μ is sufficiently smooth, as specified in the following assumption.

Assumption 2.1. The function $\mu:[0,1]\to\mathbb{R}$ is twice differentiable with Lipschitz continuous second derivative. In particular, this implies that the integrals $\int_0^1\mu^2(x)\mathrm{d}x$ and $\int_0^1\mu^2(x)\tau(\mathrm{d}x)$ are finite, thus, $\mu\in\mathcal{L}^2([0,1])$ and $\mu\in\mathcal{L}^2([0,1],\tau)$.

A natural idea for the construction of a test of the hypotheses (1.3) is to estimate the L^2-distance d_0 as defined in (2.1) and to reject the null hypothesis for large values of the corresponding estimate. For this purpose one can use the *local linear estimator*, which is defined as the first coordinate of the vector

$$(\hat\mu_{h_n}(t), \widehat{\mu'}_{h_n}(t)) = \operatorname*{argmin}_{b_0,b_1}\sum_{i=1}^n (X_{i,n}-b_0-b_1(i/n-t))^2 K_{h_n}(i/n-t), \tag{2.2}$$

to estimate the mean function μ locally (see, for example Fan and Gijbels, 1996). In order to reduce the bias we consider the *Jackknife estimator*

$$\check{\mu}_{h_n}(t) = 2\hat{\mu}_{h_n/\sqrt{2}}(t) - \hat{\mu}_{h_n}(t) \tag{2.3}$$

as proposed by Schucany and Sommers (1977) and obtain an estimate $\check{g}_n = g(\hat{\mu}_{h_n})$ of the threshold $g(\mu)$ (other estimates could be used as well). Here h_n is a positive bandwidth satisfying $h_n = o(1)$ as $n \to \infty$, $K_h(\cdot) = K(\frac{\cdot}{h})$ and K denotes a kernel function satisfying the following assumption.

Assumption 2.2. The kernel K is non-negative, symmetric, supported on the interval $[-1, 1]$. It is twice differentiable, satisfies $\int_{[-1,1]} K(x)dx = 1$ and Lipschitz continuous in an open interval containing the interval $[-1, 1]$.

The estimate of d_0 can then be defined as

$$\|\check{\mu}_{h_n} - \check{g}_n\|_{2,\tau} = \Big(\int_0^1 (\check{\mu}_{h_n}(x) - \check{g}_n)^2 \tau(\mathrm{d}x) \Big)^{1/2} . \tag{2.4}$$

To study the asymptotic properties of the statistic defined in (2.4) and alternative estimates proposed in this paper (see Section 3 for more details) we require several assumptions regarding the dependency structure of the error process in model (1.1), which will be discussed next.

2.2. **Locally stationary processes.** For the proofs of our main results we require several assumption on the dependence structure of the non-stationary time series defined in (1.1). In the following, we work with the notion of local stationarity as introduced by Zhou and Wu (2009). To be precise, 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 $\mathcal{F}_i = \{\eta_k : k \leq i\}$ and $\mathcal{F}^*_i = (\dots, \eta_{-2}, \eta_{-1}, \eta'_0, \eta_1, \dots, \eta_i)$. Let $G : [0, 1] \times \mathbb{R}^\infty \to \mathbb{R}$ denote a filter, such that $G(t, \mathcal{F}_i)$ is a properly defined random variable for all $t \in [0, 1]$.

A triangular array $\{(\varepsilon_{i,n})_{1\leq i\leq n}\}_{n\in\mathbb{N}}$ is called *locally stationary*, if there exists a filter G, which is continuous in its first argument, such that

$$\varepsilon_{i,n} = G(i/n, \mathcal{F}_i)$$

for all $i \in \{1, \dots, n\}, n \in \mathbb{N}$. The *physical dependence measure of a filter* G with $\sup_{t\in[0,1]} \|G(t, \mathcal{F}_i)\|_{q,\Omega} < \infty$ with respect to $\|\cdot\|_{q,\Omega}$ is defined by

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

A filter G is called *Lipschitz continuous with respect to* $\|\cdot\|_{q,\Omega}$, if

$$\sup_{0\leq s<t\leq 1} \|G(t, \mathcal{F}_i) - G(s, \mathcal{F}_i)\|_{q,\Omega}/|t - s| < \infty.$$

The filter G models the non-stationarity of $(\varepsilon_{i,n})$. The quantity $\delta_q(G, i)$ measures the dependence of $(\varepsilon_{i,n})$ and plays a similar role as mixing coefficients. We now state some assumptions regarding the error terms in model (1.1).

Assumption 2.3. Let the triangular array $\{(\varepsilon_{i,n})_{1\leq i\leq n}\}_{n\in\mathbb{N}}$ in (1.1) be centered and locally stationary with filter G, such that the following conditions are satisfied:

(1) There exists a constant $\gamma \in (0,1)$ such that $\delta_4(G,i) = \mathcal{O}(\gamma^i)$, as $i \to \infty$.
(2) The filter G is Lipschitz continuous with respect to $\|\cdot\|_{4,\Omega}$ and

$$\sup_{t\in[0,1]} \|G(t,\mathcal{F}_0)\|_{4,\Omega} < \infty.$$

(3) The (local) *long-run variance* of G, defined as

$$\sigma^2(t) = \sum_{i=-\infty}^{\infty} \operatorname{Cov}\big(G(t,\mathcal{F}_i), G(t,\mathcal{F}_0)\big), \tag{2.5}$$

for $t \in [0,1]$, is Lipschitz continuous and bounded away from zero, i. e.,

$$\inf_{t\in[0,1]} \sigma^2(t) > 0.$$

(4) The moments of order 8 are uniformly bounded, i. e., $\max_{1\leq i\leq n} \mathbb{E}\varepsilon_{i,n}^8 < \infty$.

2.3. **Testing for relevant differences - the problem of estimating the variance.** Continuing the discussion in Section 2.1 it follows from the results given in Section 3 that the estimator (2.4) is asymptotically normal distributed if Assumptions 2.1, 2.2, 2.3 and an additional assumption on the consistency of the statistic $\breve{g}_n$ are satisfied. More precisely, it can be shown (see Remark 3.7) that

$$\sqrt{n}(\|\breve{\mu}_{h_n} - \breve{g}_n\|_{2,\tau}^2 - \|\mu - g(\mu)\|_{2,\tau}^2) \rightsquigarrow \mathcal{N}(0, 4\|d_\omega\sigma\|_2^2), \tag{2.6}$$

where the symbol $\rightsquigarrow$ denotes weak convergence, $\sigma^2(\cdot)$ is the local long-run variance defined in (2.5) and $d_\omega(\cdot)$ denotes an unknown function, that depends on the function μ and the error process. In principle, if $\hat{\sigma}_n^2$ and $\hat{d}_\omega^2$ are estimators of the local long-run variance and the function d_ω, respectively, a reasonable strategy would be to reject the null hypothesis in (1.3) if

$$\|\breve{\mu}_{h_n} - \breve{g}_n\|_{2,\tau}^2 > \Delta^2 + z_{1-\alpha}\frac{2\|\hat{d}_\omega\hat{\sigma}_n\|_2}{\sqrt{n}}, \tag{2.7}$$

where $z_{1-\alpha}$ denotes the $(1-\alpha)$-quantile of the standard normal distribution. It will be shown in Remark 3.7 below, that this decision rule provides a consistent and asymptotic level α-test for the hypotheses in (1.3). However, it turns out that this decision rule does not provide a stable test because local estimators of the long-run variance have a rather large variability.

In order to avoid the intricate estimation of the local long-run variance we will re-define the local linear estimator in (2.2) permuting the data and consider the partial sum process of the new estimators in the following section. This approach will enable us to construct an (asymptotically) pivotal test statistic for the hypotheses in (1.3).

3. A pivotal test statistic

3.1. **Self-normalization.** A common technique to avoid estimating the long-run variance are ratio statistics or self-normalization as first introduced by Horváth et al. (2008) and Shao (2010), which are based on a convenient rescaling of the test statistic. However, these concepts are not easy to transfer to non-stationary time series as they rely on the asymptotic properties of a corresponding partial sum process. To illustrate the problems of these concepts in non-stationary time series consider the simplest case of model (1.1), where the mean function is constant and the error process is stationary. In this case the estimate of the constant mean function μ from the partial sample $X_{1,n}, \dots, X_{\lfloor \lambda n \rfloor,n}$ is its mean and under the assumptions stated in Section 2 we have the weak convergence

$$\{B_n(\lambda)\}_{\lambda\in[0,1]} = \Big\{n^{-1/2} \sum_{i=1}^{\lfloor \lambda n \rfloor} (X_{i,n} - \mu)\Big\}_{\lambda\in[0,1]} \rightsquigarrow \{\sigma W(\lambda)\}_{\lambda\in[0,1]},$$

where $\big\{W(\lambda)\big\}_{\lambda\in[0,1]}$ denotes a standard Brownian motion and the long-run variance σ^2 is defined in (2.5) and does not depend on t (because of the stationarity assumption). In this case, the factorisation of the limit into the long-run variance and a probabilistic term is used to construct a test statistic in the form of a ratio, such that σ occurs in the nominator and denominator, and therefore cancels out. On the other hand, if the error process in model (1.1) is non-stationary (but the mean function is still constant) we have the weak convergence

$$\{B_n(\lambda)\}_{\lambda\in[0,1]} \rightsquigarrow \Big\{ \int_0^\lambda \sigma(u) \mathrm{d}W(u)\Big\}_{\lambda\in[0,1]}.$$

In this case, the limiting distribution does not factorise and it is no longer possible to use the common self-normalization approach. Zhao and Li (2013) and Rho and Shao (2015) discuss locally stationary time series, but the proposed self-normalization procedures have to be combined with a wild bootstrap.

In this work, we present an alternative self-normalization procedure for non-stationary time series which does not require resampling to obtain (asymptotically) pivotal statistics. Our approach is based on the idea that in a locally stationary setting, observations from the whole interval $[0,1]$ need to be taken into account. Therefore, let b_n denote a sequence with $b_n \to \infty$ and $b_n/n \to 0$, as $n \to \infty$, and let $\ell_n = \lfloor n/b_n \rfloor$. We define a (fixed) permutation of the set $\{1, \dots, n\}$ by

$$T : \begin{cases} \{1,\dots,n\} & \to \quad \{1,\dots,n\} \\ \quad k & \mapsto \quad T_k = \begin{cases} (k-1 \mod \ell_n) b_n + \lceil k/\ell_n \rceil, & \text{if } k \le \ell_n b_n \\ k, & \text{if } k > \ell_n b_n \end{cases} \end{cases}$$

Note that for $k = i\ell_n + j$ it holds $T_k = (j-1)b_n + i + 1$, where $i \in \{0, \dots, b_n\}$ and $j \in \{1, \dots, \ell_n\}$.

Roughly speaking, the mapping T splits the set $\{1,\dots,n\}$ into ℓ_n blocks with block length b_n, that is

$$\begin{array}{rcl} \{T_1,\dots,T_{\ell_n}\} & = & \{1, b_n+1, 2b_n+1,\dots,(l_n-1)b_n+1\} \\ \{T_{\ell_n+1},\dots,T_{2\ell_n}\} & = & \{2, b_n+2, 2b_n+2,\dots,(l_n-1)b_n+2\} \\ \{T_{2\ell_n+1},\dots,T_{3\ell_n}\} & = & \{3, b_n+3, 2b_n+2,\dots,(l_n-1)b_n+3\} \\ \vdots & \vdots & \qquad\qquad\qquad\vdots \end{array}$$

where the blocks correspond to the columns in the above display.

With this notation, for $\zeta>0$ and $\lambda\in[\zeta,1]$, we define the sequential local linear estimator of the mean function μ from the sample $X_{T_1,n},\dots,X_{T_{\lfloor\lambda n\rfloor},n}$ as the first coordinate of the vector

$$(\hat\mu_{h_n}(\lambda,t),\widehat{\mu'}_{h_n}(\lambda,t)) = \underset{b_0,b_1}{\operatorname{argmin}}\sum_{i=1}^{\lfloor\lambda n\rfloor}\big(X_{T_i,n}-b_0-b_1(T_i/n-t)\big)^2K_{h_n}(T_i/n-t). \quad (3.1)$$

In the following we will work with a bias corrected version of $\hat\mu_{h_n}(\lambda,t)$ and consider the sequential Jackknife estimator

$$\tilde\mu_{h_n}(\lambda,t)=2\hat\mu_{h_n/\sqrt{2}}(\lambda,t)-\hat\mu_{h_n}(\lambda,t). \quad (3.2)$$

With the notation

$$d(x):=\mu(x)-g(\mu)$$

we can rewrite the distance in (2.1) as $d_0=\|d\|_{2,\tau}$. In order to estimate d_0 let $\hat g_n(\lambda)$ be a suitable sequential estimator of the benchmark $g(\mu)$ from the sample $X_{T_1,n},\dots,X_{T_{\lfloor\lambda n\rfloor},n}$ and define

$$\hat d_n(\lambda,x)=\tilde\mu_{h_n}(\lambda,x)-\hat g_n(\lambda)$$

and

$$\hat d_{2,n}(\lambda)=\|\hat d_n(\lambda,\cdot)\|_{2,\tau}.$$

Note that all estimates are calculated from a part of the permuted sample and that the statistic $\hat d_{2,n}(1)$ estimates d_0 from the full sample $X_{1,n},\dots,X_{n,n}$ and therefore coincides with the estimator defined in (2.4). For the proofs of our main results we need an assumption regarding the precision of the estimator $\hat g_n(\cdot)$ of the benchmark, the bandwidth h_n and the block length b_n, which are given next.

