The Annals of Statistics
2019, Vol. 47, No. 6, 3578–3608
https://doi.org/10.1214/19-AOS1811

DETECTING RELEVANT CHANGES IN THE MEAN OF NONSTATIONARY PROCESSES—A MASS EXCESS APPROACH[1]

BY HOLGER DETTE* AND WEICHI WU†,*

Ruhr-Universität Bochum and Tsinghua University*†

This paper considers the problem of testing if a sequence of means $(\mu_t)_{t=1,\dots,n}$ of a nonstationary time series $(X_t)_{t=1,\dots,n}$ 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 threshold, that is $|\mu_1 - \mu_t| \le c$ for all $t = 1, \dots, n$. 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 of the 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 nonstationary sequence from its initial value. The results are illustrated by means of a simulation study and by analyzing data examples.

1. Introduction. A frequent problem in time series analysis is the detection of structural breaks. Since the pioneering work of [32] 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 [3, 6, 9, 16, 26] and [4] and refer to the work of [5] and [25] 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 nonstationary 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 null hypothesis is then given by

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

Received February 2018; revised January 2019.

[1]Supported in part by the Deutsche Forschungsgemeinschaft (SFB 823: Statistik nichtlinearer dynamischer Prozesse, Teilprojekt A1 and C1, FOR 1735: Structural inference in statistics—adaptation and efficiency).

MSC2010 subject classifications. 62M10, 62F05, 62G08, 62G09.

Key words and phrases. Locally stationary process, change point analysis, relevant change points, local linear estimation, Gaussian approximation, rearrangement estimators.

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 nonstationarity. For example, as pointed out by [21], 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 [21] 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, e.g., [43]) but has also been used to investigate the similarity of distribution functions (see [1, 2] among others). In the context of detecting a change in a sequence of means (or other parameters of the marginal distribution) [21] assumed two stationary phases and tested if the difference before and after the change point is small, that is

$$H_0 : |\mu_{(1)} - \mu_{(2)}| \leq c \quad \text{versus} \quad H_1 : |\mu_{(1)} - \mu_{(2)}| > c, \tag{1.2}$$

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.

A similar idea can be used to specify the economic design of control charts for quality control purposes. While in change-point analysis the focus is on testing for the presence of a change and on estimating the time at which a change occurs once it has been detected, control charting has typically been focused more on detecting such a change as quickly as possible after it occurs (see, e.g., [11, 44] among many others). In particular, control charts are related to sequential change point detection, while the focus of the cited literature is on retrospective change point detection.

In the present paper, we investigate alternative relevant hypotheses in the retrospective change point problem, which are motivated by the observation that

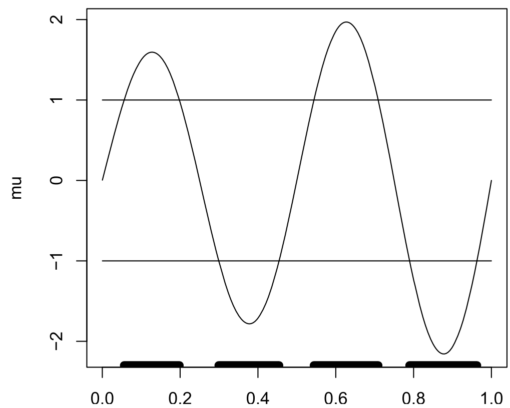

FIG. 1. *Illustration of the set $\mathcal{M}_c$ in* (1.4).

in many applications the assumption of two stationary phases (such as constant means before and after the change point) cannot be justified as the process parameters change continuously in time. For this purpose, we consider the location scale model

$$X_{i,n} = \mu(i/n) + \epsilon_{i,n}, \tag{1.3}$$

where $\{\epsilon_{i,n} : i = 1, \dots, n\}_{n\in\mathbb{N}}$ denotes a triangular array of centered random variables (note that we do not assume that the "rows" $\{\epsilon_{j,n} : j = 1, \dots, n\}$ 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 reasonably large. More precisely, for a level $c > 0$ we consider the *level set*

$$\mathcal{M}_c = \{t \in [0, 1] : |\mu(t) - \mu(0)| > c\} \tag{1.4}$$

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. The situation is illustrated in Figure 1, where the curve represents the mean function μ with $\mu(0) = 0$ and the lines in boldface represent the set $\mathcal{M}_c$ (with $c = 1$). These periods resemble in some sense popular run rules from the statistical quality control literature which signal if k of the last m standardized sample means fall in the interval (see, e.g., [11]). Define

$$T_c := \lambda(\mathcal{M}_c) \tag{1.5}$$

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

$$H_0 : T_c \le \Delta \quad \text{versus} \quad H_1 : T_c > \Delta. \tag{1.6}$$

We consider the change as *relevant*, if the Lebesgue measure $T_c = \lambda(\mathcal{M}_c)$ is larger than the threshold Δ. Note that this includes the case when a change (greater than c) occurs at some point $t_1 < 1 - \Delta$ and the mean level remains constant otherwise.

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

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

$$H_0^- : T_c^- = \lambda(\mathcal{M}_c^-) \leq \Delta \quad \text{versus} \quad H_1^- : T_c^- > \Delta. \tag{1.8}$$

The hypotheses (1.6), (1.7) and (1.8) require the specification of two parameters Δ and c and in a concrete application both parameters have to be defined after a careful discussion with the practitioners. In particular, they will be different in different fields of application. Another possibility is to investigate a relative deviation from the mean, that is: $\mu(t)$ deviates from $\mu(0)$ relative to $\mu(0)$ by at most $x\%$ (see Section 2.2.2 for a discussion of this measure).

