Change point analysis in non-stationary processes - a mass excess approach

Holger Dette, Weichi Wu

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum

Germany

January 28, 2018

Abstract

This paper considers the problem of testing if a sequence of means $(\mu_t)_{t=1,\ldots,n}$ of a non-stationary time series $(X_t)_{t=1,\ldots,n}$ is stable in the sense that the difference of the means μ_1 and μ_t between the initial time $t = 1$ and any other time is smaller than a given level, that is $|\mu_1 - \mu_t| \leq c$ for all $t = 1, \ldots, n$. A test for hypotheses of this type is developed using a bias corrected monotone rearranged local linear estimator and asymptotic normality of the corresponding test statistic is established. As the asymptotic variance depends on the location and order of the critical roots of the equation $|\mu_1 - \mu_t| = c$ a new bootstrap procedure is proposed to obtain critical values and its consistency is established. As a consequence we are able to quantitatively describe relevant deviations of a non-stationary sequence from its initial value. The results are illustrated by means of a simulation study and by analyzing data examples.

AMS subject classification: 62M10, 62F05, 62G08, 62G09

Keywords and phrases: locally stationary process, change point analysis, relevant change points, local linear estimation, Gaussian approximation, rearrangement estimators

1 Introduction

A frequent problem in time series analysis is the detection of structural breaks. Since the pioneering work of Page (1954) in quality control change point detection has become an important tool with numerous applications in economics, climatology, engineering, hydrology and many authors have developed statistical tests for the problem of detecting structural breaks or change-points in various models. Exemplarily we mention Chow (1960), Brown et al. (1975), Krämer et al. (1988), Andrews (1993), Bai and Perron (1998) and Aue et al. (2009)] and refer to the work of Aue and Horváth (2013) and Jandhyala et al. (2013) for more recent reviews.

Most of the literature on testing for structural breaks formulates the hypotheses such that in the statistical model the stochastic process under the null hypothesis of "no change-point" is stationary. For example, in the problem of testing if a sequence of means $(\mu_t)_{t=1,\dots,n}$ of a non-stationary time series $(X_t)_{t=1,\dots,n}$ is stable it is often assumed that $X_t = \mu_t + \varepsilon_t$ with a stationary error process $(\varepsilon_t)_{t=1,\dots,n}$. The hull hypothesis is given by

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_n \ , \tag{1.1}$$

while the alternative (in the simplest case of only one structural break) is defined as

$$H_1 : \mu^{(1)} = \mu_1 = \mu_2 = \cdots = \mu_k \ \neq \ \mu_{k+1} = \mu_{k+2} = \cdots = \mu_n = \mu^{(2)},$$

where $k \in \{1, \dots, n\}$ denotes the (unknown) location of the change. The formulation of the null hypothesis in the form (1.1) facilitates the analysis of the distributional properties of a corresponding test statistic substantially, because one can work under the assumption of stationarity. Consequently, it is a very useful assumption from a theoretical point of view.

On the other hand, if the differences $\{|\mu_1 - \mu_t|\}_{t=2,\dots,n}$ are rather "small", a modification of the statistical analysis might not be necessary although the test rejects the "classical" null hypothesis (1.1) and detects non-stationarity. For example, as pointed out by Dette and Wied (2016), in risk management one wants to fit a model for forecasting the Value at Risk from "uncontaminated data", that means from data after the last change-point. If the changes are small they might not yield large changes in the Value at Risk. Now using only the uncontaminated data might decrease the bias but increases the variance of a prediction. Thus, if the changes are small, the forecasting quality might not necessarily decrease and - in the best case - would only improve slightly. Moreover, any benefit with respect to statistical accuracy could be negatively overcompensated by additional transaction costs.

In order to address these issues Dette and Wied (2016) proposed to investigate *precise* hypotheses in the context of change point analysis, where one does not test for exact equality, but only looks for "similarity" or a "relevant" difference. This concept is well known in biostatistics [see, for example, Wellek (2010)] but has also been used to investigate the similarity of distribution functions [see Álvarez Esteban et al. (2008, 2012) among others]. In the context of detecting a change in a sequence of means (or other parameters of the marginal distribution) Dette and Wied (2016) assumed two stationary phases and tested if the difference before and after the change point is small, that is $H_0 : |\mu^{(1)} - \mu^{(2)}| \leq c$ versus $H_1 : |\mu^{(1)} - \mu^{(2)}| > c$, where $c > 0$ is a given constant specified by the concrete application (in the example of the previous paragraph c could be determined by the transaction costs). Their approach heavily relies on the fact that the process before and after the change point is stationary, but this assumption might also be questionable in many applications.

In the present paper we investigate alternative hypotheses in the change point problem, which are motivated by the observation that in many applications the process parameters change continuously, and - if the amount of change and the time of a substantial change are small - the statistical analysis does not have to be modified. For this purpose we consider the location scale model

$$X_{i,n} = \mu(i/n) + \epsilon_{i,n}, \tag{1.2}$$

where $\{\epsilon_{i,n} \colon i = 1, \ldots, n\}_{n \in \mathbb{N}}$ denotes a triangular array of centered random variables (note that we do not assume that the "rows" $\{\epsilon_{j,n} : j = 1, \ldots, n\}$ are stationary) and $\mu : [0,1] \to \mathbb{R}$ is the unknown mean function. We define a change as *relevant*, if the amount of the change and the time period where the change occurs are reasonable large. More precisely, for a level $c > 0$ we consider the *level set*

$$\mathcal{M}_c = \{t \in [0,1] \colon |\mu(t) - \mu(0)| > c\}$$

of all points $t \in [0,1]$, where the mean function differs from its original value at the point 0 by an amount larger than c, and define

$$T_c := \lambda(\mathcal{M}_c)$$

as the corresponding *excess* measure, where λ denotes the Lebesgue measure. We now propose to investigate the hypothesis that the relative time, where this difference is larger than c does not exceed a given constant, say $\Delta \in (0,1)$, that is

$$H_0 : T_c \leq \Delta \quad \text{versus} \quad H_1 : T_c > \Delta \ . \tag{1.5}$$

In many applications it might also be of interest to investigate one-sided hypotheses, because one wants to detect a change in certain direction. For this purpose we also consider the sets $\mathcal{M}_c^{\pm} = \{t \in [0,1] \colon \pm(\mu(t) - \mu(0)) > c\}$ and define the hypotheses

$$H_0^+ : T_c^+ = \lambda(\mathcal{M}_c^+) \leq \Delta \quad \text{versus} \quad H_1^+ : T_c^+ > \Delta \ , \tag{1.6}$$

$$H_0^- : T_c^- = \lambda(\mathcal{M}_c^-) \leq \Delta \quad \text{versus} \quad H_1^- : T_c^- > \Delta \ . \tag{1.7}$$

Although the mean function in model (1.2) cannot be assumed to be monotone, we use a monotone rearrangement type estimator [see Dette et al. (2006)] to estimate the quantities T_c, T_c^+, T_c^-, and propose to reject the null hypothesis (1.5), (1.6) (1.7) for large values of the corresponding test statistic. We study the properties of these estimators and the resulting tests in a model of the form (1.2) with a locally stationary error process, which have found considerable interest in the literature [see Dahlhaus et al. (1997), Nason et al. (2000), Ombao et al. (2005), Zhou and Wu (2009) and Vogt (2012) among others]. In particular we do **not** assume that the underlying process is stationary, as the mean function can vary smoothly in time and the error process is non-stationary. Moreover, we also allow that the derivative of the mean function μ may vanish on the set of *critical roots*

$$\mathcal{C} = \{t \in [0,1] \colon |\mu(t) - \mu(0)| = c\}$$

and prove that appropriately standardized versions of the monotone rearrangement estimators are consistent for T_c, T_c^+ and T_c^-, and asymptotically normal distributed. It is demonstrated - even in the case of independent or stationary errors - that the variance of the limit distribution depends sensitively on (eventually higher order) derivatives of the regression function at the critical roots, which are very difficult to estimate. Moreover, because of the non-stationarity of the error process in (1.2) the asymptotic variance depends also in a complicated way on the unknown dependence structure. We propose a bootstrap method to obtain critical values for the test, which is motivated by a Gaussian approximation used in the proof of the asymptotic normality. This re-sampling procedure is adaptive in the sense that it avoids the direct estimation of the critical roots and the values of the derivative of the regression function at these points.

Note that T_c is the *excess* Lebesgue measure (or mass) of the time when the absolute difference between the mean trend and its initial value exceeds the level c. Thus our approach is naturally related to the concept of excess mass which has found considerable attention in the literature. Many authors used the excess mass approach to investigate multimodality of a density [see, for example, Müller and Sawitzki (1991), Polonik (1995), Cheng and Hall (1998), Polonik and Wang (2006)]. The asymptotic properties of distances between an estimated level and the "true" level set of a density have also been studied in several publications [see Baillo (2003), Cadre (2006), Cuevas et al. (2006) and Mason and Polonik (2009) among many others]. The concept of mass excess has additionally been used for discrimination between time series [see Chandler and Polonik (2006)], for the construction of monotone regression estimates [Dette et al. (2006), Chernozhukov et al. (2010)], quantile regression [Dette and Volgushev (2008), Chernozhukov et al. (2009)], clustering [Rinaldo and Wasserman (2010)] and for bandwidth selection in density estimation [see Samworth and Wand (2010)], but to our best knowledge it has not been used for change point analysis.

Most of the literature discusses regular points, that are points, where the first derivative of the density or regression function does not vanish, but there exist also references where this condition is relaxed. For example, Hartigan and Hartigan (1985) proposed a test for multimodality of a density comparing the difference between the empirical distribution function and a class of unimodal distribution functions. They observed that the stochastic order of the test statistic depends on the minimal number k, such that the kth derivative of the cumulative distribution function does not vanish. Polonik (1995) studied the asymptotic properties of an estimate of the mass excess functional of a cumulative distribution function F with density f and Tsybakov (1997) observed that the minimax risk in the problem of estimating the level set of a density depends on its "regularity". More recently, Chandler and Polonik (2006) used the excess mass functional for discrimination analysis under the additional assumption of unimodality.

The present paper differs from this literature with respect to several perspectives. First, we are interested in change point analysis and develop a test for a relevant difference in the mean of the process over a certain range of time. Therefore - in contrast to most of the literature, which deals with i.i.d. data - we consider the regression model (1.2) with a non-stationary error process. Second, we are interested in an estimate, say $\hat{T}_{N,c}$ of the Lebesgue measure T_c of the level set $\mathcal{M}_c$ and its asymptotic properties in order to construct a test for the change point problem (1.5). Therefore - in contrast to many references - we do not discuss estimates of an excess mass functional or a distance between an estimated level set and the "true" level set, but investigate the asymptotic distribution of $\hat{T}_{N,c}$. Third, as this distribution depends sensitively on the critical points and the dependence structure of the non-stationary error process, we use a Gaussian approximation to develop a bootstrap method, which allows us to find quantiles without estimating the location of the critical points and the derivatives of the regression function at these points.

The rest of paper is organized as follows. In Section 2 we define an estimator of the quantity T_c and give some basic assumptions of the non-stationary model (1.2). Section 3 is devoted to a discussion of the asymptotic properties of this estimator in the case, where all critical points are regular points, that is $\mu^{(1)}(s) \neq 0$ for all $s \in \mathcal{C}$. We focus on this case first, because here the arguments are more transparent. In particular we identify a bias problem, which makes the implementation of the test at this stage difficult. The general case is carefully investigated in Section 4, where we also address the bias problem using a Jacknife approach. The bootstrap procedure is developed in Section 5, where we also prove its validity and illustrate its finite sample properties by means of a simulation study and by analyzing data examples. Finally, most of the technical details are deferred to an appendix in Sections 7 and 8 (the latter section contains some auxiliary results).

2 Estimation and basic assumptions

Recall the definition of the testing problems (1.5), (1.6), (1.7) and note that $T_c = T_c^+ + T_c^-$, where

$$T_c^+ = \int_0^1 \mathbf{1}(\mu(t) - \mu(0) \geq c)dt \ , \quad T_c^- = \int_0^1 \mathbf{1}(\mu(t) - \mu(0) \leq -c)dt,$$

and $\mathbf{1}(B)$ denotes the indicator function of the set B. In most parts of the paper we mainly concentrate on the estimation of the quantity T_c^+ and study the asymptotic properties of an appropriately standardized estimate [see for example Theorem 3.1 and 4.1]. Corresponding results for the estimators of T_c^- and T_c can be obtained by similar methods and the joint weak convergence is established in Theorem 3.2 and Theorem 4.2 without giving detailed proofs.

We propose to estimate the mean function by a local linear estimator

$$(\hat{\mu}_{b_n}(t), \hat{\dot{\mu}}_{b_n}(t))^T = \underset{\beta_0\in\mathbb{R},\beta_1\in\mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^{n} \left(X_i - \beta_0 - \beta_1(i/n - t)\right)^2 K\Big(\frac{i/n - t}{b_n}\Big) \ , \qquad (2.1)$$

where $K(\cdot)$ denotes a continuous and symmetric kernel supported on the interval $[-1, 1]$. We define an estimator of T_c^+ by

$$\hat{T}_{N,c}^+ = \frac{1}{N}\sum_{i=1}^{N}\int_c^\infty \frac{1}{h_d}K_d\Big(\frac{\hat{\mu}_{b_n}(i/N) - \hat{\mu}_{b_n}(0) - u}{h_d}\Big)du, \qquad (2.2)$$

where $K_d(\cdot)$ is a symmetric kernel function supported on the interval $[-1, 1]$ such that $\int_{-1}^1 K_d(x)dx = 1$. In (2.2) the quantity $h_d > 0$ denotes a bandwidth and N is the number of knots in a Riemann approximation (see the discussion in the following paragraph), which does not need to coincide with the sample size n. A statistic of the type (2.2) has been proposed by Dette et al. (2006) to estimate the inverse of a strictly increasing regression function, but we use it here without assuming monotonicity of the mean function μ. We shall see later that h_d is usually chosen to be small to reduce error in the approximation of $\int_0^1 \mathbf{1}(\mu(t) - \mu(0) \geq c)dt$ by $\int_0^1 \frac{1}{h_d}\int_c^\infty K_d\big(\frac{\mu(t)-\mu(0)-u}{h_d}\big)dudt$. Observing that $\hat{\mu}_{b_n}(t)$ is a consistent estimate of $\mu(t)$ we argue (rigorous arguments are given later) that

$$\hat{T}^+_{N,c} = \frac{1}{N}\sum_{i=1}^N \int_c^\infty \frac{1}{h_d} K_d\Big(\frac{\mu(i/N)-\mu(0)-u}{h_d}\Big)du + o_P(1)$$
$$= \frac{1}{h_d}\int_0^1 \int_c^\infty K_d\Big(\frac{\mu(x)-\mu(0)-u}{h_d}\Big)dudx + o_P(1) = T_c^+ + o_P(1) \qquad (2.3)$$

as $n, N \to \infty$, $h_d \to 0$. In Section 3 we will establish the asymptotic properties of the statistic $\hat{T}^+_{N,c}$ as an estimator of T_c^+ for a twice continuously differentiable mean function. For these investigations we will make the following basic assumptions for model (1.2).

Assumption 2.1.

(a) The mean function is twice differentiable with Lipschitz continuous second derivative.

(b) There exists a positive constant ϵ_0, such that for all $\delta \in [0, \epsilon_0]$ there are k_δ closed disjoint intervals $I_{1,\delta}, \ldots, I_{k_\delta,\delta}$, such that

$$\big\{t \in [0,1] : |\mu(t) - \mu(0) - c| \le \delta\big\} \cup \big\{t \in [0,1] : |\mu(t) - \mu(0) + c| \le \delta\big\} = \bigcup_{i=1}^{k_\delta} I_{i,\delta},$$

where the number of intervals k_δ satisfies $\sup_{0\le\delta\le\epsilon_0} k_\delta \le M$ for some universal constant M. In particular there exists only a finite number of roots of the equation $\mu(t) - \mu(0) = \pm c$. We also assume that $|\mu(1) - \mu(0)| \ne c$.

Our first result makes the approximation of T_c^+ by deterministic counterpart

$$T^+_{N,c} := \frac{1}{N}\sum_{i=1}^N \int_c^\infty \frac{1}{h_d} K_d\Big(\frac{\mu(i/N)-\mu(0)-u}{h_d}\Big)du \qquad (2.4)$$

in (2.3) more precise. For this purpose let

$$m_{\gamma,\delta}(\mu) = \lambda\big(\{t \in [0,1] : |\mu(t) - \gamma| \le \delta\}\big)$$

denote the Lebesgue measure of the set of points, where the mean function lies in a δ-neighbourhood of the point γ.