Assumption 3.1. The sequential estimator $\hat g_n(\lambda)$ of the benchmark $g(\mu)$ admits a stochastic expansion

$$\lambda\sqrt{n}(\hat g_n(\lambda)-g(\mu))=\frac{1}{\sqrt n}\sum_{i=1}^{\lfloor\lambda n\rfloor}\varepsilon_{T_i,n}\omega_n(T_i/n)+o_{\mathbb P}(1),$$

uniformly with respect to $\lambda\in[\zeta,1]$ for some constant $\zeta\in(0,1)$ and functions $\omega_n,\omega\in\mathcal L^4([0,1])$ such that ω_n is Riemann-integrable for any $n\in\mathbb N$, $\|\omega_n-\omega\|_4\to0$ and

$$\sum_{j=1}^{\ell_n}\sum_{r=1}^{b_n}\big|\omega_n\big(\tfrac{jb_n}{n}\big)-\omega_n\big(\tfrac{r+jb_n}{n}\big)\big|=\mathcal O(b_nh_n^{-1}), \quad (3.3)$$

where $\|\omega_n - \omega\|_4 = \big(\int_0^1 \|\omega_n(x) - \omega(x)\|^4 dx\big)^{1/4}$.

Assumption 3.2. There exist constants $\alpha, \beta > 0$ such that the sequence of bandwidths $h_n \to 0$ satisfies $nh_n \to \infty$, $nh_n^6 \to 0$, $n^\beta = \mathcal{O}(nh_n^4)$ and the sequence $b_n \to \infty$ satisfies $b_n^3/n \to 0$, $\frac{b_n^2}{nh_n} \to 0$, $n^\alpha = \mathcal{O}(b_n)$.

Remark 3.3.

(1) Assumption 3.1 is rather mild and satisfied for many functionals as explained below. Proofs of the following statements can be found in Section A.3 of the Appendix.

- (i) Condition (3.3) is satisfied for all Lipschitz continuous functions and all step functions on the interval $[0, 1]$.
- (ii) The assumption holds for $g(\mu) = c$ with some known $c \in \mathbb{R}$, for $\hat{g}_n(\lambda) = g(\tilde{\mu}_{h_n}(\lambda, \cdot))$.
- (iii) Assumption 3.1 is satisfied for the functional defined in (1.2) and the estimator

$$\hat{g}_n(\lambda) = \frac{1}{(t_1 - t_0)\lambda n} \sum_{i=1}^{\lfloor \lambda n \rfloor} X_{T_i,n} \mathbb{1}(t_0 \le T_i/n \le t_1).$$

- (iv) Let $g : L^2([0, 1]) \to \mathbb{R}$ be a linear, bounded operator. By the Riesz-Fréchet representation theorem, there exists $\bar{h}_g \in L^2([0, 1])$ such that $g(\cdot) = \langle \cdot, \bar{h}_g \rangle$. If there exists a continuous function h_g in the equivalence class corresponding to $\bar{h}_g$, the estimator $\hat{g}_n(\lambda) = g(\tilde{\mu}_{h_n}(\lambda, \cdot))$ satisfies Assumption 3.1.
- (v) The functional $g(\mu) = \mu(t)$ (for some fixed $t \in [0, 1]$) is not covered by Assumption 3.1. Nevertheless a corresponding pivotal test can be developed as well - see Remark 3.5 for more details.

(2) Assumption 3.2 is satisfied, if $h_n = n^{-1/5}, b_n = n^{1/4}$. In this case, the constants α and β can be chosen as $\frac{1}{4}$ and $\frac{1}{5}$, respectively.

Theorem 3.4. *Let Assumptions 2.1, 2.2, 2.3, 3.1 and 3.2 be satisfied. For any $\zeta \in (0, 1)$, the process*

$$\{G_n(\lambda)\}_{\lambda \in [\zeta,1]} = \{\lambda \sqrt{n}(\hat{d}_{2,n}^2(\lambda) - d_0^2)\}_{\lambda \in [\zeta,1]} \tag{3.4}$$

converges weakly to the process

$$\{G(\lambda)\}_{\lambda \in [\zeta,1]} = \{2\big\|d_\omega \sigma\big\|_2 W(\lambda)\}_{\lambda \in [\zeta,1]} \tag{3.5}$$

in $\ell^\infty([\zeta, 1])$, where $d_\omega(\cdot) = f_\tau(\cdot)d(\cdot) + \omega(\cdot) \int_0^1 d(x)\tau(\mathrm{d}x)$ and $\{W(\lambda)\}_{\lambda \in [0,1]}$ denotes a standard Brownian motion. In particular, $G(\lambda) = 0$ if $d_0 = 0$.

Remark 3.5. If $g(\mu) = \mu(t)$, for some fixed $t \in [0, 1]$, the benchmark $g(\mu)$ needs to be estimated locally and there is no estimator satisfying Assumption 3.1. However, an analogous result as stated in Theorem 3.4 can be shown with the same arguments given in the proof of the latter theorem. More precisely, if $g(\mu) = \mu(t)$ we can use $\hat{g}_n(\lambda) = \tilde{\mu}_{h_n}(\lambda, t)$ and under Assumptions 2.1, 2.2, 2.3 and 3.2, the process

$$\{G'_n(\lambda)\}_{\lambda\in[\zeta,1]} = \{\lambda\sqrt{nh_n}(\hat{d}^2_{2,n}(\lambda) - d_0^2)\}_{\lambda\in[\zeta,1]}$$

converges weakly to

$$\{G'(\lambda)\}_{\lambda\in[\zeta,1]} = \left\{2\sigma(t)\kappa(t)\int_0^1 d(x)\tau(\mathrm{d}x)W(\lambda)\right\}_{\lambda\in[\zeta,1]}$$

in $\ell^\infty([\zeta,1])$, where the constant κ is defined by $\kappa^2(t) = \int_{-1}^1 \left(K^*(x)\right)^2 \mathrm{d}x$ if $t \in (0,1)$ and by

$$\kappa^2(t) = \frac{1}{(\kappa_{t,0}\kappa_{t,2} - \kappa_{t,1}^2)^2}\int_{-t}^{1-t}\left\{\left(\frac{\kappa_{t,2}}{\sqrt{2}} - \kappa_{t,1}x\right)K^*(x) + \left(\frac{1}{\sqrt{2}} - 1\right)\kappa_{t,2}K(x)\right\}^2 \mathrm{d}x,$$

with $\kappa_{t,j} = \int_{-t}^{1-t} x^j K(x)\mathrm{d}x$, for $j \in \{0,1,2\}$ and $t \in \{0,1\}$, and K^* is defined by $K^*(x) = 2\sqrt{2}K(\sqrt{2}x) - K(x)$.

In the following, we will develop a pivotal test for the hypotheses (1.3) on the basis of Theorem, 3.4 or Remark 3.5. For this purpose let ν be a probability measure on the interval $[\zeta, 1]$ with $\nu(\{1\}) = 0$. We propose to reject the null hypothesis if

$$\hat{d}^2_{2,n}(1) > \Delta^2 + q_{1-\alpha}\int_\zeta^1 \lambda|\hat{d}^2_{2,n}(\lambda) - \hat{d}^2_{2,n}(1)|\mathrm{d}\nu(\lambda), \tag{3.6}$$

where $q_{1-\alpha}$ denotes the $(1-\alpha)$-quantile of the distribution of the random variable

$$\frac{W(1)}{\int_\zeta^1 |W(\lambda) - \lambda W(1)|\mathrm{d}\nu(\lambda)}.$$

Corollary 3.6. *Let the assumptions of either Theorem 3.4 or of Remark 3.5 be satisfied. If $\Delta > 0$, the decision rule (3.6) defines a consistent and asymptotic level α-test for the hypotheses (1.3) of a relevant deviation of the mean function μ from the threshold $g(\mu)$, that is*

$$\mathbb{P}(\text{ the null hypothesis (1.3) is rejected by (3.6) }) \xrightarrow{n\to\infty} \begin{cases} 0, & \textit{if } d_0 < \Delta \\ \alpha, & \textit{if } d_0 = \Delta \\ 1, & \textit{if } d_0 > \Delta. \end{cases}$$

Remark 3.7. Note that the Jackknife estimator defined in (2.3) coincides with $\tilde{\mu}_{h_n}(1,\cdot)$ and $\breve{g}_n = g(\tilde{\mu}_{h_n}(1,\cdot))$. Consequently, the continuous mapping theorem and Theorem 3.4 yield the weak convergence stated in equation (2.6) of Section 2.3. Consequently, if the estimators $\hat{\sigma}^2_n$ and $\hat{d}^2_\omega$ are consistent, the decision rule in (2.7) defines a consistent and asymptotic level α test for the hypothesis (1.3), that is

$$\mathbb{P}(\text{ the null hypothesis (1.3) is rejected by (2.7) }) \xrightarrow{n\to\infty} \begin{cases} 0, & \text{if } d_0 < \Delta \\ \alpha, & \text{if } d_0 = \Delta \\ 1, & \text{if } d_0 > \Delta. \end{cases}$$

4. Finite Sample Properties

4.1. **Monte Carlo simulation study.** A large scale Monte Carlo simulation study was performed to analyse the finite-sample properties of the proposed test (3.6). The local linear estimator in (3.1) requires the specification of the kernel K and the bandwidth h_n. We used the quartic kernel $K(x) = \frac{15}{16}(1-x^2)^2$, but other kernels will yield similar results. The choice of the bandwidth h_n for the estimator $\tilde{\mu}_{h_n}$ is crucial to avoid both overfitting and oversmoothing, and we employ the following k-fold cross-validation procedure with $k = 10$ (as recommended by Hastie et al., 2009, page 242).

Algorithm 4.1 (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.
(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

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

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

As block width we chose $b_n = 20$ and as measure ν on $[0,1]$ in (3.6) we used the uniform distribution on the set $\{1/5, \dots, 4/5\}$. Preliminary simulation studies showed that different choices of b_n and the measure ν lead to similar results.

We considered two types of mean functions μ, three different error processes and four different choices of the time-dependent variance. The first class of models is based on the mean function

$$\mu_a^{(1)}(x) = 10 + \tfrac{1}{2}\sin(8\pi x) + a\big(x - \tfrac{1}{4}\big)^2 \mathbb{1}\big(x > \tfrac{1}{4}\big), \tag{4.1}$$

which is displayed in the left part of Figure 1 for various choices of the parameter a. We considered the testing problem

$$H_0 : d_0 := \left\| \mu_a^{(1)} - g(\mu) \right\|_{2,\tau} \le 1/2 \quad \text{vs.} \quad H_1 : \left\| \mu_a^{(1)} - \bar{\mu}_a^{(1)} \right\|_{2,\tau} > 1/2, \tag{4.2}$$

where

$$g(\mu) = \bar{\mu}_a^{(1)} = 2 \int_0^{1/2} \mu_a^{(1)}(x) \mathrm{d}x,$$

and $\tau(\cdot) = 2\lambda_{[1/2,1]}(\cdot)$ is the Lebesgue measure on the interval $[\frac{1}{2}, 1]$. Such a scenario might for instance be encountered and of interest in the context of analyzing climate data where measurements for a recent period are compared with an average from previous years.

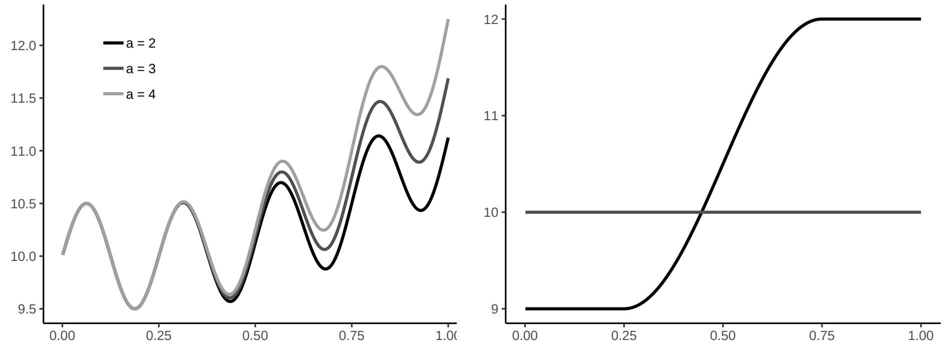

FIGURE 1. *Left: The mean function $\mu_a^{(1)}$ for three choices of a. Right: The mean function $\mu^{(2)}$.*

Note that $\|\mu_a^{(1)} - \bar{\mu}_a^{(1)}\|_{2,\tau} = 1/2$ for $a^* \approx 1.43$. We call this situation (i.e. when there is equality in (4.2)) the *boundary of the hypotheses*. On the other hand for $a < a^*$ and $a > a^*$ the null hypothesis and alternative in (4.2) are satisfied, respectively.