Proposition 2.1. *If Assumption 2.1 holds and* $m_{c+\mu(0),\delta}(\mu) = O(\delta^{\iota})$ *for some* $\iota > 0$ *as* $\delta \to 0$*, we have for the quantity* $T^{+}_{N,c}$ *in (2.4),*

$$T^{+}_{N,c} - T^{+}_{c} = O(\max\{h_d^{\iota}, N^{-1}\})$$

as $N \to \infty$*,* $h_d \to 0$*.*

Proof. By elementary calculations it follows that

$$\int_c^{\infty} \frac{1}{h_d} K_d\Big(\frac{\mu(i/N) - \mu(0) - u}{h_d}\Big) du - \mathbf{1}(\mu(i/N) - \mu(0) \geq c)$$
$$= \mathbf{1}(\{|c - (\mu(i/N) - \mu(0))| \leq h_d\}) \int_{\frac{c-\mu(i/N)+\mu(0)}{h_d}}^{\infty} K_d(x)dx$$
$$- \mathbf{1}(\{\mu(i/N) - \mu(0) - h_d \leq c \leq \mu(i/N) - \mu(0)\}).$$

Therefore, we obtain (observing that $\int_{-1}^{1} K_d(x)dx = 1$)

$$|T^{+}_{N,c} - T^{+}_{c}| = \Big|\frac{1}{N}\sum_{i=1}^{N} \int_c^{\infty} \frac{1}{h_d} K_d\Big(\frac{\mu(\frac{i}{N}) - \mu(0) - u}{h_d}\Big) du - \mathbf{1}(\mu(\tfrac{i}{N}) - \mu(0) \geq c)\Big| + O\Big(\frac{1}{N}\Big)$$
$$\leq \frac{2}{N}\sum_{i=1}^{N} \mathbf{1}(|\mu(i/N) - \mu(0) - c| \leq h_d) + O(N^{-1})$$
$$= 2m_{c+\mu(0),h_d}(\mu) + O(N^{-1}) = O\Big(\max\{h_d^{\iota}, \frac{1}{N}\}\Big).$$

as $N \to \infty$, $h_d \to 0$. □

For $q \geq 1$ let $||X||_q = \big(\mathbb{E}|X|^q\big)^{1/q}$ denote the $\mathcal{L}_q$-norm of the random variable X. We begin recalling some basic definitions on physical dependence and locally stationary processes.

Definition 2.1. *Let* $\eta = (\eta_i)_{i\in\mathbb{Z}}$ *be a sequence of indendent identically distributed random variables,* $\mathcal{F}_i = \{\eta_s \colon s \leq i\}$*, denote by* $\eta' = (\eta'_i)_{i\in\mathbb{Z}}$ *an independent copy of* η *and define* $\mathcal{F}^*_i = (\ldots, \eta_{-2}, \eta_{-1}, \eta'_0, \eta_1, \ldots, \eta_i)$*. For* $t \in [0,1]$ *let* $G : [0,1] \times \mathbb{R}^{\infty} \to \mathbb{R}$ *denote a nonlinear filter, that is a measurable function, such that* $G(t, \mathcal{F}_i)$ *is a properly defined random variable for all* $t \in [0,1]$*.*

(1) *A sequence* $(\epsilon_{i,n})_{i=1,\dots,n}$ *is called locally stationary process, if there exists a filter* G *such that* $\epsilon_{i,n} = G(i/n, \mathcal{F}_i)$ *for all* $i = 1, \dots, n$.

(2) *For a nonlinear filter* G *with* $\sup_{t\in[0,1]} \|G(t,\mathcal{F}_i)\|_q < \infty$, *the physical dependence measure of* G *with respect to* $\|\cdot\|_q$ *is defined by*

$$\delta_q(G,k) = \sup_{t\in[0,1]} \|G(t,\mathcal{F}_k) - G(t,\mathcal{F}_k^*)\|_q. \tag{2.5}$$

(3) *The filter* G *is called Lipschitz continuous with respect to* $\|\cdot\|_q$ *if and only if*

$$\sup_{0\le s<t\le 1} \|G(t,\mathcal{F}_i) - G(s,\mathcal{F}_i)\|_q / |t-s| < \infty. \tag{2.6}$$

The filter G is used to model non-stationarity. The quantity $\delta_q(G,k)$ measures the dependence of $G(t,\mathcal{F}_k)$ on η_0' over the interval $[0,1]$. When $\delta_q(G,k)$ converges sufficiently fast to 0 such that $\sum_k \delta_q(G,k) < \infty$, we speak of a short range dependent time series. Condition (2.6) means that the data generating mechanism G is varying smoothly in time. We refer to Zhou and Wu (2009) for more details, in particular for examples of locally stationary linear and nonlinear time series, calculations of the dependence measure (2.5) and for the verification of (2.6). With this notation we make the following assumptions regarding the error process in model (1.2).

Assumption 2.2. The error process $(\epsilon_{i,n})_{i=1,\dots,n}$ in model (1.2) is a zero-mean locally stationary process with filter G, which satisfies the following conditions:

(a) There exists a constant $\chi \in (0,1)$, such that $\delta_4(G,k) = O(\chi^k)$ as $k \to \infty$.

(b) The filter G is Lipschitz continuous with respect to $\|\cdot\|_4$ and $\sup_{t\in[0,1]} \|G(t,\mathcal{F}_0)\|_4 < \infty$

(c) The *long-run variance*

$$\sigma(t) := \sum_{i=-\infty}^{\infty} \operatorname{cov}(G(t,\mathcal{F}_i), G(t,\mathcal{F}_0)), \quad t \in [0,1]$$

of the filter G is Lipschitz continuous on the interval $[0,1]$ and *non-degenerate*, that is $\inf_{t\in[0,1]} \sigma(t) > 0$.

Condition (a) of Assumption 2.2 means that the error process $\{\epsilon_{i,n}\}_{i=1,\dots,n}$ in model (1.2) is locally stationary with geometrically decaying dependence measure. The theoretical results of the paper can also be derived under the assumption of a polynomially decaying dependence measure with substantially more complicated bandwidth conditions and proofs. Conditions (b) and (c) are standard in the literature of locally stationary time series. They are used later for a strong Gaussian approximation of the locally stationary time series; see for example Zhou and Wu (2010).

3 Twice continuously differentiable mean functions

In this section we briefly consider the situation, where the derivatives of the mean function at the critical set to do not vanish. These assumptions are quite common in the literature [see, for example, condition (B.ii) in Mason and Polonik (2009) or assumption (A1) in Samworth and Wand (2010).] We discuss this case separately because of (at least) two reasons. First, the results and required assumptions are slightly simpler here. Second, and more important, we use this case to demonstrate that the estimates of T_c, T_c^+ and T_c^- have a bias, which is asymptotically not negligible and makes their direct application for testing the hypotheses (1.5), (1.6) and (1.7) difficult. The general case is postponed to Section 4, where we also solve the bias problem. We do not provide proofs of the results in this section, as they can be obtained by similar (but substantially simpler) arguments as given in the proofs of Theorem 4.1 and 4.2 below.

Recall the definition of the statistic $\hat{T}^+_{N,c}$ in (2.2), where $\hat{\mu}_{b_n}(t)$ is the local linear estimate of the mean function with bandwidth b_n. Our first result specifies its asymptotic distribution, and for its statement we make the following additional assumption on the bandwidths.

Assumption 3.1. *The bandwidth b_n of the local linear estimator satisfies $b_n \to 0, nb_n \to \infty$, $b_n/h_d \to \infty$, $\sqrt{n}b_n/\log^4 n \to \infty$, and $\pi_n^*/h_d \to 0$ where*

$$\pi_n^* := (b_n^2 + (nb_n)^{-1/2} \log n) \log n.$$

Theorem 3.1. *Suppose that Assumptions 2.1, 2.2 and 3.1 hold, that there exist roots*

$t_1^+, \ldots, t_{k^+}^+$ *of the equation* $\mu(t) - \mu(0) = c$ *satisfying* $\dot{\mu}(t_j^+) \neq 0$ *for* $1 \leq j \leq k^+$, *and define*

$$\bar{R}_{1,n} = \frac{n^{1/4}\log^2 n}{nb_n}, \ \ \bar{R}_{2,n} = \Big(\frac{1}{Nb_n} + \frac{1}{Nh_d}\Big)(b_n \wedge h_d),$$
$$\bar{\chi}_n = (b_n^4 + \frac{1}{nb_n})(h_d^{-1} + \pi_n h_d^{-2}).$$

If $Nb_n \to \infty$, $Nh_d \to \infty$, $\sqrt{nb_n}(\bar{\chi}_n + \bar{R}_{1,n} + \bar{R}_{2,n}) = o(1)$, $\sqrt{nb_n}h_d = o(1)$, *then*

$$\sqrt{nb_n}\Big(\hat{T}_{N,c}^+ - T_c^+ - \mu_{2,K} b_n^2 \sum_{j=1}^{k^+} \frac{\ddot{\mu}(t_j^+)}{|\dot{\mu}(t_j^+)|} + \frac{b_n^2 c_{2,K}\ddot{\mu}(0)}{2c_{0,K}} \sum_{j=1}^{k^+} \frac{1}{|\dot{\mu}(t_j^+)|}\Big) \stackrel{\mathcal{D}}{\Longrightarrow} \mathcal{N}(0, \tau_1^{2,+} + \tau_2^{2,+}),$$

where

$$\tau_1^{2,+} = \sum_{s=1}^{k^+} \frac{\sigma^2(t_s^+)}{\dot{\mu}(t_s^+)^2} \int K^2(x)dx,$$
$$\tau_2^{2,+} = \frac{\sigma^2(0)}{c_{0,K}^2}\Big(\sum_{j=1}^{k^+} \frac{1}{|\dot{\mu}(t_j^+)|}\Big)^2 \int_0^1 \big(\mu_{2,K} - t\mu_{1,K}\big)^2 K^2(t)dt,$$

the constants $c_{0,K}$ *and* $c_{2,K}$ *are given by*

$$c_{0,K} = \mu_{0,K}\mu_{2,K} - \mu_{1,K}^2 \ , \quad c_{2,K} = \mu_{2,K}^2 - \mu_{1,K}\mu_{3,K}$$

and $\mu_{l,K} = \int_0^1 x^l K(x)dx$ *for* $(l = 1, 2, \ldots)$.

Theorem 3.1 establishes asymptotic normality under the scenario that $\dot{\mu}(t) \neq 0$ for all points $t \in \mathcal{C}^+ = \{t \in [0,1] \colon \mu(t) - \mu(0) = c\}$. This condition guarantees that the mean function μ is strictly monotone in a neighbourhood of the roots. Moreover, Assumptions 2.1(b) and 3.1 imply the asymptotic independence of the estimators of $\mu(0)$ and $\mu(t)$ for any $t \in \mathcal{C}^+$.

We conclude this section presenting a corresponding weak convergence result for the joint distribution of $(\hat{T}_{n,c}^+, \hat{T}_{n,c}^-)$, where

$$\hat{T}_{N,c}^- = \frac{1}{N}\sum_{i=1}^N \int_{-\infty}^{-c} \frac{1}{h_d} K_d\left(\frac{\hat{\mu}_{b_n}(i/N) - \hat{\mu}_{b_n}(0) - u}{h_d}\right) du$$

denotes an estimate of the quantity T_c^- defined in (1.7).

Theorem 3.2. *Suppose that Assumptions 2.1, 2.2 and 3.1 are satisfied and that the bandwidth conditions of Theorem 3.1 hold. If there also exist roots $t_1^-, \ldots, t_{k^-}^-$ of the equation $\mu(t) - \mu(0) = -c$, such that $\dot{\mu}(t^-) \neq 0$ $(l = 1, \ldots, k^-)$, then, as $n \to \infty$,*

$$\sqrt{nb_n}\left(\hat{T}_{N,c}^+ - T_c^+ - b_c^+, \hat{T}_{N,c}^- - T_c^- - b_c^-\right)^T \overset{\mathcal{D}}{\Longrightarrow} \mathcal{N}(0, \tilde{\Sigma}),$$

where

$$b_c^{\pm} = \frac{\mu_{2,K} b_n^2}{2} \sum_{j=1}^{k^{\pm}} \frac{\ddot{\mu}(t_j^{\pm})}{|\dot{\mu}(t_j^{\pm})|} + \frac{b_n^2 c_{2,K} \ddot{\mu}(0)}{2c_{0,K}} \sum_{j=1}^{k^{\pm}} \frac{1}{|\dot{\mu}(t_j^{\pm})|}, \tag{3.1}$$

and the elements in the matrix $\tilde{\Sigma} = (\tilde{\Sigma}_{ij})_{i,j=1,2}$ are given by $\tilde{\Sigma}_{11} = \tau_1^{2,+} + \tau_2^{2,+}$, $\tilde{\Sigma}_{22} = \tau_1^{2,-} + \tau_2^{2,-}$ and

$$\tilde{\Sigma}_{12} = \tilde{\Sigma}_{21} = -c_{0,K}^{-2}\sigma^2(0)\Big(\sum_{j=1}^{k^+} \frac{1}{|\dot{\mu}(t_j^+)|}\Big)\Big(\sum_{j=1}^{k^-} \frac{1}{|\dot{\mu}(t_j^-)|}\Big) \int_0^1 (\mu_{2,K} - t\mu_{1,K})^2 K^2(t)dt.$$

where $\tau_1^{2,+}$ and $\tau_2^{2,+}$ are defined in a similar way as $\tau_1^{2,+}$ and $\tau_2^{2,+}$ in Theorem 3.1.

Remark 3.1. The representation of the bias in (3.1) has some similarity with the approximation of the risk of an estimate of the highest density region investigated in Samworth and Wand (2010). We suppose that similar arguments as given in the proofs of our main results can be used to derive asymptotic normality of this estimate [see also Mason and Polonik (2009)].

Theorem 3.1 and 3.2 can be used to construct tests for the hypotheses (1.6) and (1.7). Similarly, by the continuous mapping theorem we also obtain from Theorem 3.2 the asymptotic distribution of the the statistic $\hat{T}_{N,c} = \hat{T}_{N,c}^+ + \hat{T}_{N,c}^-$, which could be used to construct a test for the hypotheses (1.5). However, such tests would either require undersmoothing or estimation of the bias b_c^+ and b_c^- in (3.1), which is not an easy task. Therefore we next investigate an alternative procedure based on the Jackknife principle, which will be the basic tool for the bootstrap test discussed in Section 5.

4 Jack-Knife bias corrected test

In this section we will address the bias problem mentioned in the previous paragraph adopting the Jackknife bias reduction technique proposed by Schucany and Sommers (1977). Moreover, we also relax the main assumption that the derivative of the mean function does not vanish at critical roots $t \in \mathcal{C}$. Recalling the definition $\hat{\mu}_{b_n}(t)$ of the local linear estimator in (2.1) with bandwidth b_n we define the Jackknife estimator by

$$\tilde{\mu}_{b_n}(t) = 2\hat{\mu}_{b_n/\sqrt{2}}(t) - \hat{\mu}_{b_n}(t) \tag{4.1}$$

for $0 \le t \le 1$. It has been shown in Wu and Zhao (2007) that the bias of the estimator (4.1) is of order $o(b_n^3 + \frac{1}{nb_n})$, whenever $b_n \le t \le 1 - b_n$, and Zhou and Wu (2010) showed that the estimate $\tilde{\mu}_{b_n}$ is asymptotically equivalent to a local linear estimate with kernel

$$K^*(x) = 2\sqrt{2}K(\sqrt{2}x) - K(x). \tag{4.2}$$

In order to use these bias corrected estimators for the construction of tests for the hypotheses defined in (1.5) - (1.7), we also need to study the estimate $\tilde{\mu}_{b_n}(0)$, which is not asymptotically equivalent to a local linear estimate with kernel $K^*(x)$. However, as a consequence of Lemma 8.2 in the Appendix we obtain the stochastic expansion

$$\Big|\tilde{\mu}_{b_n}(0) - \mu(0) - \frac{1}{nb_n}\sum_{i=1}^{n} \bar{K}^*(\frac{i}{nb_n})e_i\Big| = O(b_n^3 + \frac{1}{nb_n}),$$

where the kernel $\bar{K}^*(x)$ is given by

$$\bar{K}^*(x) = 2\sqrt{2}\bar{K}(\sqrt{2}x) - \bar{K}(x) \tag{4.3}$$

with $\bar{K}(x) = (\mu_{2,K} - x\mu_{1,K})K(x)/c_{0,K}$. Since the kernel $\bar{K}^*(x)$ is not symmetric, the bias of $\tilde{\mu}_{b_n}(0)$ is of the order $O(b_n^3 + \frac{1}{nb_n})$. The corresponding estimators of the quantities T_c^+ and T_c^- are then defined as in Section 2, where the local linear estimator $\hat{\mu}_{b_n}$ is replaced by its bias corrected version $\tilde{\mu}_{b_n}$. For example, the analogue of the statistic in (2.2) is given by

$$\tilde{T}^{+}_{N,c} = \frac{1}{N}\sum_{i=1}^{N}\int_{c}^{\infty}\frac{1}{h_d}K_d\Big(\frac{\tilde{\mu}_{b_n}(i/N)-\tilde{\mu}_{b_n}(0)-u}{h_d}\Big)du. \tag{4.4}$$

The investigation of the asymptotic properties of these estimators in the general case requires some preparations, which are discussed next.

We call a point $t \in [0,1]$ a *regular* point of the mean function μ, if the derivative μ' does not vanish at t. A point $t \in \mathcal{C}$ is called a *critical point of* μ *of order* $k \geq 1$ if the first k derivatives of μ at t vanish while the $(k+1)$st derivative of μ at t is non zero, that is $\mu^{(s)}(t) = 0$ for $1 \leq s \leq k$ and $\mu^{(k+1)}(t) \neq 0$. Regular points are critical points of order 0. Theorem 3.1 or 3.2 are not valid if any of the roots of the equation $\mu(t) - \mu(0) = c$ or $\mu(t) - \mu(0) = -c$ is a critical point of order larger or equal than 1. The following result provides the asymptotic distribution in this case and also solves the bias problem mentioned in Section 3. For its statement we make the following additional assumptions.