The second model has the mean function

$$\mu^{(2)}(x) = \begin{cases} 9 & \text{for } x \leq \frac{1}{4} \\ -\frac{3}{2}\sin(2\pi x) + 10.5 & \text{for } \frac{1}{4} < x \leq \frac{3}{4} \\ 12 & \text{for } \frac{3}{4} < x \end{cases} \tag{4.3}$$

which is displayed in the right part of Figure 1. For models involving this mean function, we considered the testing problem

$$H_0 : d_0 = \|\mu^{(2)} - g(\mu)\|_{2,\tau} \leq \Delta \quad \text{vs.} \quad H_1 : \|\mu^{(2)} - g(\mu)\|_{2,\tau} > \Delta \tag{4.4}$$

for various choices of the threshold $\Delta > 0$, where $g(\mu) \equiv 10$ and $\tau(\cdot) = \lambda_{[0,1]}(\cdot)$ is the Lebesgue measure on the interval $[0, 1]$. 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 $\|\mu^{(2)} - 10\|_{2,\tau} \leq \Delta$ for $\Delta \geq 1.392$, whereas $\|\mu^{(2)} - 10\|_{2,\tau} > \Delta$ for $\Delta < 1.392$.

We consider four different choices of time-dependent variance $\tilde{\sigma}^2(t) = \mathbb{E}[G^2(t, \mathcal{F}_0)]$, that is

$$\tilde{\sigma}_0^2(t) = 1, \qquad \tilde{\sigma}_1^2(t) = \tfrac{1}{2} + t,$$
$$\tilde{\sigma}_2^2(t) = 1 - \tfrac{1}{2}\cos(2\pi t), \qquad \tilde{\sigma}_3^2(t) = \tfrac{1}{2} + \mathbb{1}(t \geq 1/2)\ ,$$

and three classes of error processes $\{\varepsilon_{i,n} : 1 \leq i \leq n\}_{n\in\mathbb{N}}$ in model (1.1), that is

$$\text{(IID)} \quad \varepsilon_{i,n} = \tilde{\sigma}_k(i/n)\eta_i$$
$$\text{(MA)} \quad \varepsilon_{i,n} = \tilde{\sigma}_k(i/n)(\eta_i + \tfrac{1}{2}\eta_{i-1})/2$$
$$\text{(AR)} \quad \varepsilon_{i,n} = \tilde{\sigma}_k(i/n)(\eta_i + \tfrac{1}{2}\varepsilon_{i-1,n})/2,$$

for $k \in \{0, 1, 2, 3\}$, where $(\eta_i)_{i\in\mathbb{Z}}$ is an i.i.d. sequence of standard normal distributed random variables.

$\mu_a^{(1)}$		$\tilde{\sigma}_0^2$			$\tilde{\sigma}_1^2$			$\tilde{\sigma}_2^2$			$\tilde{\sigma}_3^2$		
a	$d_0 - \frac{1}{2}$	200	500	1000	200	500	1000	200	500	1000	200	500	1000
Panel A: iid errors													
0.37	-0.15	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0
0.89	-0.10	0.0	0.1	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.2	0.1
1.18	-0.05	0.2	0.9	0.6	0.0	1.0	0.3	0.0	1.4	0.4	0.1	0.9	0.6
1.43	**0.00**	**0.7**	**3.9**	**4.5**	**0.4**	**2.8**	**3.4**	**0.5**	**5.5**	**4.6**	**0.3**	**2.7**	**3.4**
1.86	0.10	2.7	23.6	43.1	2.1	21.6	33.5	3.0	32.2	47.0	1.0	13.4	26.5
2.26	0.20	7.4	54.6	83.9	8.2	46.3	73.0	9.5	66.5	88.6	3.5	33.4	61.4
2.64	0.30	14.1	76.6	95.8	12.1	68.7	91.8	19.5	88.8	98.0	6.2	51.1	83.1
Panel B: MA errors													
0.37	-0.15	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.89	-0.10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1.18	-0.05	0.1	0.4	0.2	0.0	0.8	0.1	0.0	0.2	0.2	0.1	0.2	0.4
1.43	**0.00**	**0.3**	**4.8**	**4.9**	**0.1**	**4.1**	**4.2**	**0.5**	**7.0**	**5.3**	**0.1**	**4.4**	**4.1**
1.86	0.10	4.3	38.3	58.7	3.7	30.9	53.3	5.5	45.4	69.0	3.0	21.9	39.2
2.26	0.20	12.6	76.8	94.8	11.6	69.3	91.4	16.3	83.8	97.8	9.4	52.5	79.5
2.64	0.30	29.5	93.7	99.9	28.6	87.6	98.8	33.4	97.4	99.7	21.4	75.4	94.0
Panel C: AR errors													
0.37	-0.15	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
0.89	-0.10	0.0	0.3	0.0	0.2	0.1	0.0	0.1	0.3	0.0	0.3	0.6	0.1
1.18	-0.05	0.4	1.3	1.1	0.1	1.5	0.9	0.0	2.3	1.3	0.4	2.4	1.3
1.43	**0.00**	**1.3**	**8.6**	**6.5**	**1.2**	**7.1**	**7.5**	**1.0**	**7.8**	**7.0**	**0.8**	**5.6**	**5.8**
1.86	0.10	5.3	36.3	52.8	5.4	29.5	47.9	6.8	41.1	57.0	3.9	22.6	35.4
2.26	0.20	16.4	67.8	90.0	14.6	59.9	85.7	18.0	78.2	92.6	10.0	45.0	70.1
2.64	0.30	30.6	86.6	98.9	24.1	80.1	96.1	35.4	92.9	99.4	18.7	65.6	89.5

TABLE 1. *Empirical rejection rates of the test* (3.6) *for the hypotheses* (4.2). *The mean function is given by* (4.1)*, where different values for the parameter* a*, different error processes, and sample sizes are considered. The lines in boldface correspond to the boundary of the hypotheses.*

The empirical rejection rates of the test (3.6) for the hypotheses $H_0 : d_0 \leq \Delta$ vs. $H_0 : d_0 > \Delta$ are calculated by $N = 1000$ simulation runs and displayed in Table 1 and Table 2. The sample size is chosen as $n = 200$, 500 and 1000 and the nominal level is 5%. Table 1 shows the rejection probabilities for different values of a in the function $\mu_a^{(1)}$ defined in (4.1), which yields to different values of d_0 in the hypotheses (4.2). On the other hand, in Table 2 the function $\mu^{(2)}$ and therefore the value d_0 is fixed and the threshold Δ in the hypotheses is varied. The lines marked in boldface indicate the boundary of the null hypothesis, that is, the parameter where $d_0 = \Delta$. More precisely, note that the null hypothesis in (4.2) holds if and only if $d_0 \leq 0.5$ and we display exemplary results for the cases $d_0 = 0.35$, 0.4, 0.45 and 0.5 in Table 1, where the last case corresponds to the boundary of the null hypotheses. The remaining cases $d_0 = 0.6$, 0.7 and 0.8 represent three scenarios of the alternative in (4.2). Similarly, in Table 2 the function $\mu^{(2)}$ is fixed with $d_0 = 1.39$. Therefore, the null hypothesis in

$\mu^{(2)}$		$\tilde{\sigma}_0^2$			$\tilde{\sigma}_1^2$			$\tilde{\sigma}_2^2$			$\tilde{\sigma}_3^2$		
Δ	$\Delta - d_0$	200	500	1000	200	500	1000	200	500	1000	200	500	1000
Panel A: iid errors													
1.30	-0.09	13.8	41.8	62.8	12.5	36.3	59.4	14.8	47.0	68.4	9.1	23.4	44.3
1.34	-0.05	6.8	20.8	34.3	6.9	19.1	29.8	7.2	22.8	37.6	4.8	11.7	23.3
1.38	-0.01	2.1	8.9	11.2	4.0	8.2	8.3	3.1	8.4	11.2	2.7	4.8	7.9
1.39	**0.00**	**2.0**	**5.0**	**5.1**	**2.2**	**5.2**	**5.4**	**2.4**	**5.1**	**5.8**	**2.4**	**3.3**	**4.5**
1.41	0.02	1.0	2.6	1.1	1.8	3.0	1.0	1.2	2.6	1.5	1.8	2.0	2.4
1.48	0.09	0.3	0.0	0.0	0.1	0.0	0.0	0.1	0.1	0.0	0.5	0.5	0.0
1.55	0.16	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
Panel B: MA errors													
1.30	-0.09	27.3	58.0	84.6	20.7	54.4	79.7	32.5	67.6	86.3	15.8	40.2	66.1
1.34	-0.05	12.6	30.7	52.3	9.5	26.9	45.9	15.2	37.2	55.7	7.2	18.8	34.5
1.38	-0.01	5.5	11.1	13.7	3.8	8.5	12.6	5.3	11.1	14.8	2.8	7.6	11.7
1.39	**0.00**	**3.0**	**5.0**	**7.7**	**3.4**	**5.1**	**6.2**	**3.6**	**6.9**	**7.5**	**2.2**	**4.4**	**5.9**
1.41	0.02	1.6	2.0	1.5	1.3	1.2	0.9	1.0	2.0	0.8	1.3	2.3	1.5
1.48	0.09	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1.55	0.16	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Panel C: AR errors													
1.30	-0.09	23.8	52.2	75.0	20.7	43.5	67.8	26.6	58.6	79.2	15.1	36.7	53.6
1.34	-0.05	13.9	28.3	42.8	11.8	23.2	39.3	13.2	32.8	50.1	9.3	22.0	31.1
1.38	-0.01	7.9	10.1	15.3	6.3	8.9	12.8	5.8	12.7	16.8	6.0	10.1	12.5
1.39	**0.00**	**5.7**	**7.5**	**9.0**	**4.5**	**7.5**	**8.6**	**6.3**	**8.0**	**9.6**	**4.2**	**7.3**	**9.9**
1.41	0.02	3.2	2.6	2.5	2.9	2.4	2.1	2.2	3.3	2.1	3.5	3.7	3.0
1.48	0.09	0.6	0.0	0.0	0.2	0.0	0.0	0.3	0.1	0.0	0.8	0.3	0.1
1.55	0.16	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0

TABLE 2. *Empirical rejection rates of the test* (3.6) *for the hypotheses* (4.4). *The mean function is given by* (4.3), *where different values for the threshold* Δ, *different error processes, and sample sizes are considered. The lines in boldface correspond to the boundary of the hypotheses.*

(4.4) holds if and only if the threshold satisfies $\Delta \geq 1.39$. We observe in most cases a good approximation of the nominal level at the boundary of the hypotheses and the test is also able to detect alternatives with reasonable power. These empirical findings corresponds with the theoretical results derived in Section 3.

We conclude this section with a comparison of the new test (3.6) with the test (2.7) which relies on the estimation of the (local) long-run variance. For this purpose we use the long-run variance estimator as proposed in equation (4.7) of Dette and Wu (2019) with bandwidths as suggested in this reference. In Table 3 and 4 we display the rejection probabilities for both tests for some of the models considered in Table 1 and Table 2, where we use the Lebesgue measure on the interval $[0, 1]$ for the calculation of the L^2-distances and the benchmark is given by $g(\mu) = \int_0^1 \mu(x)\mathrm{d}x$. For the sake of brevity we restrict ourselves to the sample size $n = 500$ and the variance function $\tilde{\sigma}_0^2(t) = 1$.

	errors	i.i.d		MA		AR	
a	$d_0 - \Delta$	(3.6)	(2.7)	(3.6)	(2.7)	(3.6)	(2.7)
0.13	-0.15	0.0	0.0	0.0	0.0	0.0	0.0
1.60	-0.10	0.0	0.0	0.0	0.0	0.0	0.0
2.13	-0.05	0.0	0.0	0.1	0.0	0.1	0.0
2.57	**0.00**	**2.2**	**0.0**	**1.6**	**0.0**	**4.3**	**0.4**
2.97	0.05	14.0	3.0	20.1	0.9	22.6	3.5
3.35	0.10	35.0	24.2	58.1	20.5	46.3	26.8
3.71	0.15	61.9	67.1	84.8	75.7	72.9	71.7

TABLE 3. *Empirical rejection rates of tests* (3.6) *and* (2.7) *for the hypotheses* (4.2). *The mean function is given by* (4.1), *where different values for the parameter a and different error processes are considered. The variance is $\tilde{\sigma}_0^2(t) = 1$, the sample size is $n = 500$ and the line in boldface corresponds to the boundary of the hypotheses.*

	errors	i.i.d		MA		AR	
Δ	$\Delta - d_0$	(3.6)	(2.7)	(3.6)	(2.7)	(3.6)	(2.7)
1.15	-0.15	73.8	83.9	90.3	93.1	81.8	86.2
1.20	-0.10	48.0	45.0	69.3	51.7	58.3	50.4
1.25	-0.05	22.9	10.5	32.1	6.7	31.3	12.4
1.30	**0.00**	**5.5**	**0.5**	**4.9**	**0.1**	**8.0**	**0.7**
1.35	0.05	0.5	0.0	0.1	0.0	1.4	0.1
1.40	0.10	0.0	0.0	0.0	0.0	0.1	0.0
1.45	0.15	0.0	0.0	0.0	0.0	0.0	0.0

TABLE 4. *Empirical rejection rates of tests* (3.6) *and* (2.7) *for the hypotheses* (4.4). *The mean function is given by* (4.3), *where different values for the threshold Δ and different error processes are considered. The variance is $\tilde{\sigma}_0^2(t) = 1$, the sample size is $n = 500$ and the line in boldface corresponds to the boundary of the hypotheses.*

We observe that the test (2.7) is conservative at the boundary of the hypotheses. As a consequence the proposed test (3.6) based on self-normalization is usually more powerful.

4.2. **Case Study.** Time series with possibly smoothly varying 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 in eight different places in Australia. At each station we tested for relevant deviations of the temperature from the mean temperature calculated for an historic reference period ranging from the late 19th century to 1925 at that station. As a threshold Δ, we chose 0.25, 0.5 and 0.75 degrees Celsius. Exemplary, the observed temperature curves at the weather station in Cape Otway, Gayndah and Melbourne and

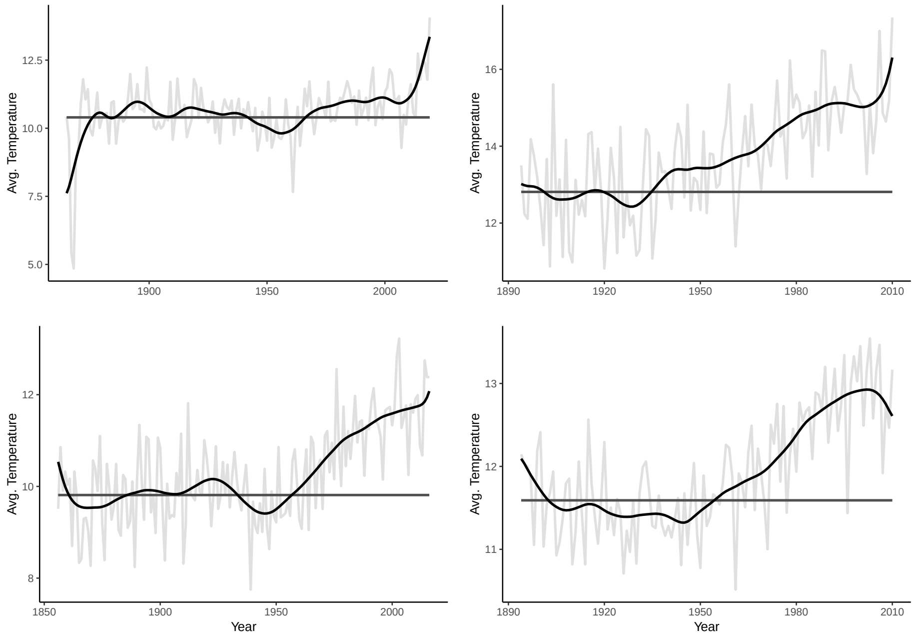

FIGURE 2. *Raw data of the temperature (light grey), estimated benchmark (dark grey) and estimated smooth mean function for different weather stations. Top left: Cape Otway. Top right: Gayndah. Bottom left: Melbourne. Bottom right: Australia (mean).*

the mean over all weather stations are plotted in Figure 2, alongside with their estimated smooth mean curves $\tilde{\mu}$ and the estimated benchmarks $\hat{g}$.

The results for all stations under consideration can be found in Table 5. For test (3.6), most p-values are significant for $\Delta = 0.25$ degrees Celsius. Further, two p-values for $\Delta = 0.5$ are significant. The test (2.7) does not yield a significant p-value below 0.05 at any station. Test (3.6) based on the proposed self-normalization procedure seems to be more powerful than (2.7), which confirms the numerical findings of the simulation study.

Acknowledgements This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Project A1, C1) of the German Research Foundation (DFG).

References

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.

Test	(3.6)			(2.7)		
Δ	0.25	0.5	0.75	0.25	0.5	0.75
Boulia Airport	**4.2**	7.8	24.2	22.1	28.5	40.8
Cape Otway Lighthouse	7.3	99.3	100.0	41.6	70.1	96.1
Gayndah Post Office	**0.7**	**1.5**	8.4	11.7	18.3	33.3
Gunnedah Pool	**0.1**	**0.5**	11.8	13.5	22.2	41.9
Hobart	5.9	53.7	98.6	25.5	51.1	87.9
Melbourne	**2.3**	59.1	99.7	29.5	51.5	84.1
Robe	48.8	99.6	100.0	49.7	85.8	99.8
Sydney	**2.3**	98.4	100.0	39.0	62.6	90.6
Australia (mean)	**0.4**	99.8	100.0	37.4	65.5	94.5

TABLE 5. *p-values of tests (3.6) and (2.7) for the respective null hypotheses in percent. Significant p-values (below 0.05) are in boldface.*

Bücher, A., H. Dette, and F. Heinrichs (2020). Are deviations in a gradually varying mean relevant? a testing approach based on sup-norm estimators. *arXiv preprint arXiv:2002.06143*.

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

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 (2020). Multiscale change point detection for dependent data. *To appear in: Scandinavian Journal of Statistics; 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.

Dette, H., W. Wu, and Z. Zhou (2019). Change point analysis of correlation in non-stationary time series. *Statist. Sinica 29*(2), 611–643.

Elsner, J. B., J. P. Kossin, and T. H. Jagger (2008). The increasing intensity of the strongest tropical cyclones. *Nature 455*(7209), 92.

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

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.

Guillaumin, A. P., A. M. Sykulski, S. C. Olhede, J. J. Early, and J. M. Lilly (2017). Analysis of non-stationary modulated time series with applications to oceanographic surface flow measurements. *Journal of Time Series Analysis 38*(5), 668–710.

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., Z. Horváth, and M. Hušková (2008). Ratio tests for change point detection. In N. Balakrishnan, E. Peña, and M. J. Silvapulle (Eds.), *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen*, Volume 1, pp. 293–304. Beachwood, Ohio, USA: Institute of Mathematical Statistics.

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.

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.

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

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.

Rho, Y. and X. Shao (2015). Inference for time series regression models with weakly dependent and heteroscedastic errors. *Journal of Business & Economic Statistics 33*(3), 444–457.

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

Shao, X. (2010). A self-normalized approach to confidence interval construction in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72*(3), 343–366.

Shao, X. (2015). Self-normalization for time series: A review of recent developments. *Journal of the American Statistical Association 110*(512), 1797–1817.

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.

Stărică, C. and C. Granger (2005). Nonstationarities in stock returns. *Review of Economics and Statistics 87*(3), 503–522.

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.

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

Wu, W. B. and M. Pourahmadi (2009). Banding sample autocovariance matrices of stationary processes. *Statistica Sinica*, 1755–1768.

Wu, W. B. and Z. Zhou (2011). Gaussian approximations for non-stationary multiple time series. *Statistica Sinica 21*(3), 1397–1413.

Zhao, Z. and X. Li (2013). Inference for modulated stationary processes. *Bernoulli: official journal of the Bernoulli Society for Mathematical Statistics and Probability 19*(1), 205.

Zhou, Z. and W. B. Wu (2009). Local linear quantile estimation for nonstationary time series. *Ann. Statist. 37*(5B), 2696–2729.

Appendix A. Proofs of main results

In this section we will provide proofs of the theoretical statements in this paper. We begin with some preliminary results regarding the uniform approximation of the sequential estimators of the regression function, which are of own interest and required for the proofs of the main results in Section 3, which will be given in Section A.3 and A.2.

A.1. Sequential estimators of the regression function.

Recall the definition of the sequential Jackknife estimator $\tilde{\mu}_{h_n}(\lambda, t)$ in (3.2) and define

$$K^*(x) = 2\sqrt{2}K(\sqrt{2}x) - K(x) \tag{A.1}$$

as the corresponding kernel. The following two results provide stochastic expansions for the difference $\tilde{\mu}_{h_n} - \mu$ uniformly with respect to λ and t. Lemma A.1 considers the case where the argument t stays away from the boundary, while a stochastic expansion for the other case is derived in Lemma A.2 below.

Lemma A.1. *Let $h_n \to 0$ and $b_n \to \infty$ be sequences with $b_n = o(nh_n)$ and define $I_n = [h_n, 1 - h_n]$. If Assumptions 2.1, 2.2 and 2.3 are satisfied, we have*

$$\sup_{t \in I_n, \lambda \in [\zeta, 1]} \Big| \lambda(\tilde{\mu}_{h_n}(\lambda, t) - \mu(t)) - \frac{1}{nh_n} \sum_{i=1}^{\lfloor \lambda n \rfloor} \varepsilon_{T_i,n} K^*_{h_n}(T_i/n - t) \Big| = \mathcal{O}(h_n^3 + \tfrac{b_n}{nh_n}),$$

where $K^*_{h_n}(\cdot) = K^*(\cdot/h_n)$ and K^* is defined in (A.1)

Proof. Define

$$S_{n,j}(\lambda, t) = \sum_{i=1}^{\lfloor \lambda n \rfloor} (\tfrac{T_i - nt}{nh_n})^j K_{h_n}(\tfrac{T_i}{n} - t) \quad \text{and} \quad R_{n,j}(\lambda, t) = \sum_{i=1}^{\lfloor \lambda n \rfloor} X_{T_i,n}(\tfrac{T_i - nt}{nh_n})^j K_{h_n}(\tfrac{T_i}{n} - t),$$

for $j \in \{0, 1, 2\}$. Note that for the calculation of the local linear estimator $\hat{\mu}_{h_n}(\lambda, t)$ in (3.1) we have to minimize the function

$$f(b_0, b_1) = \sum_{i=1}^{\lfloor \lambda n \rfloor} \big(X_{T_i,n} - b_0 - b_1(T_i/n - t)\big)^2 K_{h_n}(T_i/n - t),$$

which is differentiable with partial derivatives

$$\frac{\partial f}{\partial b_j}(b_0, b_1) = -2h_n^j\big(R_{n,j}(\lambda, t) - b_0 S_{n,j}(\lambda, t) - b_1 h_n S_{n,j+1}(\lambda, t)\big),$$

for $j \in \{0, 1\}$ and Hessian matrix

$$\mathbf{H}_f = 2 \begin{pmatrix} S_{n,0}(\lambda, t) & h_n S_{n,1}(\lambda, t) \\ h_n S_{n,1}(\lambda, t) & h_n^2 S_{n,2}(\lambda, t) \end{pmatrix}. \tag{A.2}$$

In the following discussion we will show that

$$\sup_{\lambda \in [\zeta, 1]} \Big| \frac{1}{nh_n} S_{n,j}(\lambda, t) - \lambda \int_{-1}^{1} x^j K(x) \mathrm{d}x \Big| = \mathcal{O}(\tfrac{b_n}{nh_n}), \tag{A.3}$$

for $j \in \{0,1,2\}$. If this result is true, the proof follows by arguments similar to those used in the proof of Lemma B.1 of Dette et al. (2019). To be precise, note that

$$S_{n,0}(\lambda,t)S_{n,2}(\lambda,t) - S_{n,1}^2(\lambda,t) > 0$$

for any $\lambda \in [\zeta,1]$ and almost every $n \in \mathbb{N}$. This means, that the Hessian matrix $\mathbf{H}_f$ is positive definite and both partial derivatives vanish if any only if

$$\begin{pmatrix} b_0 \\ b_1 \end{pmatrix} = \begin{pmatrix} S_{n,0}(\lambda,t) & h_n S_{n,1}(\lambda,t) \\ S_{n,1}(\lambda,t) & h_n S_{n,2}(\lambda,t) \end{pmatrix}^{-1} \begin{pmatrix} R_{n,0}(\lambda,t) \\ R_{n,1}(\lambda,t) \end{pmatrix}. \tag{A.4}$$

By (A.3) and a Taylor expansion, it follows that

$$\sup_{t\in I_n,\lambda\in[\zeta,1]} \Big| \lambda(\hat{\mu}_{h_n}(\lambda,t) - \mu(t)) - \frac{1}{nh_n}\sum_{i=1}^{\lfloor \lambda n \rfloor} \varepsilon_{T_i,n} K_{h_n}(T_i/n - t) - \frac{\lambda}{2} h_n^2 \mu''(t)\int_{-1}^1 x^2 K(x)\mathrm{d}x \Big| = \mathcal{O}(h_n^3 + \tfrac{b_n}{nh_n}).$$

The statement of Lemma A.1 now follows from the definition of the Jackknife estimator $\tilde{\mu}_{h_n}$ in (3.2).