Assumption 4.1. The mean function μ is three times continuously differentiable. Let $t_1^-, \dots, t_{k^-}^-$ and $t_1^+, \dots, t_{k^+}^+$ denote the roots of the equations $\mu(t) - \mu(0) = c$ and $\mu(t) - \mu(0) = -c$, respectively. For each t_s^- $(s = 1, \dots, k^-)$ and each t_s^+ $(s = 1, \dots, k^+)$ there exists a neighbourhood of t_s^- and t_s^+ such that μ is $(v_s^- + 1)$ and $(v_s^+ + 1)$ times differentiable in these neighbourhoods with corresponding critical order ν_s^- and ν_s^+, respectively $(1 \leq s \leq k^-,\ 1 \leq s \leq k^+)$. We also assume that the $(v_s^- + 1)$st and $(v_s^+ + 1)$st derivatives of the mean function are Lipschitz continuous on these neighbourhoods.

Assumption 4.2. There exist q points $0 = s_0 \leq s_1 \leq \ldots \leq s_q = 1$ such that the mean function μ is strictly monotone on each interval $(s_i, s_{1+1}]$ $(0 \leq i \leq q)$.

It is shown in Lemma 8.1 that under the assumptions made so far the set $\{t : |\mu(t) - c| \leq h_n, t \in [0,1]\}$ can be decomposed as a union of disjoint "small" intervals around the critical roots $t_i^+ \in \mathcal{C}^+$ and $t_i^- \in \mathcal{C}^-$, whose Lebesgue measure is of order $h_n^{1/(\nu_i^+ + 1)}$ and $h_n^{1/(\nu_i^- + 1)}$, respectively, and therefore depends on the order of the corresponding root. In the appendix we prove the following result, which clarifies the distributional properties of the estimator $\tilde{T}^{+}_{N,c}$ defined in (4.4) if the sample size converges to infinity.

Theorem 4.1. *Suppose that $k^+ \geq 1$, and that Assumptions 2.1, 2.2, 4.1 and Assumption 4.2 are satisfied. Define $v^+ = \max_{1\leq l\leq k^+} v_l^+$ as the maximum critical order of the roots of the equation $\mu(t) - \mu(0) = c$ and introduce the notation*

$$\chi_n^+ = \Big(b_n^6 + \frac{1}{nb_n}\Big) h_d^{-2}(h_d + \pi_n)^{\frac{1}{v^++1}} \ , \quad R_{1,n}^+ = h_d^{-\frac{v^+}{v^++1}}\Big(b_n^3 + \frac{1}{nb_n}\Big) \ ,$$
$$R_{2,n}^+ = \frac{n^{1/4}\log^2 n}{nb_n} h_d^{-\frac{v^+}{v^++1}} \ , \quad R_{3,n}^+ = \Big(\frac{1}{Nb_n} + \frac{1}{Nh_d}\Big)\Big(b_n \wedge h_d^{\frac{1}{v^++1}}\Big).$$

Assume further that the bandwidth conditions $h_d \to 0$, $nb_nh_d \to \infty$, $b_n \to 0$, $nb_n^2 \to \infty$, $Nb_n \to \infty$, $Nh_d \to \infty$ and $\pi_n = o(h_d)$ hold, where

$$\pi_n := (b_n^3 + (nb_n)^{-1/2}\log n)\log n = o(h_d), \tag{4.5}$$

then we have the following results.

(a) If $b_n^{v^++1}/h_d \to \infty$, $\sqrt{nb_n} h_d^{\frac{v^+}{v^++1}}(\chi_n^+ + R_{1,n}^+ + R_{2,n}^+ + R_{3,n}^+) = o(1)$, $\sqrt{nb_n}h_d = o(1)$, $\sqrt{nb_n}h_d^{\frac{v^+}{v^++1}}/N = o(1)$, then

$$\sqrt{nb_n} h_d^{\frac{v^+}{v^++1}} \left(\tilde{T}_{N,c}^+ - T_c^+\right) \overset{\mathcal{D}}{\Longrightarrow} \mathcal{N}(0, \sigma_1^{2,+} + \sigma_2^{2,+}),$$

where

$$\sigma_1^{2,+} = \Big(\int K_d(z^{v^++1})dz\Big)^2((v^++1)!)^{\frac{2}{v^++1}} \sum_{\{t_l^+ :\, v_l^+ = v^+\}} \frac{\sigma^2(t_l^+)}{|\mu^{(v^++1)}(t_l^+)|^{\frac{2}{v^++1}}} \int (K^*(x))^2 dx, \tag{4.6}$$

$$\sigma_2^{2,+} = \sigma^2(0)((v^++1)!)^{\frac{2}{v^++1}} \int_0^1 (\bar{K}^*(t))^2 dt \Big(\sum_{\{t_l^+ :\, v_l^+ = v^+\}} |\mu^{(v^++1)}(t_l^+)|^{\frac{-1}{v^++1}} \int K_d(z^{v^++1})dz\Big)^2. \tag{4.7}$$

(b) If $b_n/h_d^{\frac{1}{v^++1}} = r \in [0, \infty)$, $\sqrt{nh_d} h_d^{\frac{v^+}{2(v^++1)}}(\chi_n^+ + R_{1,n}^+ + R_{2,n}^+ + R_{3,n}^+) = o(1)$, then

$$\sqrt{nh_d} h_d^{\frac{v^+}{2(v^++1)}} \left(\tilde{T}_{N,c}^+ - T_{N,c}^+\right) \overset{\mathcal{D}}{\Longrightarrow} \mathcal{N}(0, \rho_1^{2,+} + \rho_2^{2,+}),$$

where

$$\rho_1^{2,+} = |(v^+ + 1)!|^{\frac{1}{v^+ + 1}} \sum_{\{t_l^+ : \, v_l^+ = v^+\}} \frac{\sigma^2(t_l^+)}{|\mu^{(v^+ + 1)}(t_l)|^{\frac{2}{v^+ + 1}}} \int\int\int K^*(u) K^*(v) K_d(z_1^{v^+ + 1})$$
$$\times K_d\Big(\Big(z_1 + r\Big|\frac{(v^+ + 1)!}{\mu^{(v^+ + 1)}(t_l^+)}\Big|^{\frac{-1}{v^+ + 1}}(v - u)\Big)^{v^+ + 1}\Big) du dv dz_1,$$

and $\rho_2^{2,+} = r^{-1}\sigma_2^{2,+}$*, where* $\sigma_2^{2,+}$ *is defined in* (4.7)

In general the rate of convergence of the estimator $\tilde{T}_{N,c}^+$ is determined by the maximal order of the critical points, and only critical points of maximal order appear in the asymptotic variance. The rate of convergence additionally depends on the relative order of the bandwidths b_n and h_d. Theorem 4.1 also covers the case $v^+ = 0$, where all roots of the equation $\mu(t) - \mu(0) = c$ are regular. Moreover, the use of the Jackknife corrected estimate $\tilde{\mu}_{b_n}$ avoids the bias problem observed in Theorem 3.1.

It is also worthwhile to mention that there exists a slight difference in the statement of part (a) and (b) of Theorem 4.1. While part (a) gives the asymptotic distribution of $\tilde{T}_{N,c}^+ - T_c^+$ (appropriately standardized), part (b) describes the weak convergence of $\tilde{T}_{N,c}^+ - T_{N,c}^+$. The replacement of $T_{N,c}^+$ by its limit T_c^+ is only possible under additional bandwidth conditions. In fact, if $b_n / h_d^{\frac{1}{v^+ + 1}} = c \in [0, \infty)$, Theorem 4.1 and Proposition 2.1 give

$$\sqrt{n h_d} h_d^{\frac{v^+}{2(v^+ + 1)}} \Big(\tilde{T}_{n,c}^+ - T_c^+\Big) - R_n \overset{\mathcal{D}}{\Longrightarrow} \mathcal{N}(0, \rho_1^{2,+} + \rho_2^{2,+}),$$

where $\rho_1^{2,+}$ and $\rho_2^{2,+}$ are defined in Theorem 4.1, and R_n is a an additional bias term of order

$$O(\sqrt{n h_d} h_d^{\frac{v^+ + 2}{2(v^+ + 1)}}),$$

which does not necessarily vanish asymptotically. For example, in the regular case $v^+ = 0$ this bias is of order $o(1)$ under the additional assumptions $n h_d^3 = o(1)$ and $b_n / h_d < \infty$. Note that these bandwidth conditions do not allow for the MSE-optimal bandwidth $b_n \sim n^{-1/5}$. These considerations give some arguments for using small bandwidths h_d in the estimator (4.4) such that condition (a) of Theorem 4.1 holds, that is $h_d = o(b_n^{\nu^+ + 1})$. Moreover, in numerical experiments we observed that smaller bandwidths h_d usually yield a substantially better performance of the estimator $\tilde{T}^+_{N,c}$ and in the remaining part of this section we concentrate on this case as this is most important from a practical point of view.

The next result gives a corresponding statement of the joint asymptotic distribution of $(\tilde{T}^+_{N,c}, \tilde{T}^-_{N,c})$ and as a consequence that of $\tilde{T}_{N,c} = \tilde{T}^+_{N,c} + \tilde{T}^-_{N,c}$, where the statistic $\tilde{T}^-_{N,c}$ is defined by

$$\tilde{T}^-_{N,c} = \frac{1}{N} \sum_{i=1}^{N} \int_{-\infty}^{-c} \frac{1}{h_d} K_d \left(\frac{\tilde{\mu}_{b_n}(i/N) - \tilde{\mu}_{b_n}(0) - u}{h_d} \right) du. \tag{4.8}$$

Theorem 4.2. *Assume that the conditions of Theorem 4.1 are satisfied, that $k^- \geq 1$ and define $v^- = \max_{1 \leq l \leq k^-} v_l^-$ as the maximum order of the critical roots $\{t_l^- : 1 \leq l \leq k^-\}$. If, additionally, the bandwidth conditions (a) of Theorem 4.1 hold and similar bandwidth conditions are satisfied for the level $-c$, we have*

$$\sqrt{nb_n} \Big(h_d^{\frac{v^+}{v^+ + 1}} (\hat{T}^+_{N,c} - T^+_c), h_d^{\frac{v^-}{v^- + 1}} (\hat{T}^-_{N,c} - T^-_c) \Big)^T \Rightarrow \mathcal{N}(0, \Sigma),$$

where the matrix $\Sigma = (\Sigma_{ij})_{i,j=1,2}$ has the entries $\Sigma_{11} = \sigma_1^{2,+} + \sigma_2^{2,+}$, $\Sigma_{22} = \sigma_1^{2,-} + \sigma_2^{2,-}$,

$$\Sigma_{12} = \Sigma_{21} = -\sigma^2(0)((v^+ + 1)!)^{\frac{1}{v^+ + 1}} ((v^- + 1)!)^{\frac{1}{v^- + 1}} \int_0^1 (\bar{K}^*(t))^2 dt$$
$$\times \sum_{\{t_l^+ : v_l^+ = v^+\}} \frac{\int K_d(z^{v^+ + 1}) dz}{|\mu^{(v^+ + 1)}(t_l^+)|^{1/(v^+ + 1)}} \sum_{\sum_{\{t_l^- : v_l^- = v^-\}}} \frac{\int K_d(z^{v^- + 1}) dz}{|\mu^{(v^- + 1)}(t_l^-)|^{1/(v^- + 1)}},$$

and $\sigma_1^{2,-}$, $\sigma_2^{2,-}$ are defined similarly as $\sigma_1^{2,+}$, $\sigma_2^{2,+}$ in (4.6), (4.7), *respectively.*

The continuous mapping theorem and Theorem 4.2 imply the weak convergence of the estimator $\hat{T}_{N,c}$ of T_c, that is $\sqrt{nb_n} h_d^{\frac{v}{v+1}} (\hat{T}_{N,c} - T_c) \to N(0, \sigma^2)$, where $v = \max\{v^+, v^-\}$ and the asymptotic variance is given by $\sigma^2 = \Sigma_{1,1} \mathbf{1}(v^+ \geq v^-) + \Sigma_{2,2} \mathbf{1}(v+ \leq v^-) + 2\Sigma_{1,2} \mathbf{1}(v^+ = v^-)$.

Remark 4.1. In applications one might also be interested if there exist relevant deviations of the sequence $(\mu(i/n))_{i=\lfloor nt_0 \rfloor + 1, \ldots, n}$ from an average trend formed from the previous period

$(\mu(i/n))_{i=1,\ldots,\lfloor nt_0\rfloor}$. This question can be addressed estimating the quantity

$$\int_{t_0}^{1} \mathbf{1}\Big(\mu(t) - \int_0^{t_0} \mu(s)ds \geq c\Big)dt = \lambda\Big(\Big\{t \in [t_0,1]: \ \mu(t) - \int_0^{t_0} \mu(s)ds \geq c\Big\}\Big).$$

Using similar mathematical arguments as given in Section 7 and 8 one can prove consistency and derive the asymptotic distribution of the estimate

$$\frac{1}{N}\sum_{i=\lfloor Nt_0\rfloor}^{N} \int_{c_0}^{\infty} \frac{1}{h_d} K_d\Big(\frac{\tilde{\mu}_{b_n}(i/N) - \int_0^{t_0} \tilde{\mu}_{b_n}(s)ds - u}{h_d}\Big)du$$

where $\tilde{\mu}_{b_n}$ is the bias-corrected local linear estimator of μ. The details are omitted for the sake of brevity.

5 Bootstrap

Although Theorem 4.1 is interesting from a theoretical point of view and avoids the bias problem described in Section 3, it can not be easily used to construct a test for the hypotheses (1.5). The asymptotic variance of the statistics $T^+_{N,c}$ and $T^-_{N,c}$ depends on the long-run variance $\sigma^2(\cdot)$ and the set $\mathcal{C}$ of critical points, which are difficult to estimate. Moreover, the order of the critical roots is usually unknown and not estimable. Therefore it is not clear which derivatives have to be estimated (the estimation of higher order derivatives of the mean function is a hard problem anyway). In this section, we develop an adaptive methodology to address this problem. In particular, we propose a bootstrap test which does not require the estimation of the derivatives of the mean trend at the critical roots.

The bootstrap procedure is motivated by an essential step in the proof of Theorem 4.1, which gives a stochastic approximation for the difference

$$\tilde{T}^+_{N,c} - T^+_c = I' + o_p\Big(\big(\sqrt{nb}h_d^{\frac{v^+}{v^++1}}\big)^{-1}\Big),$$

where the statistic I' is defined as

$$\frac{-1}{nNb_nh_d}\sum_{j=1}^{n}\sum_{i=1}^{N}K_d\Big(\frac{\mu(i/N)-\mu(0)-c}{h_d}\Big)\sigma\big(\frac{j}{n}\big)\Big(K^*\Big(\frac{i/N-j/n}{b_n}\Big)-\bar{K}^*\Big(\frac{j}{nb_n}\Big)\Big)V_j, \quad (5.1)$$

and $(V_j)_{j\in\mathbb{N}}$ is a sequence of independent standard normal distributed random variables. Based on this approximation we propose the following bootstrap to calculate critical values.

Algorithm 5.1.
(1) Choose bandwidths b_n, h_d and an estimator of the long-run variance, say $\hat{\sigma}^2(\cdot)$, which is uniformly consistent on the set $\cup_{k=1}^{\nu^+}\mathcal{U}_\varepsilon(t_k^+)$ for some $\varepsilon > 0$, where $\mathcal{U}_\varepsilon(t)$ denotes a ε-neighbourhood of the point t.
(2) Calculate the bias corrected local linear estimate $\tilde{\mu}_{b_n}(t)$ and the statistic $\tilde{T}^+_{N,c}$ defined in (4.1) and (4.4), respectively.
(3) Calculate

$$\bar{V}=\sum_{j=1}^{n}\hat{\sigma}^2\big(\frac{j}{n}\big)\Big[\sum_{i=1}^{N}K_d\Big(\frac{\tilde{\mu}_{b_n}(i/N)-\tilde{\mu}_{b_n}(0)-c}{h_d}\Big)\Big\{K^*\Big(\frac{i/N-j/n}{b_n}\Big)-\bar{K}^*\Big(\frac{j}{nb_n}\Big)\Big\}\Big]^2.$$

(4) Let $q^+_{1-\alpha}$ denote the the $1-\alpha$ quantile of a centred normal distribution with variance $\bar{V}$, then the null hypothesis in (1.6) is rejected, whenever

$$nNb_nh_d\big(\tilde{T}^+_{N,c}-\Delta\big) > q^+_{1-\alpha}. \quad (5.2)$$

Theorem 5.1. *Assume that the conditions of Theorem 4.1 (a) are satisfied, then the test* (5.2) *defines a consistent and asymptotic level α test for the hypotheses* (1.6).

Remark 5.1.
(a) It follows from the proof of Theorem 5.1 in the appendix that

$$\mathbb{P}\big(\text{ test (5.2) rejects }\big)\longrightarrow\begin{cases}1 & \text{if } T_c^+>\Delta\\ \alpha & \text{if } T_c^+=\Delta\\ 0 & \text{if } T_c^+<\Delta\end{cases}.$$

Moreover, these arguments also show that the test (5.2) is able to detect local alternatives converging to the null at a rate $O(n{b_n}^{-1/2} h_d^{-\frac{v^+}{v^++1}})$. When the level c decreases, the value of T_c^+ increases and the rejection probabilities also increase. On the other hand, for any given level c, the rejection probability will increase when the threshold Δ decreases (see equation (7.33) in the appendix).