To finish the proof, we now show the remaining estimate (A.3). For this purpose let $\lambda \in [\zeta,1]$ and define $B(\lambda) = \bigcup_{j=1}^{\ell_n} B_j(\lambda)$, where the sets $B_j(\lambda)$ are defined by

$$B_j(\lambda) = \Big\{ (j-1)b_n + 1, \dots, (j-1)b_n + \lfloor \tfrac{\lambda n - 1}{\ell_n} \rfloor + \mathbb{1}(j - 1 \le \lfloor \lambda n \rfloor - 1 \mod \ell_n) \Big\}.$$

Note that (by Assumption 2.2) the kernel K is Lipschitz continuous with support $[-1,1]$, which implies

$$K\big(\tfrac{i-nt}{nh_n}\big) = \begin{cases} K\big(\tfrac{jb_n - nt}{nh_n}\big) + \mathcal{O}\big(\tfrac{b_n}{nh_n}\big) & \text{if } |i - nt| \le nh_n, \\ 0 & \text{else,} \end{cases}$$

for $i \in B_j(\lambda)$, where the error term $\mathcal{O}(\frac{b_n}{nh_n})$ only depends on the function K and, in particular, does not depend on λ. Thus, if follows that

$$\begin{aligned} S_{n,k}(\lambda,t) &= \sum_{i=1}^{\lfloor \lambda n \rfloor} \big(\tfrac{T_i - nt}{nh_n}\big)^k K_{h_n}\big(\tfrac{T_i}{n} - t\big) = \sum_{i\in B(\lambda)} \big(\tfrac{i-nt}{nh_n}\big)^j K_{h_n}\big(\tfrac{i}{n} - t\big) \\ &= \sum_{j=1}^{\ell_n} \sum_{i \in B_j(\lambda)} \big(\tfrac{i-nt}{nh_n}\big)^k K_{h_n}\big(\tfrac{i}{n} - t\big) \\ &= \sum_{j=1}^{\ell_n} \Big(\lfloor \tfrac{\lambda n - 1}{\ell_n} \rfloor + \mathbb{1}(j-1 \le \lfloor \lambda n \rfloor - 1 \mod \ell_n) \Big) \big(\tfrac{jb_n - nt}{nh_n}\big)^k K_{h_n}\big(\tfrac{jb_n}{n} - t\big) + \mathcal{O}(b_n), \end{aligned}$$

uniformly in λ. As the kernel K has support $[-1,1]$ the only non-zero summands in the last expression are those with index $j \in [\frac{n(t-h_n)}{b_n}, \frac{n(t+h_n)}{b_n}]$. Moreover, it holds that

$$\sup_{\lambda\in[\zeta,1]} \Big| \frac{\lambda}{\ell_n h_n} \sum_{j=-\lfloor \ell_n h_n \rfloor}^{\lfloor \ell_n h_n \rfloor} \big(\tfrac{jb_n}{nh_n}\big)^j K\big(\tfrac{jb_n}{nh_n}\big) - \lambda \int_{-1}^1 x^j K(x)\mathrm{d}x \Big| = \mathcal{O}\big(\tfrac{b_n}{nh_n}\big),$$

which implies (A.3). □

Lemma A.2. *Let* $I_n^c = [0,1] \setminus I_n$*,* $\kappa_{j,h_n}(t) = \int_{-t/h_n}^{(1-t)/h_n} x^j K(x)\mathrm{d}x$*, for* $j \in \mathbb{N}_0$*, and*

$$c_{h_n,i}(t) = \frac{\kappa_{2,h_n}(t) - \kappa_{1,h_n}(t)\left(\frac{T_i - nt}{nh_n}\right)}{\kappa_{0,h_n}(t)\kappa_{2,h_n}(t) - \kappa_{1,h_n}^2(t)},$$

for $i \in \{1, \dots, n\}$*. If the assumptions of Lemma A.1 are satisfied, it holds*

$$\sup_{\substack{t \in I_n^c, \\ \lambda \in [\zeta,1]}} \left| \lambda(\tilde{\mu}_{h_n}(\lambda,t) - \mu(t)) - \frac{1}{nh_n} \sum_{i=1}^{\lfloor \lambda n \rfloor} \varepsilon_{T_i,n} \left\{ 2c_{\frac{h_n}{\sqrt{2}},i}(t) K_{\frac{h_n}{\sqrt{2}}}\left(\tfrac{T_i}{n} - t\right) - c_{h_n,i}(t) K_{h_n}\left(\tfrac{T_i}{n} - t\right) \right\} \right|$$
$$= \mathcal{O}\left(h_n^2 + \tfrac{b_n}{nh_n}\right). \quad \text{(A.5)}$$

Proof. By similar arguments as given for the approximation in (A.3) it follows that

$$\sup_{t \in I_n^c} \sup_{\lambda \in [\zeta,1]} \left| \frac{1}{nh_n} S_{n,j}(\lambda,t) - \lambda \kappa_{j,h_n}(t) \right| = \mathcal{O}\left(\tfrac{b_n}{nh_n}\right), \quad \text{(A.6)}$$

for $j \in \{0,1,2\}$. Note that $S_{n,0}(\lambda,t)S_{n,2}(\lambda,t) - S_{n,1}^2(\lambda,t) > 0$ for any $\lambda \in [\zeta,1]$ and almost every $n \in \mathbb{N}$ since, by Assumption 2.2,

$$\int_{\mathcal{A}} K(x)\mathrm{d}x \int_{\mathcal{A}} x^2 K(x)\mathrm{d}x - \left(\int_{\mathcal{A}} xK(x)\mathrm{d}x \right)^2 = \frac{1}{2} \int_{\mathcal{A}^2} (x-y)^2 K(x)K(y)\mathrm{d}(x,y),$$

is positive for any set $\mathcal{A} \subset [-1,1]$ with positive Lebesgue measure. Thus, the Hessian matrix $\mathbf{H}_f$, as defined in (A.2), is positive definite and the partial derivatives vanish if and only if (A.4) holds true. Therefore, by (A.6) and similar arguments as given in the proof of Lemma B.2 in Dette et al. (2019) we obtain that

$$\sup_{t \in I_n^c, \lambda \in [\zeta,1]} \left| \lambda(\hat{\mu}_{h_n}(\lambda,t) - \mu(t)) - \frac{1}{nh_n} \sum_{i=1}^{\lfloor \lambda n \rfloor} \varepsilon_{T_i,n} c_{h_n,i}(t) K_{h_n}^*(T_i/n - t) \right| = \mathcal{O}\left(h_n^2 + \tfrac{b_n}{nh_n}\right)$$

Finally the statement (A.5) follows from the definition of the Jackknife estimator in (3.2).

□

A.2. **Proof of Theorem 3.4.** The following two Lemmas A.3 and A.4 establish the convergence of the finite dimensional distributions and equicontinuity of the process $\{G_n(\lambda)\}_{\lambda \in [\zeta,1]}$ in (3.4) (for any $\zeta > 0$). The assertion of Theorem 3.4 then follows directly from Theorems 1.5.4 and 1.5.7 of van der Vaart and Wellner (1996).

Lemma A.3. *Let Assumptions 2.1, 2.2, 2.3, 3.1 and 3.2 be satisfied and* $\zeta \le \lambda_1 \le \dots \le \lambda_p \le 1$*. Then,*

$$\left(G_n(\lambda_1), \dots, G_n(\lambda_p)\right)^\top \rightsquigarrow \left(G(\lambda_1), \dots, G(\lambda_p)\right)^\top$$

in $\mathbb{R}^p$*, where the Gaussian process* $\{G(\lambda)\}_{\lambda \in [\xi,1]}$ *is defined in* (3.5)

Proof. First observe that

$$G_n(\lambda) = \lambda\sqrt{n}(\|\hat{d}_n(\lambda,\cdot) - d\|_{2,\tau}^2 + 2\langle d, \hat{d}_n(\lambda,\cdot) - d\rangle_\tau),$$

and note that

$$\hat{g}_n(\lambda) - g(\mu) = \mathcal{O}_{\mathbb{P}}(n^{-1/2}) \tag{A.7}$$

by Assumptions 2.3 and 3.1. Recall the definition of the interval $I_n = [h_n, 1-h_n]$ and denote $\langle f, g\rangle_{I_n} = \int_{I_n} f(x)g(x)\tau(\mathrm{d}x)$ and $\|f\|_{2,I_n} = \langle f, f\rangle_{I_n}^{1/2}$, for any $f, g \in L^2([0,1],\tau)$. In the following let

$$\tilde{G}_n(\lambda) = \lambda\sqrt{n}(\|\hat{d}_n(\lambda,\cdot) - d\|_{2,I_n}^2 + 2\langle d, \hat{d}_n(\lambda,\cdot) - d\rangle_{I_n})\ , \tag{A.8}$$

then the assertion follows from the statements

$$(\tilde{G}_n(\lambda_1), \ldots, \tilde{G}_n(\lambda_p))^\top \rightsquigarrow (G(\lambda_1), \ldots, G(\lambda_p))^\top \tag{A.9}$$

$$\sup_{\lambda\in[\zeta,1]} |\tilde{G}_n(\lambda) - G_n(\lambda)| = o_{\mathbb{P}}(1) \tag{A.10}$$

For a proof of (A.9) note that by the Cramér-Wold device, it is sufficient to prove

$$\sum_{i=1}^p a_i\tilde{G}_n(\lambda_i) \rightsquigarrow \sum_{i=1}^p a_i G(\lambda_i)$$

for all $a_1, \ldots, a_p \in \mathbb{R}$. Define $S_k = \sum_{i=1}^k \varepsilon_{i,n}$ and $\tilde{S}_k = \sum_{i=1}^k \tilde{\varepsilon}_{i,n}$ with

$$\tilde{\varepsilon}_{i,n} = \mathbb{E}[\varepsilon_{i,n}|\eta_i, \ldots, \eta_{i-m_n}] \tag{A.11}$$

for $m_n \in \mathbb{N}$ and $k = 1, \ldots, n$. By Assumption 2.3 (1), Assumption 3.2 and equation (3.2) of Wu and Zhou (2011) we have

$$\|\max_{1\le k\le n} |\tilde{S}_k - S_k|\|_{8,\Omega} = \mathcal{O}(n^{1/2}m_n^{-c}), \tag{A.12}$$

where $m_n = n^\gamma$ with $\gamma < \min\{\beta/4, \alpha/2\}$ and the constant c is given by $c = \frac{5}{6\gamma}$. In particular, the sequences $\frac{m_n^4}{nh_n^4}$, $\frac{m_n^2}{b_n}$ and $\frac{b_n n^{1/2}}{m_n^c}$ are all of order $o(1)$ as n tends to infinity. From Lemma A.1 it follows that

$$\sup_{t\in[h_n,1-h_n],\lambda\in[\zeta,1]} \Big|\lambda(\tilde{\mu}_{h_n}(\lambda,t) - \mu(t)) - \frac{1}{nh_n}\sum_{i=1}^{\lfloor\lambda n\rfloor} \varepsilon_{T_i,n}K_{h_n}^*(T_i/n - t)\Big| = \mathcal{O}(h_n^3 + \tfrac{b_n}{nh_n}). \tag{A.13}$$

Observe that by Assumption 3.1, (A.7), (A.12) and (A.13)

$$\begin{aligned}
\lambda\sqrt{n}\|\hat{d}_n(\lambda,\cdot) - d\|_{2,I_n}^2 &= \frac{\sqrt{n}}{\lambda}\Big\|\Big(\frac{1}{n}\sum_{j=1}^{\lfloor\lambda n\rfloor} \varepsilon_{T_j,n}\frac{1}{h_n}K_{h_n}^*(T_j/n - \cdot)\Big) + \lambda(\hat{g}_n(\lambda) - g(\mu))\Big\|_{2,I_n}^2 + o_{\mathbb{P}}(1) \\
\le\ & 2\frac{\sqrt{n}}{\lambda}\Big\|\frac{1}{n}\sum_{j=1}^{\lfloor\lambda n\rfloor} \varepsilon_{T_j,n}\frac{1}{h_n}K_{h_n}^*(T_j/n - \cdot)\Big\|_{2,I_n}^2 + 2\sqrt{n}\lambda(\hat{g}_n(\lambda) - g(\mu))^2 + o_{\mathbb{P}}(1) \\
=\ & 2\frac{\sqrt{n}}{\lambda}\Big\|\frac{1}{n}\sum_{j=1}^{\lfloor\lambda n\rfloor} \tilde{\varepsilon}_{T_j,n}\frac{1}{h_n}K_{h_n}^*(T_j/n - \cdot)\Big\|_{2,I_n}^2 + o_{\mathbb{P}}(1)
\end{aligned} \tag{A.14}$$

uniformly with respect to $\lambda \in [\zeta, 1]$, where the random variables $\tilde{\varepsilon}_{T_j,n}$ are defined in (A.11) and m_n-dependent in the sense that $\tilde{\varepsilon}_{T_{j_1},n}$ and $\tilde{\varepsilon}_{T_{j_2},n}$ are independent if $|T_{j_1} - T_{j_2}| > m_n$. For the estimation of the first term, let $K_{i,j}$ denote the integral $\int_{I_n} K^*_{h_n}(T_i/n - x)K^*_{h_n}(T_j/n - x)\tau(\mathrm{d}x)$. By the Cauchy-Schwarz inequality and absolute continuity of τ, $K_{i,j}$ can be bounded from above by Ch_n. Tith implies

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{\lambda\in[\zeta,1]} n^2 \Big\|\frac{1}{nh_n}\sum_{j=1}^{\lfloor \lambda n\rfloor} \tilde{\varepsilon}_{T_j,n} K^*_{h_n}(T_j/n - \cdot)\Big\|^8_{2,I_n}\Big] \\
&\le \sum_{k=1}^{n} \frac{n^2}{n^8 h_n^8}\mathbb{E}\Big[\Big\|\sum_{j=1}^{k} \tilde{\varepsilon}_{T_j,n} K^*_{h_n}(T_j/n - \cdot)\Big\|^8_{2,I_n}\Big] \qquad \text{(A.15)}\\
&= \sum_{k=1}^{n} \frac{1}{n^6 h_n^8}\sum_{j_1,\dots,j_8=1}^{k} \mathbb{E}\Big[\prod_{r=1}^{8} \tilde{\varepsilon}_{T_{j_r},n}\Big] K_{j_1,j_2}K_{j_3,j_4}K_{j_5,j_6}K_{j_7,j_8} = \mathcal{O}\big(\tfrac{m_n^4}{nh_n^4}\big),
\end{aligned}$$

where the last estimate follows observing $\max_{1\le i\le n}\mathbb{E}\varepsilon^8_{i,n} < \infty$ by Assumption 2.3 (4) and the fact that only $\mathcal{O}(n^4 m_n^4)$ summands of the inner sum are non-zero due to the m_n-dependency of the random variables $\tilde{\varepsilon}_{j,n}$. Thus, by (A.14),

$$\lambda\sqrt{n}\|\hat{d}_n(\lambda,\cdot) - d\|^2_{2,I_n} = o_{\mathbb{P}}(1) \qquad \text{(A.16)}$$

uniformly with respect to $\lambda \in [\zeta, 1]$. If

$$d_0 = \int_0^1 d^2(x)dx = 0,$$

it follows that $G_n(\lambda) = o_{\mathbb{P}}(1)$ and therefore we assume $d_0 > 0$ in the following discussion. In this case we have from (A.8) and (A.13) that

$$\sum_{i=1}^{p} a_i \tilde{G}_n(\lambda_i) = 2\sum_{i=1}^{p} a_i\lambda_i\sqrt{n}\langle d, \hat{d}_n(\lambda,\cdot) - d\rangle_{I_n} + o_{\mathbb{P}}(1) = Z_n + o_{\mathbb{P}}(1), \qquad \text{(A.17)}$$

where

$$Z_n = \frac{2}{\sqrt{n}}\sum_{i=1}^{p} a_i \sum_{j=1}^{\lfloor \lambda_i n\rfloor} \tilde{\varepsilon}_{T_j,n}\langle d, h_n^{-1}K^*_{h_n}(T_j/n - \cdot) + \omega_n(T_j/n)\rangle_{I_n}.$$

By Lipschitz continuity of d and $\mathrm{supp}(K) = [-1, 1]$ it follows that

$$\int_{I_n} d(x)K^*_{h_n}(T_j/n - x)\tau(\mathrm{d}x) = \Big(d(\tfrac{T_j}{n}) + \mathcal{O}(h_n)\Big)\int_{I_n} K^*_{h_n}(T_j/n - x)\tau(\mathrm{d}x).$$

We obtain for any point of continuity y of the piecewise continuous density f_τ of the measure τ that

$$\begin{aligned}
\frac{1}{h_n}\int_{I_n} K^*_{h_n}(y - x)\tau(\mathrm{d}x) &= \frac{1}{h_n}\int_{h_n}^{1-h_n} K^*_{h_n}(y - x)f_\tau(x)\mathrm{d}x \\
&= \int_{1-y/h_n}^{1/h_n - 1 - y/h_n} K^*(x)f_\tau(xh_n + y)\mathrm{d}x \\
&= \begin{cases} f_\tau(y) + o(1), & \text{if } y \in [2h_n, 1 - 2h_n], \\ \mathcal{O}(1), & \text{else.}\end{cases}
\end{aligned}$$

Therefore, for $T_j \in \{2nh_n, \dots, n-2nh_n\}$, it holds

$$\frac{1}{h_n}\int_{I_n} d(x)K^*_{h_n}(T_j/n-x)\tau(\mathrm{d}x) = f_\tau(T_j/n)d(T_j/n) + o(1), \tag{A.18}$$

which leads to

$$Z_n = \sum_{j=1}^n Y_j + o_{\mathbb{P}}(1) \tag{A.19}$$

where the random variables $Y_1, \dots, Y_n$ are defined by

$$Y_j = \frac{2}{\sqrt{n}}\Big(\sum_{i=1}^p a_i \mathbb{1}(j \le \lfloor \lambda_i n \rfloor)\Big)\tilde{\varepsilon}_{T_j,n} d_{\omega_n}(T_j/n) \qquad (j = 1, \dots, n)$$

and

$$d_{\omega_n}(T_j/n) = f_\tau(T_j/n)d(T_j/n) + \omega_n(T_j/n)\int_0^1 d(x)\tau(\mathrm{d}x).$$

Observe that $Y_1, \dots, Y_n$ centred and m_n-dependent random variables in the sense that Y_{j_1} and Y_{j_2} are independent if $|T_{j_1} - T_{j_2}| > m_n$. Define the big blocks $B_j = \{k \in \mathbb{N} : (j-1)b_n + 1 \le k \le jb_n - m_n\}$ and the small blocks $S_j = \{k \in \mathbb{N} : jb_n - m_n + 1 \le k \le jb_n\}$, for $j = 1 \dots, \ell_n$, and the remainder $R = \{k \in \mathbb{N} : \ell_n b_n + 1 \le k \le n\}$. In the following, we will show that the small blocks and the remainder are negligible and the asymptotic behaviour of Z_n is determined by the big blocks. First observe that $\tilde{\varepsilon}_{T_{k_1},n} \in S_{j_1}$ and $\tilde{\varepsilon}_{T_{k_2},n} \in S_{j_2}$ are independent for $j_1 \ne j_2$. Thus,

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\sum_{j=1}^{\ell_n}\sum_{k:T_k\in S_j} Y_k\Big)^2\Big] \\
&= \sum_{j=1}^{\ell_n}\sum_{k_1:T_{k_1}\in S_j}\sum_{k_2:T_{k_2}\in S_j}\mathbb{E}[Y_{k_1}Y_{k_2}] \\
&= \frac{4}{n}\sum_{j=1}^{\ell_n}\sum_{i_1,i_2=1}^{p} a_{i_1}a_{i_2}\sum_{k_1=1}^{\lfloor\lambda_{i_1}n\rfloor}\sum_{k_2=1}^{\lfloor\lambda_{i_2}n\rfloor}\mathbb{1}(T_{k_1}, T_{k_2} \in S_j)\mathbb{E}[\tilde{\varepsilon}_{T_{k_1},n}\tilde{\varepsilon}_{T_{k_2},n}]d_{\omega_n}(T_{k_1}/n)d_{\omega_n}(T_{k_2}/n).
\end{aligned} \tag{A.20}$$

Further, $T_k \in S_j$ for some $k \le \lfloor \lambda n \rfloor$, if and only if $k = r\ell_n + j$ for some $r \ge b_n - m_n$ and $r \le \lfloor \frac{\lambda n - j}{\ell_n} \rfloor$ and we obtain

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\sum_{j=1}^{\ell_n}\sum_{k:T_k\in S_j} Y_k\Big)^2\Big] \\
&= \sum_{i_1,i_2=1}^{p} a_{i_1}a_{i_2}\frac{4}{n}\sum_{j=1}^{\ell_n}\sum_{r_1=b_n-m_n}^{\lfloor\frac{\lambda_{i_1}n-j}{\ell_n}\rfloor}\sum_{r_2=b_n-m_n}^{\lfloor\frac{\lambda_{i_2}n-j}{\ell_n}\rfloor} d_{\omega_n}\big(\tfrac{(j-1)b_n+r_1+1}{n}\big)d_{\omega_n}\big(\tfrac{(j-1)b_n+r_2+1}{n}\big) \\
&\qquad \times \mathbb{E}[\tilde{\varepsilon}_{(j-1)b_n+r_1+1,n}\tilde{\varepsilon}_{(j-1)b_n+r_2+1,n}].
\end{aligned}$$

For $\lambda < 1$, it holds that $\lambda \frac{n}{b_n \ell_n} \to \lambda$ and $1 - \frac{m_n}{b_n} \to 1$, so for almost every $n \in \mathbb{N}$, $b_n - m_n \geq \lfloor \frac{\lambda n - j}{\ell_n} \rfloor$. Thus, if $\lambda_{i_1} < 1$ or $\lambda_{i_2} < 1$, the sums indexed by r_1 and r_2 on the right-hand side of the previous display are empty sums for almost every $n \in \mathbb{N}$. For $\lambda = 1$, there are m_n summands in both sums, thus, the right-hand side of the previous display is of order $\mathcal{O}(m_n^2/b_n)$ which vanishes by assumption. Thus the small blocks are asymptotically negligible, and analogously,

$$\mathbb{E}\Big[\Big(\sum_{k \in \bar{R}} Y_k\Big)^2\Big] = \mathcal{O}\Big(\frac{b_n^2}{n}\Big).$$

The sums over the big blocks are independent, and we have analogously to (A.20),

$$\sum_{j=1}^{\ell_n} \mathbb{E}\Big[\Big(\sum_{k: T_k \in B_j} Y_k\Big)^2\Big] = \sum_{i_1, i_2 = 1}^{p} a_{i_1} a_{i_2} \frac{4}{n} \sum_{j=1}^{\ell_n} \sum_{r_1 \in \bar{B}_{i_1, j}} \sum_{r_2 \in \bar{B}_{i_2, j}} d_{\omega_n}(\tfrac{r_1}{n}) d_{\omega_n}(\tfrac{r_2}{n}) \mathbb{E}[\tilde{\varepsilon}_{r_1, n} \tilde{\varepsilon}_{r_2, n}]$$

where

$$\bar{B}_{i,j} = \{(j-1)b_n + 1, \ldots, (j-1)b_n + 1 + \lfloor \tfrac{\lambda_i n - j}{\ell_n} \rfloor \wedge (b_n - m_n - 1)\},$$

for $\lambda_i \in [\zeta, 1]$. Note that, for almost every $n \in \mathbb{N}$, $\bar{B}_{i,j} = \{(j-1)b_n + 1, \ldots, (j-1)b_n + 1 + \lfloor \frac{\lambda_i n - j}{\ell_n} \rfloor\}$, if $\lambda < 1$ and $\bar{B}_{i,j} = \{(j-1)b_n + 1, \ldots, j b_n - m_n\}$, if $\lambda = 1$. By Lipschitz continuity of d and Assumption 3.1,

$$\sum_{j=1}^{\ell_n} \mathbb{E}\Big[\Big(\sum_{k: T_k \in B_j} Y_k\Big)^2\Big] = \sum_{i_1, i_2 = 1}^{p} a_{i_1} a_{i_2} \frac{4}{n} \sum_{j=1}^{\ell_n} d_{\omega_n}^2(\tfrac{j b_n}{n}) \sum_{r_1 \in \bar{B}_{i_1, j}} \sum_{r_2 \in \bar{B}_{i_2, j}} \mathbb{E}[\tilde{\varepsilon}_{r_1, n} \tilde{\varepsilon}_{r_2, n}] + \mathcal{O}(\tfrac{b_n^2}{n h_n}). \tag{A.21}$$

Now, by (A.12),

$$\max_{1 \leq r_1, r_2 \leq n} \big|\mathbb{E}[\tilde{\varepsilon}_{r_1, n} \tilde{\varepsilon}_{r_2, n}] - \mathbb{E}[\varepsilon_{r_1, n} \varepsilon_{r_2, n}]\big| = \mathcal{O}(n^{1/2} m_n^{-c}). \tag{A.22}$$