(b) The bootstrap procedure can easily be modified to test the hypothesis (1.5) referring to the quantity T_c. In step (2), we additionally calculate the statistic $\hat{T}_{N,c}^-$ defined in (4.8), $\hat{T}_{N,c} = \hat{T}_{N,c}^+ + \hat{T}_{N,c}^-$ and the quantity

$$V^* = \sum_{j=1}^{n} \hat{\sigma}^2(j/n) \Big(\sum_{i=1}^{N} K_d^\dagger \Big(\frac{\tilde{\mu}_{b_n}(i/N) - \tilde{\mu}_{b_n}(0) - c}{h_d} \Big) \Big(K^* \Big(\frac{i/N - j/n}{b_n} \Big) - \bar{K}^* \Big(\frac{j}{nb_n} \Big) \Big) \Big)^2,$$

where

$$K_d^\dagger \Big(\frac{\tilde{\mu}_{b_n}(i/N) - \tilde{\mu}_{b_n}(0) - c}{h_d} \Big) = K_d \Big(\frac{\tilde{\mu}_{b_n}(i/N) - \tilde{\mu}_{b_n}(0) - c}{h_d} \Big) - K_d \Big(\frac{\tilde{\mu}_{b_n}(i/N) - \tilde{\mu}_{b_n}(0) + c}{h_d} \Big).$$

Finally, the null hypothesis (1.5) is rejected if $nNb_nh_d\big(\hat{T}_{N,c} - \Delta\big) > q_{1-\alpha}$, where $q_{1-\alpha}$ denotes the $(1-\alpha)$th quantile of a centered normal distribution with variance V^*

For the estimation of the the long-variance we define $S_{k,r} = \sum_{i=k}^{r} X_i$ and for $m \geq 2$

$$\Delta_j = \frac{S_{j-m+1,j} - S_{j+1,j+m}}{m},$$

and for $t \in [m/n, 1-m/n]$

$$\hat{\sigma}^2(t) = \sum_{j=1}^{n} \frac{m\Delta_j^2}{2} \omega(t,j), \tag{5.3}$$

where for some bandwidth $\tau_n \in (0,1)$,

$$\omega(t,i) = K\Big(\frac{i/n - t}{\tau_n}\Big) \Big/ \sum_{i=1}^{n} K\Big(\frac{i/n - t}{\tau_n}\Big).$$

For $t \in [0, m/n)$ and $t \in (1-m/n, 1]$ we define $\hat{\sigma}^2(t) = \hat{\sigma}^2(m/n)$ and $\hat{\sigma}^2(t) = \hat{\sigma}^2(1-m/n)$, respectively. Note that the estimator (5.3) does not involve estimated residuals. The following result shows that $\hat{\sigma}^2$ is consistent and can be used in Algorithm 5.1.

Theorem 5.2. *Let Assumption 2.1 - 2.2 be satisfied and assume* $\tau_n \to 0, n\tau_n \to \infty$, $m \to \infty$ *and* $\frac{m}{n\tau_n} \to 0$. *If, additionally, the function* σ^2 *is twice continuously differentiable, then the estimate defined in* (5.3) *satisfies*

$$\sup_{t\in[\gamma_n,1-\gamma_n]} |\hat{\sigma}^2(t) - \sigma^2(t)| = O_p\Big(\sqrt{\frac{m}{n\tau_n^2}} + \frac{1}{m} + \tau_n^2 + m^{5/2}/n\Big), .$$

where $\gamma_n = \tau_n + m/n$. *Moreover, we have*

$$\hat{\sigma}^2(t) - \sigma^2(t) = O_p\Big(\sqrt{\frac{m}{n\tau_n}} + \frac{1}{m} + \tau_n^2 + m^{5/2}/n\Big). \tag{5.4}$$

for any fixed $t \in (0, 1)$ *and for* $s = \{0, 1\}$

$$\hat{\sigma}^2(s) - \sigma^2(s) = O_p\Big(\sqrt{\frac{m}{n\tau_n}} + \frac{1}{m} + \tau_n + m^{5/2}/n\Big).$$

Note that error term $\sqrt{\frac{m}{n\tau_n}} + \frac{1}{m} + \tau_n^2$ in (5.4) is minimized at the rate of $O(n^{-2/7})$ by $m \asymp n^{2/7}$ and $\tau_n \asymp n^{-1/7}$, where we write $r_n \asymp s_n$ if $r_n = O(s_n)$ and $s_n = O(r_n)$. For this choice the estimator (5.3) achieves a better rate than the long-run variance estimator proposed in Zhou and Wu (2010) (see Theorem 5 in this reference).

6 Finite sample properties

In this section we investigate the finite sample properties of the bootstrap tests proposed in the previous sections. For the sake of brevity we restrict ourselves to the test (5.2) for the hypotheses (1.6). Similar results can be obtained for the corresponding tests for the hypotheses (1.5) and (1.7).

The selection of the bandwidth b_n in the local linear estimator is of particular importance in our approach, and for this purpose we use the generalized cross validation (GCV) method. To be precise, let $\tilde{e}_{i,b} = X_{i,n} - \tilde{\mu}_b(i/n)$ be the residual obtained from a bias corrected local linear fit with bandwidth b and define $\tilde{\mathbf{e}}_b = (\tilde{e}_{1,b}, \ldots, \tilde{e}_{n,b})$. Throughout this section we use the bandwidth

$$\hat{b}_n = \operatorname*{argmin}_b GCV(b) := \operatorname*{argmin}_b \frac{n^{-1}\hat{\mathbf{e}}_b \hat{\Gamma}_n^{-1} \hat{\mathbf{e}}_b}{(1 - K^*(0)/(nb))^2} \ ,$$

where $\hat{\Gamma}_n$ is an estimator of the covariance matrix $\Gamma_n := \{\mathbb{E}(\epsilon_{i,n}\epsilon_{j,n})\}_{1 \leq i,j \leq n}$, which is obtained by the banding techniques as described in Wu and Pourahmadi (2009).

It turns out that Algorithm 5.1 is not very sensitive with respect to the choice of the bandwidth h_d as long as it is chosen sufficiently small. As a rule of thumb satisfying the bandwidth conditions of Theorem 4.1(a), we use $h_d = n^{-1/2}/2$ throughout this section. In order to save computational time we use $m = \lfloor n^{2/7} \rfloor$ and $\tau_n = n^{-1/7}$ for the estimator $\hat{\sigma}^2$ in the simulation study [see the discussion at the end of Section 5]. For the data analysis in Section 6.2 we suggest a data-driven procedure and use a slight modification of the minimal volatility method as proposed by Zhou and Wu (2010). To be precise - in order to avoid choosing too large values for m and τ - we penalize the quantity

$$ISE_{h,j} = ise[\cup_{r=-2}^{2} \hat{\sigma}^2_{m_h,\tau_{j+r}}(t) \cup_{r=-2}^{2} \hat{\sigma}^2_{m_{h+r},\tau_j}(t)]$$

in their selection criteria by the term $2(\tau_j + m_h/n)IS$, where $\hat{\sigma}^2_{m_h,\tau_j}(\cdot)$ is the estimator (5.3) of the long-run variance with parameters m_h and τ_j and IS is the average of the quantities $ISE_{h,j}$.

6.1 Simulation results

All simulation results presented in this section are based on 2000 simulation runs. We consider the model (1.2) with errors $\epsilon_{i,n} = G(i/n, \mathcal{F}_i)/5$, where

(I) : $G(t, \mathcal{F}_i) = 0.25|\sin(2\pi t)|G(t, \mathcal{F}_{i-1}) + \eta_i$;

(II) : $G(t, \mathcal{F}_i) = 0.6(1 - 4(t - 0.5)^2)G(t, \mathcal{F}_{i-1}) + \eta_i$,

and the filtration $\mathcal{F}_i = (\eta_{-\infty}, \dots, \eta_i)$ is generated by a sequence $\{\eta_i, i \in \mathbb{Z}\}$ of independent standard normal distributed random variables. For the mean trend we consider the following two cases

(a): $\mu(t) = 8(-(t - 0.5)^2 + 0.25)$;

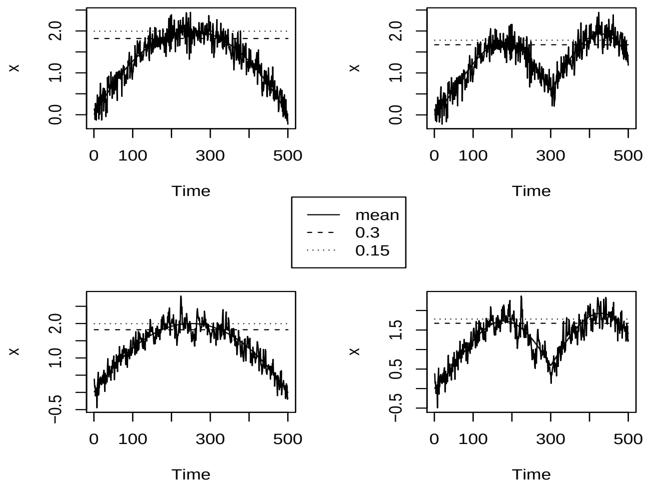

Figure 1: *Simulated sample paths for the four models under consideration. The horizontal lines display the level c which is given by* 1.82 *and* 1.995 *for the mean function (a) and by* 1.672 *and* 1.78 *for the mean function (b).*

(b): $\mu(t) = \sin(2|t - 0.6|\pi)(1 + 0.4t)$.

Typical sample paths of these processes are depicted in Figure 1. Note that the mean trend (b) is not differentiable at the point 0.6. However, using similar but more complicated arguments as given in Section 7 and 8, it can be shown that the results of this paper also hold if $\mu(\cdot)$ is Lipschitz continuous outside of an open set containing the critical roots $t_1^+, \ldots, t_{k^+}^+, t_1^-, \ldots, t_{k^-}^-$.

The threshold is given by $\Delta = 0.3$ and $\Delta = 0.15$. Following the discussion in Remark 5.1(a) we display in Tables 1 the simulated type 1 error at the boundary of the null hypothesis in (1.6), that is $T_c^+ = \Delta$. A good approximation of the nominal level at this point is required as the rejection probabilities for $T_c^+ < \Delta$ or $T_c^+ > \Delta$ are usually smaller or larger than this value, respectively. The values of c corresponding to $T_c^+ = 0.3$ and $T_c^+ = 0.15$ are given by $c = 1.82$ and $c = 1.955$ for the mean function (a) and by $c = 1.672$ and $c = 1.78$ or the mean function (b). We observe a rather precise approximation of the nominal level, which is improved with increasing sample size. For the sample size $n = 200$ the GCV method selects the bandwidths b_{cv} for 0.25, 0.26, 0.23, 0.19 for the models $((I), (a))$, $((I), (b))$, $((II), (a))$, and $((II), (b))$, respectively. Similarly, for the sample size $n = 500$ the GCV method selects the bandwidths 0.2, 0.17, 0.21, 0.14 for the models $((I),(a))$, $((I),(b))$, $((II),(a))$ and $((II),(b))$, respectively. In order to study the robustness of the test with respect to the choice of b_n we investigate the bandwidths $b_{cv}^- = b_{cv} - 0.05, b_{cv}, b_{cv}^+ = b_{cv} + 0.05$. For this range of bandwidths the approximation of the nominal level is remarkably stable.

Table 1: *Simulated level of the test* (5.2) *at the boundary of the null hypothesis* (1.6)*. The sample size is* $n = 200$ *(upper part) and* $n = 500$ *(lower part) and various bandwidths are considered. The bandwidth* b_{cv} *is chosen by GCV, and* $b_{cv}^- = b_{cv} - 0.05$*,* $b_{cv}^+ = b_{cv} + 0.05$.

n		model	(a,I)		(b,I)		(a,II)		(b,II)	
	Δ	b_n	5%	10%	5%	10%	5%	10%	5%	10%
200	0.3	b_{cv}^-	4	8.95	5.35	10.1	4.9	8.8	5.6	9.35
		b_{cv}	3.5	8.2	4.15	8.05	4	8	6	10.7
		b_{cv}^+	4.15	7.6	2.85	5.3	3.75	6.85	4.85	9.15
	0.15	b_{cv}^-	5.45	8.75	5.8	9.25	6.9	10	6.45	11.55
		b_{cv}	6.45	10.8	5.35	8.7	6.45	10.7	7.25	11.05
		b_{cv}^+	5.65	10.05	2.45	4.55	6.4	10.15	5.75	9.95
500	0.3	b_{cv}^-	5.2	9.45	5.85	10.1	5.85	10.05	5.55	9.9
		b_{cv}	4.6	9.55	5.45	9.85	5.65	9.25	6	10.1
		b_{cv}^+	5.15	9.1	5	8.95	3.65	7.15	5.45	9.85
	0.15	b_{cv}^-	7.6	12.1	6.5	9.6	7,7	11.15	7.5	11.3
		b_{cv}	6.55	11.25	5.1	9.15	7.75	12.2	5.15	9.25
		b_{cv}^+	6.85	10.6	4.4	7.5	6.6	11.05	4.6	8.3

In Figure 2, we investigate the properties of the test (5.2) as a function of the threshold Δ and level c, where we restrict ourselves to the scenario $((I),(a))$. For the other cases the observations are similar. The bandwidth is $b_n = 0.2$. In the left part of the figure the level c is fixed as 1.82 and Δ varies from 0 to 0.4 (where the true threshold is $\Delta = 0.3$). As expected the rejection probabilities decrease with an increasing threshold Δ. Similarly, in the right part of Figure 2 we display the rejection probabilities for fixed $\Delta = 0.3$ when c varies between 1.44 and 2. Again the rejection rates decrease when c increases.

We finally investigate the power of the test (5.2) for the hypotheses (1.6) with $c = 1.82$ and $\Delta = 0.3$, where the bandwidth is chosen as $b_n = 0.2$. The model is given by (1.2) with error (I) and different mean functions

$$\mu(t) = a(-(t-0.5)^2 + 0.25) \ , \quad a \in [7.5, 9.5] \tag{6.1}$$

are considered (here the case $a = 8$ corresponds to the boundary of the hypotheses). The results are presented in Figure 3, which demonstrate that the test (5.2) has decent power.

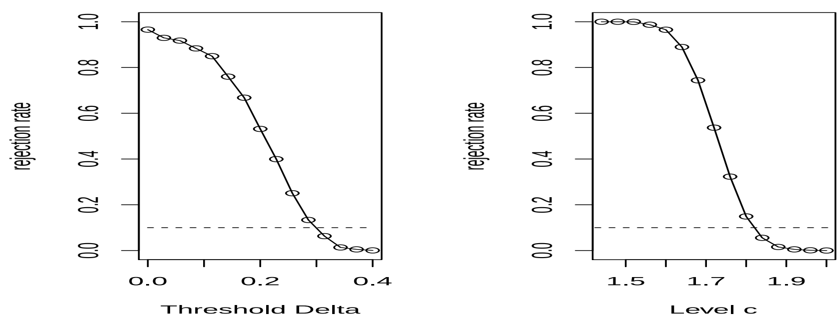

Figure 2: *Simulated rejection probabilities of the test* (5.2) *in model* (1.2) *for varying values of* c *and* Δ*. Left:* $c = 1.82$, $\Delta \in [0, 0.4]$ *(the case* $\Delta = 0.3$ *corresponds to the boundary of the null hypothesis). Right:* $\Delta = 0.3$, $c \in [1.44, 2]$ *(the case* $c = 1.82$ *corresponds to the boundary of the null hypothesis).The dashed horizontal line represents the nominal level 10%.*

6.2 Empirical Studies

6.2.1 Global temperature data

Global temperature data has been extensively studied in the statistical literature under the assumption of stationarity [see for example Bloomfield and Nychka (1992), Vogelsang (1998) and Wu and Zhao (2007) among others]. We consider here a series from http://cdiac.esd.ornl.gov/ftp/trends/temp/jonescru/ with global monthly temperature anomalies from January 1850 to April 2015, relative to the $1961 - 1990$ mean. The data and a local linear estimate of the mean function are depicted in left panel of Figure 4. The figure indicates a non-constant higher order structure of the series and analyzing this series under the assumption of stationarity might be questionable. In fact, the test of Dette et al.

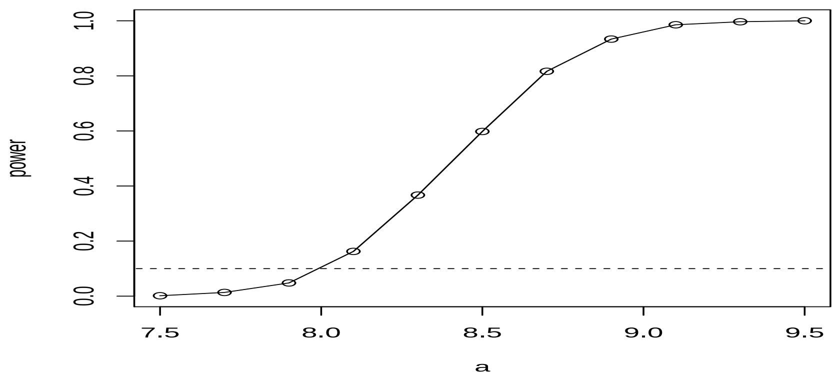

Figure 3: *Simulated power of the test* (5.2) *in model* (1.2) *for the hypothesis* (1.6) *with* $c = 1.82$ *and* $\Delta = 0.3$. *The mean functions are given by* (6.1) *and the case* $a = 8$ *corresponds to the boundary of the null hypothesis. The dashed horizontal line represents the nominal level 10%.*

(2015) for a constant lag-1 correlation yields a p-value of 1.6% supporting a non-stationary model for data analysis.