Applying Assumption 2.3 (2), yields

$$\mathbb{E}[\varepsilon_{r_1, n} \varepsilon_{r_2, n}] = \mathbb{E}\big[G(\tfrac{j b_n}{n}, \mathcal{F}_{r_1}) G(\tfrac{j b_n}{n}, \mathcal{F}_{r_2})\big] + \mathcal{O}(b_n/n),$$

for any $r_1 \in \bar{B}_{i_1, j}, r_2 \in \bar{B}_{i_2, j}$. By the same arguments as in the proof of Theorem 1 in Wu and Pourahmadi (2009) and Assumption 2.3 (1), it follows that

$$\mathbb{E}\big[G(\tfrac{i}{n}, \mathcal{F}_{r_1}) G(\tfrac{i}{n}, \mathcal{F}_{r_2})\big] = \mathcal{O}(\gamma^{|r_2 - r_1|}),$$

for any $1 \leq i \leq n$, in particular $i = j b_n$. Let $b := |\bar{B}_{i_1, j} \cap \bar{B}_{i_2, j}|$, then,

$$\begin{aligned}\sum_{r_1 \in \bar{B}_{i_1, j}} \sum_{r_2 \in \bar{B}_{i_2, j}} \mathbb{E}[\varepsilon_{r_1, n} \varepsilon_{r_2, n}] &= b \sum_{k=-b}^{b} (1 - \tfrac{|k|}{b}) \mathbb{E}\big[G(\tfrac{j b_n}{n}, \mathcal{F}_0) G(\tfrac{j b_n}{n}, \mathcal{F}_k)\big] + \mathcal{O}(b_n^3/n + b_n \gamma^{b_n} + 1) \\ &= (\lambda_{i_1} \wedge \lambda_{i_2}) b_n \sigma^2\Big(\frac{j b_n}{n}\Big) + \mathcal{O}(b_n^3/n + b_n \gamma^{b_n} + 1).\end{aligned} \tag{A.23}$$

Thus, by (A.22),

$$\sum_{r_1\in\bar{B}_{i_1,j}}\sum_{r_2\in\bar{B}_{i_2,j}}\mathbb{E}[\tilde{\varepsilon}_{r_1,n}\tilde{\varepsilon}_{r_2,n}]=(\lambda_{i_1}\wedge\lambda_{i_2})b_n\sigma^2\Big(\frac{jb_n}{n}\Big)+\mathcal{O}\Big(\frac{b_n^3}{n}+b_n\gamma^{b_n}+1+\frac{b_n^2n^{1/2}}{m_n^c}\Big).$$

Plugging this into (A.21) and observing $\ell_n=\lfloor n/b_n\rfloor$ leads to

$$\sum_{j=1}^{\ell_n}\mathbb{E}\Big[\Big(\sum_{k:T_k\in B_j}Y_k\Big)^2\Big]=\sum_{i_1,i_2=1}^{p}4a_{i_1}a_{i_2}(\lambda_{i_1}\wedge\lambda_{i_2})\|d_\omega\sigma\|_2^2+o(1) \tag{A.24}$$
$$=\mathrm{Var}\Big(\sum_{i=1}^{p}a_iG(\lambda_i)\Big)+o(1).$$

Finally, observe that by Jensen's inequality and Assumption 2.3 (4), for some constant $C>0$,

$$\sum_{j=1}^{\ell_n}\mathbb{E}\Big[\Big(\sum_{k:T_k\in B_j}Y_k\Big)^4\Big]\le\sum_{j=1}^{\ell_n}b_n^3\mathbb{E}\Big[\sum_{k:T_k\in B_j}Y_k^4\Big]\le C\frac{b_n^3}{n}\max_{k=1}^{n}\mathbb{E}\,\tilde{\varepsilon}_{T_k,n}^4=\mathcal{O}(b_n^3/n).$$

By Lyapunov's central limit theorem, it follows that

$$Z_n\rightsquigarrow\mathcal{N}\Big(0,\mathrm{Var}\Big(\sum_{i=1}^{p}a_iG(\lambda_i)\Big)\Big)\overset{\mathcal{D}}{=}\sum_{i=1}^{p}a_iG(\lambda_i)$$

and the statement (A.9) is a consequence of (A.17), (A.19) and the Cramér-Wold device. For the proof of the remaining statement (A.10) we note that this assertion is a consequence of the estimate

$$\sup_{\lambda\in[\zeta,1]}\int_{I_n^c}\lambda\sqrt{n}(\tilde{\mu}_{h_n}(\lambda,t)-\mu(t))^j\mathrm{d}t=o_{\mathbb{P}}(1)\qquad j\in\{1,2\}. \tag{A.25}$$

To prove this statement, we note that by Lemma A.2

$$\sup_{\lambda\in[\zeta,1]}\int_{I_n^c}\lambda\sqrt{n}(\tilde{\mu}_{h_n}(\lambda,t)-\mu(t))^j\mathrm{d}t=$$
$$\sup_{\lambda\in[\zeta,1]}\int_{I_n^c}\frac{1}{\lambda^{j-1}}\sqrt{n}\Big(\frac{1}{nh_n}\sum_{i=1}^{\lfloor\lambda n\rfloor}\varepsilon_{T_i,n}\Big(2c_{\frac{h_n}{\sqrt{2}},i}(t)K_{\frac{h_n}{\sqrt{2}}}(\tfrac{T_i}{n}-t)-c_{h_n,i}(t)K_{h_n}(\tfrac{T_i}{n}-t)\Big)\Big)^j\mathrm{d}t+o(1),$$

for $j\in\{1,2\}$. The case $j=2$ follows by similar arguments as given in (A.15). For the case $j=1$ recall from the previous discussion that the random variables $\varepsilon_{i,n}$ can be approximated by m_n-dependent random variables $\tilde{\varepsilon}_{i,n}$. Thus,

$$\begin{aligned}
&\sup_{\lambda\in[\zeta,1]}\int_{I_n^c}\lambda\sqrt{n}(\tilde{\mu}_{h_n}(\lambda,t)-\mu(t))\mathrm{d}t \\
&=\sup_{\lambda\in[\zeta,1]}\int_{I_n^c}\frac{1}{\sqrt{n}h_n}\sum_{i=1}^{\lfloor\lambda n\rfloor}\tilde{\varepsilon}_{T_i,n}\Big(2c_{\frac{h_n}{\sqrt{2}},i}(t)K_{\frac{h_n}{\sqrt{2}}}\big(\tfrac{T_i}{n}-t\big)-c_{h_n,i}(t)K_{h_n}\big(\tfrac{T_i}{n}-t\big)\Big)\mathrm{d}t+o_{\mathbb{P}}(1) \\
&=\sup_{\lambda\in[\zeta,1]}\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor\lambda n\rfloor}\tilde{\varepsilon}_{T_i,n}\int_{I_n^c}\frac{1}{h_n}\Big(2c_{\frac{h_n}{\sqrt{2}},i}(t)K_{\frac{h_n}{\sqrt{2}}}\big(\tfrac{T_i}{n}-t\big)-c_{h_n,i}(t)K_{h_n}\big(\tfrac{T_i}{n}-t\big)\Big)\mathrm{d}t+o_{\mathbb{P}}(1) \\
&=\sup_{\lambda\in[\zeta,1]}\frac{1}{\sqrt{n}}\sum_{j\in B}\sum_{i=1}^{\lfloor\lambda b_n\rfloor}\tilde{\varepsilon}_{(j-1)b_n+i,n} \\
&\qquad\times\int_{I_n^c}\frac{1}{h_n}\Big(2c_{\frac{h_n}{\sqrt{2}},i}(t)K_{\frac{h_n}{\sqrt{2}}}\big(\tfrac{(j-1)b_n+i}{n}-t\big)-c_{h_n,i}(t)K_{h_n}\big(\tfrac{(j-1)b_n+i}{n}-t\big)\Big)\mathrm{d}t+o_{\mathbb{P}}(1),
\end{aligned} \tag{A.26}$$

where B denotes the set $\{1,\dots,\lfloor 2\ell_n h_n\rfloor\}\cup\{\lfloor\ell_n(1-2h_n)\rfloor,\dots,\ell_n\}$. Note that the integral on the right-hand side of (A.26) is bounded and by similar arguments as used in the proof of (A.15), the right-hand side of (A.26) is of order $\mathcal{O}(b_n h_n^4 m_n^4)$, which converges to 0 by the definition of m_n and Assumption 3.2.

Therefore (A.25) follows and the proof of Lemma A.3 is completed. □

Lemma A.4. *Let Assumptions 2.1, 2.2, 2.3, 3.1 and 3.2 be satisfied. Then,*

$$\lim_{\rho\searrow 0}\lim_{n\to\infty}\mathbb{P}\big(\sup_{|\lambda_1-\lambda_2|\le\rho}|G_n(\lambda_1)-G_n(\lambda_2)|>\varepsilon\big)=0,$$

for any $\varepsilon>0$.

Proof. By (A.10), it follows that

$$G_n(\lambda_1)-G_n(\lambda_2)=\tilde{G}_n(\lambda_1)-\tilde{G}_n(\lambda_2)+o_{\mathbb{P}}(1).$$

uniformly with respect to $\lambda\in[\zeta,1]$, where $\tilde{G}_n(\lambda)$ is defined in (A.8). Therefore the assertion of the Lemma follows from

$$\lim_{\rho\searrow 0}\lim_{n\to\infty}\mathbb{P}\big(\sup_{|\lambda_1-\lambda_2|\le\rho}|\tilde{G}_n(\lambda_1)-\tilde{G}_n(\lambda_2)|>\varepsilon\big)=0, \tag{A.27}$$

To prove this statement note that we obtain from (A.16)

$$\tilde{G}_n(\lambda_1)-\tilde{G}_n(\lambda_2)=2\sqrt{n}\langle\lambda_1(\hat{d}_n(\lambda_1,\cdot)-d)-\lambda_2(\hat{d}_n(\lambda_2,\cdot)-d),d\rangle_{I_n}+o_{\mathbb{P}}(1)$$

uniformly with respect to $\lambda_1,\lambda_2\in[\zeta,1]$. By Lemma A.1, Assumption 3.1 and (A.18) we have the expansion

$$\tilde{G}_n(\lambda_1)-\tilde{G}_n(\lambda_2)=\frac{2}{\sqrt{n}}\sum_{i=\lfloor(\lambda_1\wedge\lambda_2)n\rfloor+1}^{\lfloor(\lambda_1\vee\lambda_2)n\rfloor}\varepsilon_{T_i,n}\Big\langle\frac{1}{h_n}K_{h_n}^*\Big(\frac{T_i}{n}-\cdot\Big)+\omega_n\Big(\frac{T_i}{n}\Big),d\Big\rangle_{I_n}+o_{\mathbb{P}}(1)$$

$$= \frac{2}{\sqrt{n}} \sum_{i=\lfloor(\lambda_1\wedge\lambda_2)n\rfloor+1}^{\lfloor(\lambda_1\vee\lambda_2)n\rfloor} \varepsilon_{T_i,n} d_{\omega_n}(T_i/n) + o_{\mathbb{P}}(1),$$

uniformly in $\lambda_1, \lambda_2 \in [\zeta, 1]$, where d_{ω_n} is defined in the proof of Lemma A.3. Further, recalling the definition of $\tilde{\varepsilon}_{i,n}$ in (A.11), it follows by (A.12), that

$$\tilde{G}_n(\lambda_1) - \tilde{G}_n(\lambda_2) = \frac{2}{\sqrt{n}} \sum_{i=\lfloor(\lambda_1\wedge\lambda_2)n\rfloor+1}^{\lfloor(\lambda_1\vee\lambda_2)n\rfloor} \tilde{\varepsilon}_{T_i,n} d_{\omega_n}(T_i/n) + o_{\mathbb{P}}(1)$$

uniformly in $\lambda_1, \lambda_2 \in [\zeta, 1]$. In particular, we obtain

$$\mathbb{P}\Big(\sup_{|\lambda_1-\lambda_2|\le\rho} |\tilde{G}_n(\lambda_1) - \tilde{G}_n(\lambda_2)| > \varepsilon \Big)$$
$$= \mathbb{P}\Big(\sup_{|\lambda_1-\lambda_2|\le\rho} \Big| \frac{2}{\sqrt{n}} \sum_{i=\lfloor(\lambda_1\wedge\lambda_2)n\rfloor+1}^{\lfloor(\lambda_1\vee\lambda_2)n\rfloor} \tilde{\varepsilon}_{T_i,n} d_{\omega_n}(T_i/n) \Big| > \varepsilon \Big) + o(1). \quad \text{(A.28)}$$

Now, for some $\zeta \le \lambda_1 \le \lambda_2 \le 1$, define the sets $\tilde{B}_j = \tilde{B}_j(\lambda_1, \lambda_2)$ by

$$\Big\{ i \in \mathbb{N} : (j-1)b_n + \Big\lfloor \frac{\lambda_1 n}{\ell_n} \Big\rfloor + \mathbb{1}(j < \lfloor \lambda_1 n \rfloor \mod \ell_n) + 1$$
$$\le i \le (j-1)b_n + \Big\lfloor \frac{\lambda_2 n}{\ell_n} \Big\rfloor - \mathbb{1}(\lfloor \lambda_2 n \rfloor \mod \ell_n < j) + 1 \Big\},$$

for $j = 1, \ldots, \ell_n$. In particular, $|\tilde{B}_j| \le \lfloor \frac{\lambda_2 n}{\ell_n} \rfloor - \lfloor \frac{\lambda_1 n}{\ell_n} \rfloor + 1 \le \lfloor \frac{|\lambda_2-\lambda_1|n}{\ell_n} \rfloor + 2$. With this notation, it holds

$$R_n = \mathbb{E}\Big[\Big| \frac{1}{\sqrt{n}} \sum_{i=\lfloor(\lambda_1\wedge\lambda_2)n\rfloor+1}^{\lfloor(\lambda_1\vee\lambda_2)n\rfloor} \tilde{\varepsilon}_{T_i,n} d_{\omega_n}(T_i/n) \Big|^4 \Big] = \frac{1}{n^2} \mathbb{E}\Big[\Big| \sum_{j=1}^{\ell_n} \sum_{i\in\tilde{B}_j} \tilde{\varepsilon}_{i,n} d_{\omega_n}(i/n) \Big|^4 \Big].$$