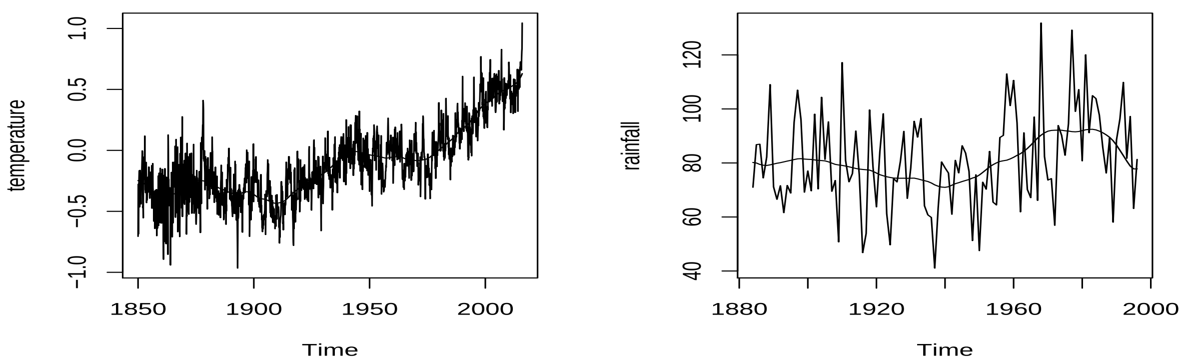

Figure 4: *Left panel: deseasonalized global temperature 1850–2015 and its fitted mean-trend. Right panel: Yearly Rainfall of Tucumán Province, Argentina, 1884–1996.*

We are interested in the question if the deseasonalized monthly temperature exceeds the temperature in January 1850 by more than $c = 0.15$ degrees Celsius in more than $100\Delta\%$ of the considered period. For this purpose we run the test (5.2) for the hypothesis (1.6), where the bandwidth (chosen by GCV) is $b_n = 0.105$ and $h_d = 0.011$ (we note again that the procedure is rather stable with respect to the choice of h_d). For the estimate

(5.3) of the long-run variance σ^2, we use the procedure described at the beginning of this section, which yields $m = 30$ and $\tau = 0.202$. For a threshold $\Delta = 43.4\%$ we obtain a p-value of 4.82%.

Next we investigate the same question for the sub-series from January 1850 to December 1974. The GCV method yields the bandwidth $b_n = 0.135$ and we chose $h_d = 0.013$ and $m = 36$, $\tau = 0.234$ for the estimate of the time-varying long-run variance (see the discussion at the beginning of this section). We find that for $\Delta = 26\%$ and $c = 0.15$ the p-value is 6.6%. Comparing the results for the series and sub-series shows that relevant deviations of more than $c = 0.15$ degrees Celsius arise more frequently between 1975 and 2015. The conclusions of this short data analysis are similar to those of many authors, but by our method we are able to quantitatively describe relevant deviations. For example, if we reject the hypothesis that in less than 26% of the time between January 1850 and April 2015 the mean function exceeds its value from January 1850 by more than $c = 0.15$ degrees Celsius, the type I error of this conclusion is less or equal than 5%.

6.2.2 Rainfall data

In this example we analyze the yearly rainfall data (in millimeters) from 1884 to 1996 in the Tucumán Province, Argentina, which is a predominantly agriculture region. Therefore its economy well-being depends sensitively on timely rainfall. The series with a local linear estimate of the mean trend are depicted in right panel of Figure 4 (note that the range of estimated mean function is [71.0mm, 92.5mm]) and it has been studied by several authors in the context of change point analysis with different conclusions. For example, the null hypothesis of no change point is rejected by the conventional CUSUM test, isotonic regression approach of Wu et al. (2001) rejects the hypothesis with p-value smaller than 0.1%, and the robust bootstrap test of Zhou (2013) with a p-value smaller than 2%. On the other hand a self-normalization method considered in Shao and Zhang (2010) reports a p-value about 10%.

Meanwhile, there is some belief that there exists a change point because of the construction of a dam near the region during 1952 – 1962. As a result, a more practical question is whether the construction of the dam has a relevant influence on the economic well-being of the region via affecting the annual rainfall. To investigate this question, we are testing the hypotheses (1.6) with a threshold $\Delta = 0.05$ (here we calculated $b_n = 0.235$, $m = 11$, $\tau = 0.24$ and $h_d = 0.047$ as described at the beginning of this section). For the level $c = 7$ the p-value is 6.05%. In other words the hypothesis that in less than 5% of the 113 years the mean annual rainfall is at least 7mm higher than the rainfall in the year 1880 can not be rejected. This result indicates that the effect of the new dam on the change of the amount of rainfall is small.

Acknowledgements. The authors would like to thank Martina Stein who typed this manuscript with considerable technical expertise. The work of the authors was supported by the Deutsche Forschungsgemeinschaft (SFB 823: Statistik nichtlinearer dynamischer Prozesse, Teilprojekt A1 and C1).

References

Álvarez Esteban, P. C., Barrio, E. D., Cuesta-Albertos, J. A., and Matran, C. (2008). Trimmed comparison of distributions. *Journal of the American Statistical Association*, 103(482):697–704.

Álvarez Esteban, P. C., del Barrio, E., Cuesta-Albertos, J. A., and Matran, C. (2012). Similarity of samples and trimming. *Bernoulli*, 18(2):606–634.

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4):128–156.

Aue, A., Hörmann, S., Horváth, L., and Reimherr, M. (2009). Break detection in the covariance structure of multivariate time series models. *Annals of Statistics*, 37(6B):4046–4087.

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

Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 66(1):47–78.

Baillo, A. (2003). Total error in a plug-in estimator of level sets. *Statistics & Probability Letters*, 65(4):411 – 417.

Bloomfield, P. and Nychka, D. (1992). Climate spectra and detecting climate change. *Climatic Change*, 21(3):275–287.

Brown, R., Durbin, J., and Evans, J. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society Series B*, 37(2):149–163.

Cadre, B. (2006). Kernel estimation of density level sets. *Journal of Multivariate Analysis*, 97(4):999 – 1023.

Chandler, G. and Polonik, W. (2006). Discrimination of locally stationary time series based on the excess mass functional. *Journal of the American Statistical Association*, 101(473):240–253.

Cheng, M.-Y. and Hall, P. (1998). Calibrating the excess mass and dip tests of modality. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(3):579–589.

Chernozhukov, V., Fernandéz-Val, I., and Galichon, A. (2009). Improving point and interval estimators of monotone functions by rearrangement. *Biometrika*, 96(3):559–575.

Chernozhukov, V., Fernandéz-Val, I., and Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica*, 78(3):1093–1125.

Chow, G. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, 28(3):591–605.

Cuevas, A., González-Manteiga, W., and Rodríguez-Casal, A. (2006). Plug-in estimation of general level sets. *Australian & New Zealand Journal of Statistics*, 48(1):7–19.

Dahlhaus, R. et al. (1997). Fitting time series models to nonstationary processes. *The annals of Statistics*, 25(1):1–37.

Dette, H., Neumeyer, N., and Pilz, K. F. (2006). A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli*, 12:469–490.

Dette, H. and Volgushev, S. (2008). Non-crossing non-parametric estimates of quantile curves. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(3):609–627.

Dette, H. and Wied, D. (2016). Detecting relevant changes in time series models. *Journal of the Royal Statistical Society, Ser. B*, 78:371–394.

Dette, H., Wu, W., and Zhou, Z. (2015). Change point analysis of second order characteristics in non-stationary time series. *arXiv preprint arXiv:1503.08610*.

Hartigan, J. A. and Hartigan, P. M. (1985). The dip test of unimodality. *The Annals of Statistics*, pages 70–84.

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

Krämer, W., Ploberger, W., and Alt, R. (1988). Testing for structural change in dynamic models. *Econometrica*, 56(6):1355–1369.

Mason, D. M. and Polonik, W. (2009). Asymptotic normality of plug-in level set estimates. *Ann. Appl. Probab.*, 19(3):1108–1142.

Müller, D. W. and Sawitzki, G. (1991). Excess mass estimates and tests for multimodality. *Journal of the American Statistical Association*, 86(415):738–746.

Nason, G. P., von Sachs, R., and Kroisandt, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *Journal of the Royal Statistical Society, Ser. B*, 62:271–292.

Ombao, H., von Sachs, R., and Guo, W. (2005). SLEX analysis of multivariate non-stationary time series. *Journal of the American Statistical Association*, 100:519–531.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, 41((1-2)).

Polonik, W. (1995). Measuring mass concentrations and estimating density contour clusters – an excess mass approach. *Annals of Statistics*, 23(3):855–881.

Polonik, W. and Wang, Z. (2006). Estimation of regression contour clusters – an application of the excess mass approach to regression. *Journal of Multivariate Analysis*, 94(2):227–249.

Rinaldo, A. and Wasserman, L. (2010). Generalized density clustering. *The Annals of Statistics*, pages 2678–2722.

Samworth, R. J. and Wand, M. P. (2010). Asymptotics and optimal bandwidth selection for highest density region estimation. *Ann. Statist.*, 38(3):1767–1792.

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

Shao, X. and Zhang, X. (2010). Testing for change points in time series. *Journal of the American Statistical Association*, 105(491):1228–1240.

Tsybakov, A. B. (1997). On nonparametric estimation of density level sets. *The Annals of Statistics*, 25(3):948–969.

Vogelsang, T. J. (1998). Trend function hypothesis testing in the presence of serial correlation. *Econometrica*, pages 123–148.

Vogt, M. (2012). Nonparametric regression for locally stationary time series. *Annals of Statistics*, 40(5):2601–2633.

Wellek, S. (2010). *Testing Statistical Hypotheses of Equivalence and Noninferiority*. CRC Press.

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

Wu, W. B., Woodroofe, M., and Mentz, G. (2001). Isotonic regression: Another look at the changepoint problem. *Biometrika*, 88(3):793–804.

Wu, W. B. and Zhao, Z. (2007). Inference of trends in time series. *Journal of the Royal Statistical*

Society: Series B (Statistical Methodology), 69(3):391–410.

Zhou, Z. (2010). Nonparametric inference of quantile curves for nonstationary time series. *The Annals of Statistics*, 38(4):2187–2217.

Zhou, Z. (2013). Heteroscedasticity and autocorrelation robust structural change detection. *Journal of the American Statistical Association*, 108(502):726–740.

Zhou, Z. and Wu, W. B. (2009). Local linear quantile estimation for nonstationary time series. *The Annals of Statistics*, 37(5):2696–2729.

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

7 Proofs of main results

In this section we will prove the main results of this paper. For the sake of a simple notation we write $e_i := \epsilon_{i,n}$ throughout this section, where $\epsilon_{i,n}$ is the nonstationary error process in model (1.2). Moreover, in all arguments given below M denotes a sufficiently large constant which may vary from line to line. For the sake of brevity we will restrict ourselves to proofs of the results in Section 4, while the details for the proofs of the results in Section 3 are omitted as they follow by similar arguments as presented here.

Proof of Theorem 4.1. It follows from Assumption 2.1 that there exist $k^+ \geq 1$ roots $t_1^+ < \ldots < t_{k^+}^+$ of the equation $\mu(t) = \mu(0) + c$. Define $\gamma^+ = \min_{0 \leq i \leq k^+}(t_{i+1}^+ - t_i^+) > 0$, with the convention that $t_0^+ = 0$ and $t_{k^++1}^+ = 1$. Recalling the definition of the statistic $\tilde{T}_{N,c}^+$ and the quantity $T_{N,c}^+$ in (4.4) and (2.4), respectively, we obtain the decomposition

$$\tilde{T}_{N,c}^+ - T_{N,c}^+ = \Delta_{1,N} + \Delta_{2,N}, \tag{7.1}$$

where the random variables $\Delta_{1,N}$ and $\Delta_{1,N}$ are defined by

$$\begin{aligned} \Delta_{1,N} &= \frac{1}{N}\sum_{i=1}^{N}\int_c^\infty \frac{1}{h_d^2}K_d'\Big(\frac{\mu(\frac{i}{N}) - \mu(0) - u}{h_d}\Big)(\tilde{\mu}_{b_n}(\tfrac{i}{N}) - \mu(\tfrac{i}{N}) - (\tilde{\mu}_{b_n}(0) - \mu(0)))du, \\ \Delta_{2,N} &= \frac{1}{2N}\sum_{i=1}^{N}\int_c^\infty \frac{1}{h_d^3}K_d''\Big(\frac{\zeta_i - u}{h_d}\Big)(\tilde{\mu}_{b_n}(\tfrac{i}{N}) - \mu(\tfrac{i}{N}) - (\tilde{\mu}_{b_n}(0) - \mu(0)))^2du \end{aligned} \tag{7.2}$$

(note that we do reflect the dependence of $\Delta_{\ell,N}$ on n in our notation) and ζ_i denotes a random variable satisfying $|\zeta_i - (\mu(i/N) - \mu(0))| \leq |\tilde{\mu}_{b_n}(i/N) - \mu(i/N) - (\tilde{\mu}_{b_n}(0) - \tilde{\mu}(0))|$ and $|\zeta_i - (\tilde{\mu}_{b_n}(i/N) - \tilde{\mu}_{b_n}(0))| \leq |\tilde{\mu}_{b_n}(i/N) - \mu(i/N) - (\tilde{\mu}_{b_n}(0) - \mu(0))|$. It is easy to see that

$$|2\Delta_{2,N}| = \Big|\frac{1}{N}\sum_{i=1}^{N}\frac{1}{h_d^2}K_d'\Big(\frac{\zeta_i - c}{h_d}\Big)(\tilde{\mu}_{b_n}(i/N) - \mu(i/N) - (\tilde{\mu}_{b_n}(0) - \mu(0)))^2du\Big|. \tag{7.3}$$

Recall the definition of π_n in (4.5) and define

$$A_n = \Big\{\sup_{t\in[b_n,1-b_n]\cup\{0\}}|\tilde{\mu}_{b_n}(t) - \mu(t)| \leq \pi_n, \sup_{t\in[0,b_n)\cup(1-b_n,b_n]}|\tilde{\mu}_{b_n}(t) - \mu(t)| \leq b_n^2 \vee \pi_n\Big\}, \tag{7.4}$$

where we denote $\max\{a,b\}$ by $a \vee b$. By Lemma 8.3 in Section 8, we have $\lim_{n\to\infty} \mathbb{P}(A_n) = 1$ and Lemma 8.1 yields

$$\begin{aligned}\sharp\{i : |\tilde{\mu}_{b_n}(i/N) - \tilde{\mu}_{b_n}(0) - c| \le h_d, |\tilde{\mu}_{b_n}(i/N) - \mu(i/N) - (\tilde{\mu}_{b_n}(0) - \mu(0))| \le 2\pi_n\} \\ \le \sharp\{i : |\mu(i/N) - \mu(0) - c| \le h_d + 2\pi_n\} = O(N(h_d + \pi_n)^{1/(v^+ +1)}) \quad (7.5)\end{aligned}$$

almost surely, where $\sharp A$ denotes the number of points in the set A. Observing the definition of ζ_i and (7.5) we obtain that the number of non-vanishing terms on the right hand side of equality (7.3) is bounded by $O(N(h_d + \pi_n)^{\frac{1}{v+1}})$. Therefore the triangle inequality yields for a sufficiently large constant M

$$\|\Delta_{2,N}\mathbf{1}(A_n)\|_2 \le M\Big(b_n^6 + \frac{1}{nb_n}\Big)h_d^{-2}((h_d + \pi_n)^{\frac{1}{v^+ +1}}).$$

Now Proposition B.3 of Dette et al. (2015) (note that $\lim_{n\to\infty} \mathbb{P}(A_n) = 1$) yields the estimate

$$\Delta_{2,N} = O_p\Big((b_n^6 + \frac{1}{nb_n})h_d^{-2}((h_d + \pi_n)^{\frac{1}{v^+ +1}}\Big).$$

Notice that the assumptions regarding bandwidths guarantee that

$$\sqrt{nh_d}\, h_d^{\frac{v^+}{2(v^+ +1)}} \Delta_{2,N} = o(1), \quad \text{if} \quad b_n^{v^+ +1}/h_d = c \in [0,\infty), \quad (7.6)$$

$$\sqrt{nb_n}\, h_d^{\frac{v^+}{v^+ +1}} \Delta_{2,N} = o(1), \quad \text{if} \quad b_n^{v^+ +1}/h_d \to \infty, \quad (7.7)$$

and therefore it remains to consider the term $\Delta_{1,N}$ in the decomposition (7.1).

For this purpose we recall its definition in (7.2) and obtain by an application of Lemma 8.3 and straightforward calculations the following decomposition

$$\begin{aligned}\Delta_{1,n} &= \frac{-1}{Nh_d}\sum_{i=1}^{N} K_d\Big(\frac{\mu(i/N) - \mu(0) - c}{h_d}\Big)((\tilde{\mu}_{b_n}(i/N) - \mu(i/N)) - (\tilde{\mu}_{b_n}(0) - \mu(0))) \\ &= I + R, \quad (7.8)\end{aligned}$$

where the terms I and R are defined by

$$I = \frac{-1}{nNb_nh_d}\sum_{i=1}^{N} K_d\Big(\frac{\mu(\frac{i}{N})-\mu(0)-c}{h_d}\Big)\sum_{j=1}^{n} e_j\Big(K^*\Big(\frac{\frac{i}{N}-\frac{j}{n}}{b_n}\Big)-\bar{K}^*\Big(\frac{j}{nb_n}\Big)\Big), \qquad (7.9)$$

$$R = O\Big(\frac{1}{Nh_d}\sum_{i=1}^{N} K_d\Big(\frac{\mu(\frac{i}{N})-\mu(0)-c}{h_d}\Big)\Big(b_n^3+\frac{1}{nb_n}\Big)\Big).$$

By Lemma 8.1 the term R is of order $O(h_d^{-\frac{v^+}{v^++1}}(b_n^3+\frac{1}{nb_n}))$. For the investigation of the remaining term I, we use Proposition 5 of Zhou (2013), which shows that there exist (on a possibly richer probability space), independent stand normal distributed random variables $\{V_i\}_{i\in\mathbb{Z}}$, such that

$$\max_{1\le i\le n}|\sum_{j=1}^{i} e_j - \sum_{j=1}^{i}\sigma(j/n)V_j| = o_p(n^{1/4}\log^2 n).$$

This representation and the summation by parts formula in equation (44) of Zhou (2010) yield

$$\sup_{t\in[0,1]}\Big|\sum_{j=1}^{n} e_j\tilde{K}^*\Big(\frac{t-j/n}{b_n}\Big)-\sum_{j=1}^{i}\sigma(j/n)V_j\tilde{K}^*\Big(\frac{t-j/n}{b_n}\Big)\Big| = o_p(n^{1/4}\log^2 n),$$

where we introduce the notation

$$\tilde{K}^*\Big(\frac{t-j/n}{b_n}\Big) = K^*\Big(\frac{t-j/n}{b_n}\Big)-\bar{K}^*\Big(\frac{j}{nb_n}\Big). \qquad (7.10)$$

Using these results in (7.9) and Lemma 8.1 provides an asymptotically equivalent representation of the term I, that is

$$|I'-I| = o_p\Big(\frac{n^{1/4}\log^2 n}{nb_n}h_d^{-\frac{v^+}{v^++1}}\Big). \qquad (7.11)$$

Here

$$I' := \frac{-1}{nNb_nh_d}\sum_{j=1}^{n}\sum_{i=1}^{N} K_d\Big(\frac{\mu(i/N)-\mu(0)-c}{h_d}\Big)\sigma(j/n)\tilde{K}^*\Big(\frac{i/N-j/n}{b_n}\Big)V_j$$

is a zero mean Gaussian random variable with variance

$$\begin{aligned}\mathrm{Var}(I') &= \frac{1}{n^2b_n^2h_d^2}\sum_{j=1}^{n}\Big(\frac{1}{N}\sum_{i=1}^{N}\sigma(j/n)\tilde{K}^*\Big(\frac{i/N-j/n}{b_n}\Big)K_d\Big(\frac{\mu(i/N)-\mu(0)-c}{h_d}\Big)\Big)^2 \\ &= \frac{1}{n^2b_n^2h_d^2}\sum_{j=1}^{n}\Big(\int_0^1 \sigma(j/n)\tilde{K}^*\Big(\frac{t-j/n}{b_n}\Big)K_d\Big(\frac{\mu(t)-\mu(0)-c}{h_d}\Big)dt\Big)^2+\beta_n \\ &:= \bar{\alpha}_n+\beta_n, \end{aligned} \tag{7.12}$$

and the last two equalities define the quantities $\bar{\alpha}_n$ and β_n in an obvious manner. Observing the estimates

$$\begin{aligned}&\frac{1}{N}\sum_{i=1}^{N}\sigma(\frac{j}{n})\tilde{K}^*\Big(\frac{i/N-j/n}{b_n}\Big)K_d\Big(\frac{\mu(i/N)-\mu(0)-c}{h_d}\Big)- \\ &\int_0^1\sigma(\frac{j}{n})\tilde{K}^*\Big(\frac{t-j/n}{b_n}\Big)K_d\Big(\frac{\mu(t)-\mu(0)-c}{h_d}\Big)dt = O\Big((\frac{1}{Nb_n}+\frac{1}{Nh_d})(b_n\wedge h_d^{\frac{1}{v^++1}})\Big), \\ &\frac{1}{n^2b_n^2h_d^2}\sum_{j=1}^{n}\Big(\int_0^1\sigma(j/n)\tilde{K}^*\Big(\frac{t-j/n}{b_n}\Big)K_d\Big(\frac{\mu(t)-\mu(0)-c}{h_d}\Big)dt\Big) = O\Big(\frac{h_d^{\frac{-v^+}{v^++1}}}{nb_nh_d}\Big),\end{aligned}$$

we have that

$$\beta_n = \frac{h_d^{\frac{-v^+}{v^++1}}}{nb_nh_d}\Big(\frac{1}{Nb_n}+\frac{1}{Nh_d}\Big)(b_n\wedge h_d^{\frac{1}{v^++1}})+\Big((\frac{1}{Nb_n}+\frac{1}{Nh_d})(b_n\wedge h_d^{\frac{1}{v^++1}})\Big)^2, \tag{7.13}$$

where $a\wedge b := \min(a,b)$.

For the calculation of $\bar{\alpha}_n$ we note that

$$\bar{K}^*\Big(\frac{j/n}{b_n}\Big)K^*\Big(\frac{t-j/n}{b_n}\Big)K_d\Big(\frac{\mu(t)-\mu(0)-c}{h_d}\Big)=0. \tag{7.14}$$

for sufficiently large n. This statement follows because by Lemma 8.1 the third factor vanishes outside of (shrinking) neighbourhoods $\mathcal{U}_1, \ldots, \mathcal{U}_{k^+}$ of the points $t_1^+, \ldots, t_{k^+}^+$ with Lebesgue measure of order $h_d^{\frac{1}{v_l+1}}$, $(1 \le l \le k^+)$. Consequently, the product of the first and second factor vanishes, wheneever the point j/n is **not** an element of the set

$$\Big\{ s + t \ \Big|\ t \in \cup_{j=1}^{k^+} \mathcal{U}_j \ ; s \in [-b_n, b_n] \Big\}.$$

However, if n is sufficiently large the intersection of this set with the interval $[0, b_n]$, is empty. Consequently, for sufficiently large n there exists no pair $(t, j/n)$ such that all factors in (7.14) different from zero.
Therefore, we obtain (recalling the notation of $\tilde{K}^*$ in (7.10))

$$\bar{\alpha}_n = \alpha_n + \tilde{\alpha}_n, \tag{7.15}$$

where

$$\begin{aligned} \alpha_n &= \frac{1}{n^2 b_n^2 h_d^2} \sum_{j=1}^{n} \Big(\int_0^1 \sigma(j/n) K^* \Big(\frac{t - j/n}{b_n} \Big) K_d \Big(\frac{\mu(t) - \mu(0) - c}{h_d} \Big) dt \Big)^2, \\ \tilde{\alpha}_n &= \frac{1}{n^2 b_n^2 h_d^2} \sum_{j=1}^{n} \Big(\int_0^1 \sigma(j/n) \bar{K}^* \Big(\frac{j/n}{b_n} \Big) K_d \Big(\frac{\mu(t) - \mu(0) - c}{h_d} \Big) dt \Big)^2. \end{aligned} \tag{7.16}$$

At the end of the proof we will show that

$$\alpha_n = \begin{cases} h_d^{\frac{-2v^+}{v^++1}} (n b_n)^{-1} \sigma_1^{2,+} & \text{if } b_n^{v^++1}/h_d \to \infty \\ \\ h_d^{\frac{1}{v^++1}} (n h_d^2)^{-1} \rho_1^{2,+} & \text{if } b_n^{v^++1}/h_d \to r \in [0, \infty) \end{cases} \tag{7.17}$$

$$\tilde{\alpha}_n = \begin{cases} h_d^{\frac{-2v^+}{v^++1}} (n b_n)^{-1} \sigma_2^{2,+} & \text{if } b_n^{v^++1}/h_d \to \infty \\ \\ h_d^{\frac{1}{v^++1}} (n h_d^2)^{-1} \rho_2^{2,+} & \text{if } b_n^{v^++1}/h_d \to r \in [0, \infty) \end{cases} . \tag{7.18}$$

where $\sigma_1^{2,+}$, $\sigma_2^{2,+}$, $\rho_1^{2,+}$ and $\rho_2^{2,+}$ are defined in Theorem 4.1. The assertion now follows from (7.1), (7.6), (7.7), (7.8) and (7.11) observing that the random variable I' is normally distributed, where the (asymptotic) variance can be obtained from (7.12), (7.13), (7.15), (7.17) and (7.18).

Proof of (7.17): By Lemma 8.1 with m replaced by $\mu(0)+c$, there exists a small positive number $0<\epsilon<\gamma^+/4$ such that when n is sufficiently large, we have

$$\begin{aligned}
\alpha_n &= \frac{1}{n^2 b_n^2 h_d^2}\sum_{j=1}^{n}\Big(\sum_{l=1}^{k^+}\int_{t_l^+-\epsilon}^{t_l^++\epsilon}\sigma(j/n)K^*\Big(\frac{t-j/n}{b_n}\Big)K_d\Big(\frac{\mu(t)-\mu(0)-c}{h_d}\Big)dt\Big)^2\\
&= \frac{1}{n^2 b_n^2 h_d^2}\sum_{j=1}^{n}\sum_{l=1}^{k^+}\Big(\int_{t_l^+-\epsilon}^{t_l^++\epsilon}\sigma(j/n)K^*\Big(\frac{t-j/n}{b_n}\Big)K_d\Big(\frac{\mu(t)-\mu(0)-c}{h_d}\Big)dt\Big)^2\\
&= \frac{1}{n^2 b_n^2 h_d^2}\sum_{j=1}^{n}\sum_{l=1}^{k^+}\alpha_{n,l,j}^2,
\end{aligned}$$

where the last equation defines the quantities $\alpha_{n,l,j}^2$ in an obvious manner. We now calculate α_n for the two bandwidth conditions in (7.17).

(i) We begin with the case $b_n^{v^++1}/h_d\to\infty$, which means $b_n^{v_l^++1}/h_d\to\infty$ for $l=1,\ldots,k$. By Lemma 8.1 there exists a sufficiently large constant M such that

$$\alpha_{n,l,j} = \int_{t_l^+-Mh_d^{\frac{1}{v_l^++1}}}^{t_l^++Mh_d^{\frac{1}{v_l^++1}}}\sigma(t_l^+)K^*\Big(\frac{t-j/n}{b_n}\Big)K_d\Big(\frac{\mu(t)-\mu(0)-c}{h_d}\Big)dt\Big(1+O\Big(h_d^{\frac{1}{v_l^++1}}\Big)\Big) \tag{7.19}$$

Observing the fact that the kernel $K_d(\cdot)$ is bounded and continuous we obtain by a Taylor expansion of $\mu(t)-\mu(0)-c$ around t_l^+,

$$\big|\alpha_{n,l,j}-\alpha_{n,l,j}^*\big| = O\Big(h_d^{\frac{2}{v_l^++1}}\mathbf{1}(|j/n-t_l^+|\le 2b_n)\Big) \tag{7.20}$$

uniformly with respect to $t\in[0,1]$, where

$$\alpha_{n,l,j}^* = \int_{t_l^+-Mh_d^{\frac{1}{v_l^++1}}}^{t_l^++Mh_d^{\frac{1}{v_l^++1}}}\sigma(t_l^+)K^*\Big(\frac{t-j/n}{b_n}\Big)K_d\Big(\frac{\mu^{(v_l^++1)}(t_l^+)(t-t_l^+)^{v_l^++1}}{(v_l^++1)!h_d}\Big)dt.$$

Substituting $t = t_l^+ + z\big|h_d(v_l^+ + 1)!/\mu^{(v_l^+ + 1)}(t_l^+)\big|^{\frac{1}{v_l^+ + 1}}$, observing the symmetry of $K_d(\cdot)$ and using a Taylor expansion shows that

$$\alpha^*_{n,l,j} = \Big|\frac{(v_l^+ + 1)!h_d}{\mu^{(v_l^+ + 1)}(t_l^+)}\Big|^{\frac{1}{v_l^+ + 1}}\Big(\int K_d(z^{v_l^+ + 1})dz\Big)\sigma(t_l^+)K^*\Big(\frac{t_l^+ - j/n}{b_n}\Big) \\ + O\Big(h_d^{\frac{2}{v_l^+ + 1}}(b_n)^{-1}\mathbf{1}(|j/n - t_l^+| \le 2b_n)\Big), \tag{7.21}$$

where we have used the fact that $\int zK_d(z^{v_l^+ + 1})dz < \infty$ since $K_d(\cdot)$ has a compact support. Equations (7.19)–(7.21) and the condition $\frac{b_n^{v^+ + 1}}{h_d} \to \infty$ now give

$$\begin{aligned}\alpha_n &= \frac{1}{n^2 b_n^2 h_d^2}\sum_{l=1}^{k}\sum_{j=1}^{n}\Big(\Big(\int K_d(z^{v_l^+ + 1})dz\Big)\sigma(t_l^+)\Big|\frac{(v^+ + 1)!h_d}{\mu^{(v^+ + 1)}(t_l^+)}\Big|^{\frac{1}{v^+ + 1}}K^*\Big(\frac{t_l^+ - j/n}{b_n}\Big)\Big)^2 \\ &\qquad \times \Big(1 + O\big(h_d^{\frac{1}{v^+ + 1}}{b_n}^{-1}\mathbf{1}(|j/n - t_l^+| \le 2b_n)\big)\Big) \\ &= \sum_{l=1}^{k}\frac{h_d^{\frac{-2v_l^+}{v_l^+ + 1}}}{nb_n}\Big(\int K_d(z^{v_l^+ + 1})dz\Big)^2((v_l^+ + 1)!)^{\frac{2}{v_l^+ + 1}}\Big(\frac{\sigma(t_l^+)}{|\mu^{(v_l^+ + 1)}(t_l^+)|^{\frac{1}{v_l^+ + 1}}}\Big)^2\int (K^*(x))^2 dx \\ &\qquad \times \big(1 + O\big((nb_n)^{-1} + h_d^{\frac{1}{v_l^+ + 1}}/b_n\big)\big) \\ &= h_d^{\frac{-2v^+}{v^+ + 1}}(nb_n)^{-1}\sigma_1^{2,+}\big(1 + o(1)\big)\end{aligned}$$

which proves (7.17) in the case $b_n^{v^+ + 1}/h_d \to \infty$.

Next we turn to the case $b_n/h_d^{\frac{1}{v^+ + 1}} \to c \in [0, \infty)$, introduce the notation $\alpha_{n,l} = \frac{1}{n^2 b_n^2 h_d^2}\sum_{j=1}^{n}\alpha^2_{n,l,j}$ and note that

$$\alpha_n = \sum_{l=1}^{k}\alpha_{n,l}. \tag{7.22}$$

Define $c_l = b_n^{v_l^+ + 1}/h_d$ for $l \in \{1, \ldots, k^+\}$. For those l satisfying $c_l \to \infty$, we have already shown that

$$\alpha_{n,l} = \frac{h_d^{\frac{-2v_l^+}{v_l^++1}}}{nb_n} = o\Big(\frac{h_d^{\frac{1}{v_l^++1}}}{nh_d^2}\Big) = o\Big(\frac{h_d^{\frac{1}{v^++1}}}{nh_d^2}\Big).$$

In the following discussion we prove that for those l, for which c_l does not converge to infinity, the quantity $\alpha_{n,l}$ is exactly of order $O(h_d^{\frac{1}{v_l^++1}}(nh_d^2)^{-1})$. For this purpose define

$$\alpha'_{n,l} = \frac{1}{nb_n^2h_d^2}\int_0^1 \big(G(t_l^+, s, b_n, h_d)\big)^2 ds. \tag{7.23}$$

where

$$G(t_l^+, s, b_n, h_d) = \int_{t_l^+-\epsilon}^{t_l^++\epsilon} \sigma(s)K^*\Big(\frac{t-s}{b_n}\Big)K_d\Big(\frac{\mu(t)-\mu(0)-c}{h_d}\Big)dt.$$

It follows from a Taylor expansion and an approximation by a Riemann sum that

$$|\alpha_{n,l} - \alpha'_{n,l}| \le \frac{1}{nb_n^2h_d^2}\sum_{j=1}^n \frac{1}{n^2} \sup_{\frac{j-1}{n}\le s\le\frac{j}{n}} \big|G(t_l^+, s, b_n, h_d)\big| \Big|\frac{\partial}{\partial s}G(t_l^+, s, b_n, h_d)\Big|.$$

The terms in this sum can be estimated by an application of Lemma 8.1, that is

$$\sup_{\frac{j-1}{n}<s\le\frac{j}{n}} \big|G(t_l^+, s, b_n, h_d)\big| \le C\lambda\left(\mathcal{D}_{lj}\right), \tag{7.24}$$

$$\sup_{\frac{j-1}{n}<s\le\frac{j}{n}} \Big|\frac{\partial}{\partial s}G_j(t_l^+, s, b, h_d)\Big| \le C\lambda\left(\mathcal{D}_{lj}\right)/b_n, \tag{7.25}$$

where

$$\mathcal{D}_{lj} = \Big(\frac{j-1}{n} - b_n, \frac{j+1}{n} + b_n\Big) \cap \Big(t_l^+ - Mh_d^{\frac{1}{v_l^++1}}, t_l^+ + Mh_d^{\frac{1}{v_l^++1}}\Big),$$