Observe that the distance between two blocks $\tilde{B}_j$ and $\tilde{B}_{j-1}$ is larger than $b_n - \Big\lfloor \frac{(\lambda_2-\lambda_1)n}{\ell_n} \Big\rfloor > m_n$. Thus, the sums over these blocks are independent and we obtain the representation

$$\begin{aligned} R_n =& \frac{1}{n^2} \sum_{j=1}^{\ell_n} \sum_{i_1,i_2,i_3,i_4\in\tilde{B}_j} \Big(\prod_{r=1}^4 d_{\omega_n}(i_r/n) \Big) \mathbb{E}\Big[\prod_{r=1}^4 \tilde{\varepsilon}_{i_r,n} \Big] \\ &+ \frac{3}{n^2} \sum_{\substack{j_1,j_2=1 \\ j_1\ne j_2}}^{\ell_n} \sum_{i_1,i_2\in\tilde{B}_{j_1}} \sum_{i_3,i_4\in\tilde{B}_{j_2}} \Big(\prod_{r=1}^4 d_{\omega_n}(i_r/n) \Big) \mathbb{E}[\tilde{\varepsilon}_{i_1,n}\tilde{\varepsilon}_{i_2,n}] \mathbb{E}[\tilde{\varepsilon}_{i_3,n}\tilde{\varepsilon}_{i_4,n}]. \end{aligned} \quad \text{(A.29)}$$

We first consider the first term of (A.29). Recall that the random variables $\tilde{\varepsilon}_{i,n}$ are m_n-dependent and that $|\tilde{B}_j| \le \lfloor \frac{|\lambda_1-\lambda_2|n}{\ell_n} \rfloor + 2$. Therefore, the number of non-zero summands in the inner sum can be bounded from above by $\big(\lfloor \frac{|\lambda_1-\lambda_2|n}{\ell_n} \rfloor + 2\big)^2 m_n^2$. By Assumption 2.3 (4), Assumption 3.1 and Lipschitz continuity of d, the first term in (A.29) can be bounded from above by $C|\lambda_1 - \lambda_2|^2 \frac{m_n^2}{\ell_n} \le C|\lambda_1 - \lambda_2|^2$.

The second term of (A.29) is bounded by

$$S_n = C\Big(\frac{1}{n}\sum_{j=1}^{\ell_n}\sum_{i_1,i_2\in\tilde{B}_j} d_{\omega_n}(i_1/n)d_{\omega_n}(i_2/n)\mathbb{E}[\tilde{\varepsilon}_{i_1,n}\tilde{\varepsilon}_{i_2,n}]\Big)^2$$

for some constant C. The inner sum in this term can be rewritten as $\frac{|\lambda_1-\lambda_2|n}{\ell_n}d^2_{\omega_n}(\frac{jb_n}{n})\sigma^2(\frac{jb_n}{n})+o(1)$, analogously to (A.23). Thus, S_n converges to $C|\lambda_1-\lambda_2|^2\|d_\omega\sigma\|_2^4$. Combining these arguments we obtain

$$R_n \le C|\lambda_1-\lambda_2|^2.$$

Therefore, by Theorem 2.2.4 of van der Vaart and Wellner (1996) it follows that

$$\mathbb{E}\Big[\sup_{|\lambda_1-\lambda_2|\le\rho}\Big|\frac{1}{\sqrt{n}}\sum_{i=\lfloor(\lambda_1\wedge\lambda_2)n\rfloor+1}^{\lfloor(\lambda_1\vee\lambda_2)n\rfloor}\tilde{\varepsilon}_{T_i,n}d_{\omega_n}(T_j/n)\Big|^4\Big]$$
$$\le K^4\Big\{\int_0^\eta D^{1/4}(\varepsilon)\mathrm{d}\varepsilon+\rho^{1/2}D^{1/2}(\eta)\Big\}^4 \le K^4\Big(2\eta^{1/2}+\frac{\rho^{1/2}}{\eta}\Big)^4,$$

for some constant K and any $\eta>0$, where $D(\varepsilon)$ denotes the packing number of the space $([0,1],|\cdot|^{1/2})$ and can be bounded from above by ε^{-2}. Thus, by (A.28) and Markov's inequality,

$$\lim_{\rho\searrow 0}\lim_{n\to\infty}\mathbb{P}(\sup_{|\lambda_1-\lambda_2|\le\rho}|\tilde{G}_n(\lambda_1)-\tilde{G}_n(\lambda_2)|>\varepsilon)\le 16\tfrac{K^4}{\varepsilon^4}\eta^2,$$

for any $\eta>0$, which proves (A.27) and completes the proof of the lemma. □

A.3. **Proof of the statements in Remark 3.3.** Part (i) is obvious. Part (ii) of the statement follows with $g(\mu)=c$ and $\omega\equiv 0$. For a proof of part (iii) note that

$$\lambda\Big(\hat{g}_n(\lambda)-\frac{1}{t_1-t_0}\int_{t_0}^{t_1}\mu(t)\mathrm{d}t\Big)$$
$$=\frac{1}{(t_1-t_0)}\Big\{\frac{1}{n}\sum_{i=1}^{\lfloor\lambda n\rfloor}\varepsilon_{T_i,n}\mathbb{1}(t_0\le T_i/n\le t_1)+\frac{1}{n}\sum_{i=1}^{\lfloor\lambda n\rfloor}\mu(\tfrac{T_i}{n})\mathbb{1}(t_0\le T_i/n\le t_1)-\lambda\int_{t_0}^{t_1}\mu(t)\mathrm{d}t\Big\}$$
$$=\frac{1}{(t_1-t_0)n}\sum_{i=1}^{\lfloor\lambda n\rfloor}\varepsilon_{T_i,n}\mathbb{1}(t_0\le T_i/n\le t_1)+\mathcal{O}(b_n/n)\ .$$

Consequently Assumption 3.1 holds with $\omega(x)=(t_1-t_0)^{-1}\mathbb{1}(t_0\le x\le t_1)$.

Finally, for a proof of part (iv) note that it follows from Lemma A.1 and A.2 that

$$\lambda\sqrt{n}(\hat{g}_n(\lambda)-g(\mu)) = g\Big(\lambda\sqrt{n}(\tilde{\mu}_{h_n}(\lambda,\cdot)-\mu)\Big)$$
$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor\lambda n\rfloor}\varepsilon_{T_i,n}g(\omega'(T_i/n,\cdot)/h_n)+o_{\mathbb{P}}(1),$$

by linearity of g, where the function ω' is defined by

$$\omega'(T_i/n,t)=K^*_{h_n}(\tfrac{T_i}{n}-t)\mathbb{1}_{[h_n,1-h_n]}(t)$$

$$+\Big\{2c_{\frac{h_n}{\sqrt{2}},i}(t)K_{\frac{h_n}{\sqrt{2}}}\big(\tfrac{T_i}{n}-t\big)-c_{h_n,i}(t)K_{h_n}\big(\tfrac{T_i}{n}-t\big)\Big\}\mathbb{1}_{[0,h_n)\cup(1-h_n,1]}(t).$$

Note K_{h_n} and $K^*_{h_n}$ are Lipschitz continuous with constant C_k/h_n where C_K is the Lipschitz constant of K and K^*. In particular, if $\mathrm{supp}(K)\subset[-1,1]$, it holds

$$\begin{aligned}\Big|\omega'\Big(\tfrac{jb_n}{n},t\Big)-\omega'\Big(\tfrac{jb_n+r}{n},t\Big)\Big| &\le C_k\frac{r}{nh_n}\mathbb{1}\Big(t\in[\tfrac{jb_n}{n}-h_n,\tfrac{jb_n}{n}+h_n]\cup[\tfrac{jb_n+r}{n}-h_n,\tfrac{jb_n+r}{n}+h_n]\Big)\\ &\le C_k\frac{r}{nh_n}\mathbb{1}\Big(t\in[\tfrac{jb_n}{n}-h_n,\tfrac{(j+1)b_n}{n}+h_n]\Big)\\ &= C_k\frac{r}{nh_n}\mathbb{1}\Big((t-h_n)\tfrac{n}{b_n}-1\le j\le(t+h_n)\tfrac{n}{b_n}\Big),\end{aligned}\tag{A.30}$$

for any $j\in\{1,\dots,\ell_n\}$ and $r\in\{1,\dots,b_n\}$. Since g is bounded, the function $\omega_n(x):=g(\omega'(x,\cdot)/h_n)$ therefore satisfies (3.3), that is

$$\sum_{j=1}^{\ell_n}\sum_{r=1}^{b_n}\big|\omega\big(\tfrac{jb_n}{n}\big)-\omega\big(\tfrac{r+jb_n}{n}\big)\big|\le\sum_{j=1}^{\ell_n}\sum_{r=1}^{b_n}\frac{\|g\|_{op}}{h_n}\big\|\omega'\big(\tfrac{jb_n}{n},\cdot\big)-\omega'\big(\tfrac{r+jb_n}{n},\cdot\big)\big\|_2=\mathcal{O}(b_nh_n^{-1}).$$

To complete the argument, note that the assumption $\|\omega_n-\omega\|_4\to0$ is only needed in the proof of Theorem 3.4 to establish the convergence in (A.24). However, with $\omega=h_g$, this argument can now be obtained directly noting that the continuity of h_g implies for any $j\in\{\lceil 2\ell_nh_n\rceil,\dots,\lfloor\ell_n(1-2h_n)\rfloor\}$

$$\begin{aligned}\omega_n(j/\ell_n)=\langle h_g,\omega'(j/\ell_n,\cdot)/h_n\rangle&=\frac{1}{h_n}\int_{I_n}h_g(x)\omega'(j/\ell_n,x)\mathrm{d}x+o(1)\\ &=\frac{1}{h_n}\int_{I_n}h_g(x)K^*_{h_n}(j/\ell_n-x)\mathrm{d}x+o(1)\\ &=h_g(j/\ell_n)\int_{I_n}\frac{1}{h_n}K^*_{h_n}(j/\ell_n-x)\mathrm{d}x+o(1)\\ &=h_g(j/\ell_n)\int_{-1}^{1}K^*(x)\mathrm{d}x+o(1)=h_g(j/\ell_n)+o(1).\end{aligned}$$

This gives

$$\begin{aligned}\frac{1}{\ell_n}\sum_{j=1}^{\ell_n}\sigma^2(j/\ell_n)d^2_{\omega_n}(j/\ell_n)&=\frac{1}{\ell_n}\sum_{j=1}^{\ell_n}\sigma^2(j/\ell_n)\Big(f_\tau(j/\ell_n)d(j/\ell_n)+\omega_n(j/\ell_n)\int_0^1d(x)\tau(\mathrm{d}x)\Big)^2\\ &=\frac{1}{\ell_n}\sum_{j=1}^{\ell_n}\sigma^2(j/\ell_n)\Big(f_\tau(j/\ell_n)d(j/\ell_n)+h_g(j/\ell_n)\int_0^1d(x)\tau(\mathrm{d}x)\Big)^2+o(1),\end{aligned}$$

which converges to $\|d_\omega\sigma\|_2^2$.

A.4. **Proof of Corollary 3.6.** If $\|d_\omega\sigma\|_2>0$ the corollary follows immediately from Theorem 3.4 or Remark 3.5 since

$$\mathbb{P}\Bigg(\frac{\hat d^2_{2,n}(1)-\Delta^2}{\int_\zeta^1\lambda|\hat d^2_{2,n}(\lambda)-\hat d^2_{2,n}(1)|\mathrm{d}\nu(\lambda)}>q_{1-\alpha}\Bigg)\to\begin{cases}0,&\text{if } d_0<\Delta\\ \alpha,&\text{if } d_0=\Delta\\ 1,&\text{if } d_0>\Delta.\end{cases}$$

If $\|d\sigma\|_2 = 0$, it follows $d \equiv 0$ by Assumption 2.3 (3), and in this case the probability to reject the null hypothesis by the decision rule (3.6) converges to $\mathbb{P}(0 > \Delta^2) = 0$.

Ruhr-Universität Bochum, Fakultät für Mathematik, Universitätsstr. 150, 44780 Bochum, Germany.

E-mail address: florian.heinrichs@rub.de

E-mail address: holger.dette@rub.de