M and C are sufficiently large constants and $\lambda(\cdot)$ denotes the Lebesgue measure. Straightforward calculations show that the number of indices j such that the set $\mathcal{D}_{lj}$ is not empty is of order $O(nh_d^{\frac{1}{v_l^++1}})$, while the Lebesgue measure in (7.24) and of (7.25) is of order $O(b_n)$

and $O(1)$, respectively. Combining these facts we obtain

$$\alpha_{n,l} = \alpha'_{n,l} + O\Big(nh_d^{\frac{1}{v_l^+ + 1}} b_n \frac{1}{n^3 b_n^2 h_d^2}\Big) \tag{7.26}$$

(for all $l = 1, \ldots, k^+$ such that $c_l < \infty$). As the function σ is strictly positive on a compact set it follows that

$$\alpha'_{n,l} = \alpha''_{n,l}\Big(1 + O\Big(b_n + h_d^{\frac{1}{v_l^+ + 1}}\Big)\Big), \tag{7.27}$$

where the quantity $\alpha''_{n,l}$ is defined as $\alpha'_{n,l}$ in (7.23) replacing the $\sigma(s)$ by $\sigma(t_l^+)$. Define

$$\alpha'''_{n,l} = \frac{\sigma^2(t_l^+)}{nb_n^2 h_d^2} \int_0^1 \Big(\int_{t_l^+ - \epsilon}^{t_l^+ + \epsilon} K^*\Big(\frac{t-s}{b_n}\Big) K_d\Big(\frac{\mu^{(v_l^+ + 1)}(t_l^+)(t - t_l^+)^{v_l^+ + 1}}{(v_l^+ + 1)! h_d}\Big) dt\Big)^2 ds \tag{7.28}$$

and note that the only difference between $\alpha''_{n,l}$ and $\alpha'''_{n,l}$ is the term inside $K_d(\cdot)$. A Taylor expansion around t_l^+ yields

$$\frac{\mu(t) - \mu(0) - c}{h_d} = \frac{\mu^{(v_l^+ + 1)}(t_l^*)(t - t_l^+)^{v_l^+ + 1}}{(v_l^+ + 1)! h_d}$$

for some $t_l^* \in [t_l \wedge t_l^*, t_l \vee t_l^*]$ and the mean value theorem gives

$$\begin{aligned} &K_d\Big(\frac{\mu(t) - \mu(c) - c}{h_d}\Big) - K_d\Big(\frac{\mu^{(v_l^+ + 1)}(t_l^+)(t - t_l^+)^{v_l^+ + 1}}{(v_l^+ + 1)! h_d}\Big) \\ &= K_d'\Big(\frac{((1-\theta_l)\mu^{(v_l^+ + 1)}(t_l) + \theta_l \mu^{(v_l^+ + 1)}(t_l^*))(t - t_l^+)^{v_l^+ + 1}}{(v_l^+ + 1)! h_d}\Big) \frac{(\mu^{(v_l^+ + 1)}(t_l^*) - \mu^{(v_l^+ + 1)}(t_l))(t - t_l^+)^{v_l^+ + 1}}{(v_l^+ + 1)! h_d} \end{aligned}$$

for some $\theta_l \in [-1, 1]$. Then similar arguments as used in the derivation of (7.26) show that

$$\alpha'''_{n,l} - \alpha''_{n,l} = O\Big(h_d^{-\frac{2v_l^+}{v_l^+ + 1}} n^{-1}\Big). \tag{7.29}$$

On the other hand, further expanding the squared term of (7.28) yields that

$$\alpha'''_{n,l} = \frac{\sigma^2(t_l^+)}{nb_n^2h_d^2}\int_0^1\int_{t_l^+-\epsilon}^{t_l^++\epsilon}\int_{t_l^+-\epsilon}^{t_l^++\epsilon} K^*\Big(\frac{t-s}{b_n}\Big)K_d\Big(\frac{\mu^{(v_l^++1)}(t_l^+)(t-t_l^+)^{v_l^++1}}{(v_l^++1)!h_d}\Big)$$
$$\times K^*\Big(\frac{v-s}{b_n}\Big)K_d\Big(\frac{\mu^{(v_l^++1)}(t_l^+)(v-t_l^+)^{v_l^++1}}{(v_l^++1)!h_d}\Big)dvdtds.$$

For t, v satisfying $|t-t_l^+| = O(\min\{b_n, h_d^{\frac{1}{v_l^++1}}\})$, $|v-t_l| = O(\min\{b_n, h_d^{\frac{1}{v_l^++1}}\})$, straightforward calculations show

$$\int_0^1 K^*\Big(\frac{t-s}{b_n}\Big)K^*\Big(\frac{v-s}{b_n}\Big)ds = b_n\int_{-\infty}^{\infty} K^*\Big(u\Big)K^*\Big(\frac{v-t}{b_n}+u\Big)du. \tag{7.30}$$

To move forward, we introduce the notation

$$z_1 = (t-t_l^+)\Big|\frac{\mu^{(v_l^++1)}(t_l^+)}{h_d(v_l^++1)!}\Big|^{\frac{1}{v_l^++1}},\ z_2 = (v-t_l^+)\Big|\frac{\mu^{(v_l^++1)}(t_l^+)}{h_d(v_l^++1)!}\Big|^{\frac{1}{v_l^++1}},\ \theta(v_l^+, h_d) = \Big|\frac{h_d(v_l^++1)!}{\mu^{(v_l^++1)}(t_l^+)}\Big|^{\frac{1}{v_l^++1}}.$$

By a change of variables and (7.30), we now obtain

$$\begin{aligned}\alpha'''_{n,l} =& \frac{\sigma^2(t_l^+)\theta(v_l^+,h_d)^2}{nb_nh_d^2}\int\int\int K^*(u)K^*\Big(u+\frac{1}{b_n}\theta(v_l^+,h_d)(z_2-z_1)\Big)\\ &\times K_d(z_1^{v_l^++1})K_d(z_2^{v_l^++1})dz_1dz_2du\\ =&\frac{\sigma^2(t_l^+)}{nh_d^2}h_d^{\frac{1}{v_l^++1}}\Big|\frac{(v_l^++1)!}{\mu^{(v_l^++1)}(t_l^+)}\Big|^{\frac{1}{v_l^++1}}\int\int\int K^*(u)K^*(v)K_d(z_1^{v_l^++1})\\ &\times K_d\Big(\Big(z_1+c_l\Big|\frac{(v_l^++1)!}{\mu^{(v_l^++1)}(t_l^+)}\Big|^{\frac{-1}{v_l^++1}}(v-u)\Big)^{v_l^++1}\Big)dudvdz_1\end{aligned}$$

Finally, combining (7.26), (7.27) and (7.29) we have that

$$\alpha_{n,l} = \alpha'''_{n,l}\Big(1+b_n+\frac{1}{nb_n}+h_d^{\frac{1}{v_l^++1}}\Big)\ ,$$

and, observing that $h_d^{\frac{1}{v_l^+ +1}} = o(h_d^{\frac{1}{v^+ +1}})$ whenever $v_l^+ < v^+$, we obtain from (7.22)

$$\alpha_n = \frac{|h_d(v^+ +1)!|^{\frac{1}{v^+ +1}}}{nh_d^2} \sum_{\{l: v_l^+ = v^+\}} \frac{\sigma^2(t_l^+)}{|\mu^{(v^+ +1)}(t_l^+)|^{\frac{1}{v^+ +1}}} \int\int\int K^*(u)K^*(v)K_d(z_1^{v^+ +1})$$
$$\times K_d\Big(\Big(z_1 + r\Big|\frac{(v^+ +1)!}{\mu^{(v^+ +1)}(t_l^+)}\Big|^{\frac{-1}{v^+ +1}}(v-u)\Big)^{v^+ +1}\Big) dudvdz_1(1+o(1)) \ ,$$

which proves (7.17) in the case $b_n^{v^+ +1}/h_d \to r^{v^+ +1} \in [0,\infty)$.

Proof of (7.18). Recalling the definition of $\tilde{\alpha}_n$ in (7.16) we obtain by straightforward calculations and a Taylor expansion

$$\tilde{\alpha}_n = \frac{\sigma^2(0)}{nb_nh_d^2} \int_0^1 (\bar{K}^*(t))^2 dt \Big(\int_0^1 K_d\Big(\frac{\mu(t)-\mu(0)-c}{h_d}\Big)dt\Big)^2 \Big(1 + O\Big(b_n + \frac{1}{nb_n}\Big)\Big).$$

Similar (but easier) arguments as used in the derivation of (7.20) and (7.21) show

$$\int_0^1 K_d\Big(\frac{\mu(t)-\mu(0)-c}{h_d}\Big)dt = |h_d(v^+ +1)!|^{\frac{1}{v^+ +1}}$$
$$\times \sum_{\{l: v_l^+ = v^+\}} |\mu^{(v^+ +1)}(t_l^+)|^{-\frac{1}{v^+ +1}} \int K_d(z^{v^+ +1})dz(1+o(1)),$$

which gives

$$\tilde{\alpha}_n = \frac{\sigma^2(0)h_d^{\frac{-2v^+}{v^+ +1}}((v^+ +1)!)^{\frac{2}{v^+ +1}}}{nb_n} \int_0^1 (\bar{K}^*(t))^2 dt$$
$$\times \Big(\sum_{\{l: v_l^+ = v^+\}} |\mu^{(v^+ +1)}(t_l^+)|^{-\frac{1}{v^+ +1}} \int K_d(z^{v^+ +1})dz\Big)^2 (1+o(1)).$$

Consequently, if $b_n^{v^+ +1}/h_d \to \infty$ we have

$$\tilde{\alpha}_n = \frac{h_d^{\frac{-2v^+}{v^+ +1}}}{nb_n} \sigma_2^{2,+}(1+o(1)).$$

where $\sigma_2^{2,+}$ is defined by (4.7). This proves the statement (7.18) in the case $b_n^{v^+ +1}/h_d \to \infty$, while the second case follows by similar arguments observing that we have

$$nh_d^{\frac{v^+}{v^++1}+1}\frac{h_d^{-\frac{2v^+}{v^++1}}}{nb_n} = r^{-1}$$

if $b_n^{v^++1}/h_d \to r^{v^++1} \in [0,\infty)$.

Proof of Theorem 5.1. We have to distinguish two cases:

(1) The equation $\mu(t)-\mu(0)=c$ has at least one solution. Recall the definition of the quantity I' in (5.1), then it follows from the proof of Theorem 4.1, that

$$\mathrm{Var}(\sqrt{nb_n}h_d^{\frac{v^+}{v^++1}}I') = \sigma_1^{2,+}+\sigma_2^{2,+}+o(1),$$

where $\sigma_1^{2,+}$ and $\sigma_2^{2,+}$ are defined in (4.6) and (4.7), respectively. Note that $\mathrm{Var}(I') = \frac{1}{n^2N^2b_n^2h_d^2}\tilde{V}$, where

$$\tilde{V} = \sum_{j=1}^{n}\sigma^2(j/n)\Big(\sum_{i=1}^{N}K_d\Big(\frac{\mu(i/N)-\mu(0)-c}{h_d}\Big)\Big(K^*\Big(\frac{i/N-j/n}{b_n}\Big)-\bar{K}^*\Big(\frac{j}{nb_n}\Big)\Big)\Big)^2.$$

At the end of this proof we will show that

$$\frac{(\sqrt{nb_n}h_d^{\frac{v^+}{v^++1}})^2}{n^2N^2b_n^2h_d^2}(\tilde{V}-\bar{V}) = o(1), \tag{7.31}$$

which implies that

$$\lim_{n\to\infty}\sqrt{nb_n}h_d^{\frac{v^+}{v^++1}}q_{1-\alpha}^+/(nNb_nh_d) = \Phi^{-1}(1-\alpha)\sqrt{\sigma_1^{2,+}+\sigma_2^{2,+}}. \tag{7.32}$$

Observing the identity

$$\begin{aligned}&\mathbb{P}\big(nNb_nh_d\big(\tilde{T}_{N,c}^+-\Delta\big)>q_{1-\alpha}^+\big) \qquad (7.33)\\ &=\mathbb{P}\left(\frac{\sqrt{nb_n}h_d^{\frac{v^+}{v^++1}}\big(\tilde{T}_{N,c}^+-T_c^+\big)}{\sqrt{\sigma_1^{2,+}+\sigma_2^{2,+}}} > \frac{\frac{\sqrt{nb_n}h_d^{\frac{v^+}{v^++1}}}{nNb_nh_d}q_{1-\alpha}^+ + \sqrt{nb_n}h_d^{\frac{v^+}{v^++1}}(\Delta-T_c^+)}{\sqrt{\sigma_1^{2,+}+\sigma_2^{2,+}}}\right)\end{aligned}$$

the assertion now follows from (7.32) and Theorem 4.1, which shows that the random variable

$$\frac{\sqrt{nb_n} h_d^{\frac{v^+}{v^++1}} \big(\tilde{T}^+_{N,c} - T^+_c\big)}{\sqrt{\sigma_1^{2,+} + \sigma_2^{2,+}}}$$

converges weakly to a standard normal distribution.

It remains to prove (7.31), which is a consequence of the following observations

(a) $\hat{\sigma}(t_l^+) = \sigma(t_l^+)(1 + o(1))$, uniformly with respect to $l = 1, \ldots, k^+$.

(b) The bandwidth condition $\pi_n/h_d = o(1)$, Proposition 2.1 and similar arguments as (7.5) show

$$K_d\Big(\tfrac{\mu(\frac{i}{N})-\mu(0)-c}{h_d}\Big) - K_d\Big(\tfrac{\tilde{\mu}_{b_n}(\frac{i}{N})-\tilde{\mu}_{b_n}(0)-c}{h_d}\Big) = O\Big(\sum_{\{l: v_l^+ = v^+\}} \mathbf{1}\big(\big|\tfrac{i}{N} - t_l^+\big| \le h_d^{\frac{1}{v^++1}}\big) \tfrac{\pi_n}{h_d}\Big),$$

where π_n is defined in Theorem 4.1.

This completes the proof of Theorem 5.1 in the case that there exist in fact roots of the equation $\mu(t) - \mu(0) = c$.

(2) The equation $\mu(t)-\mu(0) = c$ has no solutions. In this case we have $\mu(t)-\mu(0) < c$ where $c > 0$. Note that for two sequences of measurable sets U_n and V_n such that $\mathbb{P}(U_n) \to 1$ and $\mathbb{P}(U_n \cap V_n) \to u \in (0, 1)$, we have $\mathbb{P}(V_n) \to u$. Consequently, as the set A_n defined in (7.4) satisfies $\mathbb{P}(A_n) \to 1$ the assertion of the theorem follows from

$$\lim_{n\to\infty} \mathbb{P}(nNb_nh_d(\tilde{T}^+_{N,c} - \Delta) > q^+_{1-\alpha}, A_n, \mu(t) - \mu(0) < c) = 0. \tag{7.34}$$

However, under the event A_n and $\mu(t) - \mu(0) < c$ we have $q^+_{1-\alpha} = 0$ and $\tilde{T}^+_{N,c} = 0$, if n is sufficiently large. Thus (7.34) is obvious (note that $0 < \Delta < 1$), which finishes the proof in the case where the equation $\mu(t) - \mu(0) = c$ has in fact no roots. □

Proof of Theorem 5.2 Define $\tilde{S}_{k,r} = \sum_{i=k\vee 1}^{r\wedge n} e_i$,

$$\tilde{\Delta}_j = \frac{\tilde{S}_{j-m+1,j} - \tilde{S}_{j+1,j+m}}{m}\ , \quad \tilde{\sigma}^2(t) = \sum_{j=1}^{n} \frac{m\tilde{\Delta}_j^2}{2} w(t, j).$$

Since $\mu(\cdot) \in \mathcal{C}^2$, elementary calculations show that uniformly for $t \in [0,1]$,

$$|\tilde{\sigma}^2(t) - \hat{\sigma}^2(t)| = O_p(m^{5/2}/n) \tag{7.35}$$

Similar arguments as given in the proof of Lemma 3 of Zhou and Wu (2010) yields $\sup_j \|\tilde{\Delta}_j\|_4 = O(m^{-1/2})$. A further application of Lemma 3 of Zhou and Wu (2010) gives

$$\| \sup_{t \in [\gamma_n, 1-\gamma_n]} |\tilde{\sigma}^2(t) - \mathbb{E}(\tilde{\sigma}^2(t))| \|_2 = O(m^{1/2} n^{-1/2} \tau_n^{-1}), \tag{7.36}$$
$$\|\tilde{\sigma}^2(t) - \mathbb{E}(\tilde{\sigma}^2(t))\|_2 = O(m^{1/2} n^{-1/2} \tau_n^{-1/2})$$

Elementary calculations show that

$$\mathbb{E}(\tilde{\sigma}^2(t)) = \Lambda_1(t) + \Lambda_2(t) + \Lambda_3(t), \tag{7.37}$$

where

$$\Lambda_1(t) = \frac{1}{2m} \sum_{j=1}^{n} \tilde{S}^2_{j-m-1,j} \omega(t,j)$$
$$\Lambda_2(t) = \frac{1}{2m} \sum_{j=1}^{n} \tilde{S}^2_{j+1,j+m} \omega(t,j)$$
$$\Lambda_3(t) = -\frac{1}{2m} \sum_{j=1}^{n} \tilde{S}_{j+1,j+m} \tilde{S}_{j-m-1,j} \omega(t,j)$$

Recall the representation $e_i = G(i/n, \mathcal{F}_i)$. Define $\tilde{S}^{\diamond}_{j-m+1,j} = \sum_{r=1\vee(j-m+1)}^{j} G(j/n, \mathcal{F}_r)$, and $\tilde{S}^{\diamond}_{j-m+1} = \sum_{r=j+1}^{n\wedge(j+m)} G(j/n, \mathcal{F}_r)$. For $s = 1,2,3$, define $\Lambda^{\diamond}_s(t)$ as the quantity where the terms $\tilde{S}_{j-m+1,j}$ and $\tilde{S}_{j-m+1}$ in $\Lambda_s(t)$ are replaced by $\tilde{S}^{\diamond}_{j-m+1,j}$, $\tilde{S}^{\diamond}_{j-m+1}$, respectively.

Then by Lemma 4 of Zhou and Wu (2010), we have uniformly with resepct to $t \in [0,1]$,

$$|\mathbb{E}(\Lambda^{\diamond}_s(t)) - \mathbb{E}(\Lambda_s(t))| = O(\sqrt{m/n}), \quad s = 1,2,3.$$

By Lemma 5 of Zhou and Wu (2010), it follows for $s = 1, 2$,

$$|\mathbb{E}(\Lambda_s^\diamond(t)) - \sigma^2(t)/2| = O(m^{-1} + \tau_n^2), t \in [\gamma_n, 1-\gamma_n],$$
$$|\mathbb{E}(\Lambda_s^\diamond(t)) - \sigma^2(t)/2| = O(m^{-1} + \tau_n), t \in [0, \gamma_n) \cup (1-\gamma_n, 1].$$

Define $\Gamma(k) = \mathbb{E}(G(i/n, \mathcal{F}_0)G(i/n, \mathcal{F}_k))$, then similar arguments as given in the proof of Lemma 5 of Zhou and Wu (2010) yield $\Gamma(k) = O(\chi^{|k|})$. Elementary calculations show that for $1 \le j \le n$

$$\mathbb{E}(S_{j-m+1,j}^\diamond S_{j+1,j+m}^\diamond) = \sum_{k=1}^{m} \Gamma(k) = O(1),$$

which proves

$$\mathbb{E}(\Lambda_3^\diamond(t)) = O(m^{-1}) \tag{7.38}$$

uniformly with respect to $t \in [0, 1]$. From (7.37)–(7.38) it follows that

$$\sup_{t \in [\gamma_n, 1-\gamma_n]} |\mathbb{E}\tilde{\sigma}^2(t) - \sigma^2(t)| = O(\sqrt{m/n} + m^{-1} + \tau_n^2),$$
$$\sup_{t \in [0,\gamma_n) \cup (1-\gamma_n, 1]} |\mathbb{E}\tilde{\sigma}^2(t) - \sigma^2(t)| = O(\sqrt{m/n} + m^{-1} + \tau_n).$$

The theorem is now a consequence of these two equations and (7.36)–(7.37). □

8 Some technical results

8.1 The size of mass excess

Lemma 8.1. *Assume that the function $\mu(\cdot) - m$ has k roots $0 < t_1 < \ldots < t_k < 1$ of order v_i, $1 \le i \le k$, and define $\gamma = \frac{1}{2}\min_{0 \le i \le k}(t_{i+1} - t_i)$ (with convention that $t_0 = 0, t_{k+1} = 1$), such that*

(i) For $1 \le s \le k$, the (v_s+1)nd derivative of $\mu(\cdot)$ exists on the interval $\mathcal{I}_s := (t_s - \gamma, t_s + \gamma)$, and is Lipschitz continuous on $[t_s - \gamma, t_s + \gamma]$.

(ii) $\mu(\cdot)$ *is strictly monotone on the intervals* $\mathcal{I}_s^-$ *and* $\mathcal{I}_s^+$ *for* $1 \le s \le k$*, where* $\mathcal{I}_s^- := (t_s - \gamma, t_s]$, $\mathcal{I}_s^+ := (t_s, t_s + \gamma)$*,*

(iii) there exists a positive number ϵ*, such that* $\min_{t \in [0,1] \cap_{s=1}^k \bar{\mathcal{I}}_s} |\mu(t) - m| \ge \epsilon$*, where* $\bar{\mathcal{I}}_s := [0, t_s - \gamma] \cup [t_s + \gamma, 1]$ *is complement of* $\mathcal{I}_s$ *.*

If A_n *denotes the set*

$$A_n := \{s : |\mu(s) - m| \le h_n\}, \tag{8.1}$$

then there exists a sufficiently large constant C *such that for any sequence* $h_n \to 0$*, we have*

$$\lambda(A_n) \le C h_n^{\frac{1}{v+1}},$$

where $v = \max_{1 \le l \le k} v_l$*. Furthermore, there exists a sufficiently large constant* M*, such that*

$$A_n = \cup_{l=1}^k B_{n,l,M} \tag{8.2}$$

when n *is sufficiently large, where the sets* $B_{n,l,M}$ *are defined by*

$$B_{n,l,M} = \{s : |s - t_l|^{v_l+1} \le M h_n, |\mu(s) - m| \le h_n\}.$$

Proof. Define for $1 \le l \le k$,

$$A_{n,l} = \{s : |\mu(s) - m| \le h_n, |s - t_l| < \min\{\gamma, \zeta_n\}\}, \tag{8.3}$$

where ζ_n is a sequence of real numbers which converges to zero arbitrarily slowly. We shall show that there exists a constant $n_0 \in \mathbb{N}$, such that for $n \ge n_0$

$$A_n = \cup_{l=1}^k A_{n,l}, \tag{8.4}$$

$$A_{n,l} \subseteq B_{n,l,M}, \quad 1 \le l \le k, \tag{8.5}$$

where M is a sufficiently large constant. Note that (8.4) and (8.5) yield $A_n \subseteq \cup_{l=1}^k B_{n,l,M}$. By definition of $B_{n,l,M}$ and A_n, we have that $\cup_{l=1}^k B_{n,l,M} \subseteq A_n$, which proves (8.2). Then a straightforward calculation shows that

$$\lambda(B_{n,l,M}) \le C h_n^{\frac{1}{v_l+1}} \le C h_n^{\frac{1}{v+1}},$$

and the lemma follows.

We first prove the assertion (8.4). By definition, $A_n \supseteq \cup_{l=1}^k A_{n,l}$. We now argue that there exists a sufficiently large constant n_0, such that for $n \ge n_0$, $\cup_{l=1}^k A_{n,l} \supseteq A_n$.

Suppose this statement is not true, then there exists a sequence of points $(s_n)_{n\in\mathbb{N}}$, such that $s_n \in A_n$ and $s_n \in \cap_{l=1}^k \bar{A}_{n,l}$, where $\bar{A}_{n,l}$ is the complement set of $A_{n,l}$. Since $h_n = o(1)$ we have $h_n < \epsilon$ for sufficiently large n and by assumption (iii), there exists an $l \in \{1, \ldots, k\}$ such that

$$s_n \in \mathcal{I}_l \cap A_n \cap \bar{A}_{n,l}.$$

Without loss of generality we assume that $s_n \in \mathcal{I}_l^+ \cap A_n \cap \bar{A}_{n,l}$. The case that $s_n \in \mathcal{I}_l^- \cap A_n \cap \bar{A}_{n,l}$ can be treated similarly.

A Taylor expansion and assumption (i) yield for sufficiently large $n \in \mathbb{N}$

$$\mu(s) - \mu(t_l) = \frac{\mu^{(v_l+1)}(t_l)}{(v_l+1)!}(s-t_l)^{v_l+1} + \frac{\mu^{(v_l+1)}(t_l^*) - \mu^{(v_l+1)}(t_l)}{(v_l+1)!}(s-t_l)^{v_l+1} \qquad (8.6)$$

for $s \in A_{n,l}$, where $t_l^* \in [t_l \wedge s, t_l \vee s]$. By the definition of $A_{n,l}$ in (8.3) and the fact that $\zeta_n = o(1)$, we have that $A_{n,l} \subset \mathcal{I}_l$ for sufficiently large $n \in \mathbb{N}$. This result together with $s_n \in \mathcal{I}_l^+ \cap A_n \cap \bar{A}_{n,l}$ implies that $t_l + \zeta_n < s_n \le t_l + \gamma$. However, by assumption (ii), $\mu(\cdot)$ is strictly monotone in $\mathcal{I}_l^+$, which yields that for sufficiently large n,

$$|\mu(s_n) - \mu(t_l)| \ge |\mu(t_l + \zeta_n) - \mu(t_l)| > 2h_n, \qquad (8.7)$$

where the last $>$ is due to (8.6), the Lipschitz continuity of $\mu^{(v_l+1)}(\cdot)$ in the neighbourhood of t_l and the fact that $\zeta_n \to 0$ arbitrarily slowly. By the definition of A_n in (8.1), equation (8.7) implies that $s_n \notin A_n$. This contradicts to the assumption that $s_n \in A_n$, from which (8.4) follows.

Now we show the conclusion (8.5). Since $\mu(t_l) = m$ and the leading term in (8.6) is of order $|(s-t_l)^{v_l+1}|$, the set $A_{n,l}$ can be represented as

$$\Big\{s : |s-t_l| \leq \Big(\frac{h_n}{|M_{1,l}+M_{2,l}(s)}|\Big)^{\frac{1}{v_l+1}}, |s-t_l| \leq \zeta_n, |\mu(s)-c| \leq h_n\Big\},$$

where $M_{1,l} = \frac{\mu^{(v_l+1)}(t_l)}{(v_l+1)!}$, and $M_{2,l}(s) = \frac{\mu^{(v_l+1)}(t_l^*)-\mu^{(v_l+1)}(t_l)}{(v_l+1)!}(s-t_l)^{v_l+1}$ for some $t_l^* \in [t_l \wedge s, t_l \vee s]$. By the Lipschitz continuity of $\mu^{(v_l+1)}(\cdot)$ on the interval $[t_l-\gamma, t_l+\gamma]$, there exists a constant M_l' such that $|M_{2,l}(s)| \leq M_l'|t_l - s|$. As $\zeta_n = o(1)$ there exists an $n_l \in \mathbb{N}$ such that $|s-t_l| \leq \frac{|M_{1,l}|}{2M_l'}$ for all $s \in A_{n,l}$ whenever $n \geq n_l$. This yields

$$|M_{1,l} + M_{2,l}(s-t_l)| \geq \frac{|M_{1,l}|}{2}$$

for all $n \geq n_l$, $s \in A_{n,l}$. By choosing $n_0 = \max_{1\leq l\leq k} n_l$ and $M = \max_{1\leq l\leq k}\Big(\frac{2}{|M_{1,l}|}\Big)^{\frac{1}{v_l+1}}$, and noticing the fact that $\zeta_n \to 0$ arbitrarily slow, it follows that

$$A_{n,l} \subseteq B_{n,l,M}$$

for $n \geq n_0$. Thus (8.5) follows, which completes the proof of Lemma 8.1. □

Remark 8.1. Observe that $B_{n,i,M} \cap B_{n,j,M} = \emptyset$ for $i \neq j$ if n is sufficiently large. Moreover, $B_{n,i,M}$ can be covered by closed intervals. The Lemma shows that the set $\{t : |\mu(t)-m| \leq h_n, t \in [0,1]\}$ can be decomposed in disjoint intervals containing the root of the equation $\mu(t) = m$, with Lebesgue measure determined by the maximal critical order of the roots.

8.2 Uniform bounds for nonparametric estimates

In this section we present some results about the rate of uniform convergence of the Jackknife estimator $\tilde{\mu}_{b_n}(t)$ defined in (4.1).

Lemma 8.2. *Recall the definition of $\tilde{\mu}_{b_n}$ in* (4.1) *and suppose that Assumption 2.1(a)* *holds. If $b_n \to 0$, $nb_n \to \infty$, then*

$$\sup_{t\in[b_n,1-b_n]}\Big|\tilde\mu_{b_n}(t)-\mu(t)-\frac{1}{nb_n}\sum_{i=1}^n K^*\Big(\frac{i/n-t}{b_n}\Big)e_i\Big| = O(b_n^3+\frac{1}{nb_n}), \tag{8.8}$$

$$\Big|\tilde\mu_{b_n}(0)-\mu(0)-\frac{1}{nb_n}\sum_{i=1}^n \bar K^*\Big(\frac{i/n}{b_n}\Big)e_i\Big| = O(b_n^3+\frac{1}{nb_n}),$$

where $K^*(\cdot)$ and $\bar K^*(\cdot)$ are defined in (4.2) and (4.3), respectively

Proof. We only show the estimate (8.8). The other result follows similarly using Lemma B.2 of Dette et al. (2015). By Lemma B.1 of Dette et al. (2015) we obtain a uniform bound for the (uncorrected) local linear estimate $\hat\mu_{b_n}$ in (2.1), that is

$$\sup_{t\in[b_n,1-b_n]}\Big|\hat\mu_{b_n}(t)-\mu(t)-\frac{\mu_2\ddot\mu(t)}{2}b_n^2-\frac{1}{nb_n}\sum_{i=1}^n e_iK_{b_n}(i/n-t)\Big| = O(b_n^3+\frac{1}{nb_n}).$$

Then the lemma follows from the definition of $\tilde\mu_{b_n}(\cdot)$. □

Lemma 8.3. *If Assumption 2.1(a), Assumption 2.2 are satisfied and $\frac{nb_n^2}{\log^4 n}\to\infty$, $b_n\to 0$, then*

$$\sup_{t\in\{0\}\cup[b_n,1-b_n]}|\tilde\mu_{b_n}(t)-\mu(t)| = O_p\Big(b_n^3+\frac{\log n}{\sqrt{nb_n}}\Big). \tag{8.9}$$

$$\sup_{t\in[0,b_n)\cup(1-b_n,1]}|\tilde\mu_{b_n}(t)-\mu(t)| = O_p\Big(b_n^2+\frac{\log n}{\sqrt{nb_n}}\Big). \tag{8.10}$$

Proof. We only prove the estimate

$$\sup_{[b_n,1-b_n]}|\tilde\mu_{b_n}(t)-\mu(t)| = O_p\Big(b_n^3+\frac{\log n}{\sqrt{nb_n}}\Big).$$

The case that $t=0$ in (8.9) and the estimate (8.10) follow by similar arguments, which are omitted for the sake of brevity. By the stochastic expansion (8.8), it suffices to show that

$$\sup_{t\in[b_n,1-b_n]}\Big|\frac{1}{nb_n}\sum_{i=1}^n K^*\Big(\frac{i/n-t}{b_n}\Big)e_i\Big| = O_p\Big(\frac{\log n}{\sqrt{nb_n}}\Big).$$

Then Assumption 2.2, Proposition 5 of Zhou (2013) and the summation by parts formula (44) in Zhou (2010) yield the existence (on a possibly richer probability space) of a sequence $(V_i)_{i\in\mathbb{Z}}$ of independently standard normal distributed random variables such that

$$\sup_{t\in[b_n,1-b_n]}\Big|\frac{1}{nb_n}\sum_{i=1}^{n}K^*\Big(\frac{i/n-t}{b_n}\Big)(e_i-V_i)\Big|=O_p\Big(\frac{n^{1/4}\log^2 n}{nb_n}\Big).$$

Note that $(V_i)_{i\in\mathbb{Z}}$ is a martingale difference sequence with respect to the filtration generated by $(V_{-\infty},...,V_i)$. By Burkholder's inequality it follows that for any positive κ and a sufficiently large universal constant C the inequality

$$\Big\|\sum_{i=1}^{n}V_iK^*_{b_n}(i/n-t)\Big\|_\kappa^2\leq C\kappa\Big\|\Big(\sum_{i=1}^{n}\{V_iK^*_{b_n}(i/n-t)\}^2\Big)^{1/2}\Big\|_\kappa^2$$
$$\leq C\kappa\sum_{i=1}^{n}\big\|(V_iK^*_{b_n}(i/n-t))^2\big\|_{\frac{\kappa}{2}}=C\kappa\sum_{i=1}^{n}\big\|(V_iK^*_{b_n}(i/n-t))\big\|_\kappa^2\leq C\kappa^2(nb_n)$$

holds uniformly with respect to $t\in[b_n,1-b_n]$, where we have used that $\mathbb{E}|V_0|^\kappa\leq(\kappa-1)!!\leq\kappa^{\frac{\kappa}{2}}$ in the last inequality. This leads to

$$\sup_{t\in[b_n,1-b_n]}\Big\|\frac{1}{nb_n}\sum_{i=1}^{n}K^*_{b_n}(i/n-t)V_i\Big\|_\kappa=O\Big(\frac{\kappa}{\sqrt{nb_n}}\Big).$$

Similarly, we obtain

$$\sup_{t\in[b_n,1-b_n]}\Big\|\frac{1}{nb_n}\sum_{i=1}^{n}\frac{\partial}{\partial t}K^*_{b_n}(i/n-t)V_i\Big\|_\kappa=O\Big(\frac{\kappa b_n^{-1}}{\sqrt{nb_n}}\Big).$$

Consequently, Proposition B.1. of Dette et al. (2015) shows that

$$\Big\|\sup_{t\in[b_n,1-b_n]}\frac{1}{nb_n}\sum_{i=1}^{n}K^*_{b_n}(i/n-t)V_i\Big\|_\kappa=O\Big(\frac{\kappa b_n^{-\frac{1}{\kappa}}}{\sqrt{nb_n}}\Big)$$

The result now follows using $\kappa=\log(b_n^{-1})$ observing the conditions on the bandwidths. □