Although the mean function in model (1.3) cannot be assumed to be monotone, we use a monotone rearrangement type estimator (see [19]) to estimate the quantities T_c, T_c^+, T_c^-, and propose to reject the null hypothesis (1.6), (1.7) (1.8) 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 (1.3) with a locally stationary error process, which have found considerable interest in the literature (see [18, 30, 31, 47] and [42] 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 nonstationary. 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 normally distributed. The main challenge in this asymptotic analysis is to quantify the order of an approximation of the quantity

$$\lambda(\{t \in [0, 1] \colon |\hat{\mu}(t) - \hat{\mu}(0)| > c\}), \tag{1.9}$$

where $\hat{\mu}$ is an appropriate estimate of the regression function. While estimates of the mean trend have been already studied under local stationarity in the literature (see, e.g., [46]), the analysis of the quantity (1.9) and its approximation requires a careful localization of the effect of the estimation error around the critical roots satisfying the equation $|\mu(t) - \mu(0)| = c$.

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 nonstationarity of the error process in

(1.3) 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 derivatives 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, e.g., [13, 29, 33, 34]). 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 [7, 10, 17] and [27] among many others). The concept of mass excess has additionally been used for discrimination between time series (see [12]), for the construction of monotone regression estimates [15, 19], quantile regression [14, 20], clustering [36] and for bandwidth selection in density estimation (see [37]), 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, [24] 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. [33] studied the asymptotic properties of an estimate of the mass excess functional of a cumulative distribution function F with density f and [41] observed that the minimax risk in the problem of estimating the level set of a density depends on its "regularity." More recently, [12] 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 (1.3) with a nonstationary 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 (1.6). 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 nonstationary 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.

We also mention the differences to the work of [28] and [39], which has its focus on the detection of intervals of homogeneity of the underlying process, while the present paper investigates the problem to detect significant deviations of an inhomogeneous process from its initial distribution (here specified by different values of the mean function).

The approach proposed in this paper is also related to the sojourn time of a (real valued) stochastic process, say $\{X(t)\}_{t\in[0,1]}$, which is defined as

$$S_c = \int_0^1 \mathbf{1}\{|X(t) - X(0)| > c\}\, dt$$

and has widely been studied in probability theory under specific distributional assumptions (see, e.g., [8, 40] among many). To be precise, let $X(t) = \mu(t) + \epsilon(t)$ for some centered process $\{\epsilon(t)\}_{t\in[0,1]}$, then compared to the quantity T_c defined in (1.5), which refers to expectation $\mu(t)$, the quantity S_c is a random variable. An alternative excess-type measure is now given by the expected sojourn time

$$e_c := \mathbb{E}(S_c), \tag{1.10}$$

and the corresponding null hypotheses can be formulated as

$$H_0 : e_c \le \Delta \quad \text{versus} \quad H_1 : e_c > \Delta.$$

A further quantity of interest was mentioned by a referee to us and is defined by the probability that the sojourn time exceeds the threshold Δ, that is,

$$p_{c,\Delta} := \mathbb{P}(S_c > \Delta). \tag{1.11}$$

This quantity cannot be directly used for testing, but can be considered as a measure of a relevant deviation for a sufficiently long time from the initial state $X(0)$.

The rest of paper is organized as follows. In Section 2, we motivate our approach, define an estimator of the quantity T_c, discuss alternative measures and give some basic assumptions of the nonstationary model (1.3). 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, in this case all roots are of the same order and contribute to the asymptotic variance of the limit distribution, which simplifies the statement of the results substantially. In this case, we also 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 jackknife approach. The bootstrap procedure is developed in the second part of Section 4. In Section 5, we illustrate its finite sample properties by means of a simulation study. Finally, some discussion on multivariate data is given in Section 6. In this section, we also propose estimators of the quantities (1.10) and (1.11). All proofs and technical details are deferred to an Suuplementary Material [22], which also contains an illustration of the method analyzing two data examples.

2. Estimation and basic assumptions.

2.1. *Relevant changes via a mass excess approach.* Recall the definition of the testing problems (1.6), (1.7), (1.8) and note that $T_c = T_c^+ + T_c^-$, where

$$T_c^+ = \int_0^1 \mathbf{1}\big(\mu(t) - \mu(0) > c\big)\,dt, \qquad T_c^- = \int_0^1 \mathbf{1}\big(\mu(t) - \mu(0) < -c\big)\,dt,$$

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, e.g., Theorems 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

$$\begin{aligned}&\big(\hat{\mu}_{b_n}(t), \hat{\dot{\mu}}_{b_n}(t)\big)^T\\ &\quad = \operatorname*{argmin}_{\beta_0\in\mathbb{R},\beta_1\in\mathbb{R}} \sum_{i=1}^n \big(X_i - \beta_0 - \beta_1(i/n - t)\big)^2 K\Big(\frac{i/n - t}{b_n}\Big), \quad t \in [0,1],\end{aligned} \tag{2.1}$$

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

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

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. It turns out that the procedures proposed in this paper are not sensitive with respect to the choice of h_d and N, provided that these parameters have been chosen sufficiently small and large, respectively (see Section 5 for a further discussion).

A statistic of the type (2.2) has been proposed by [19] to estimate the inverse of a strictly increasing regression function, but we use it here without assuming monotonicity of the mean function μ. Observing that $\hat{\mu}_{b_n}(t)$ is a consistent estimate of $\mu(t)$ we argue (rigorous arguments are given later) that

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

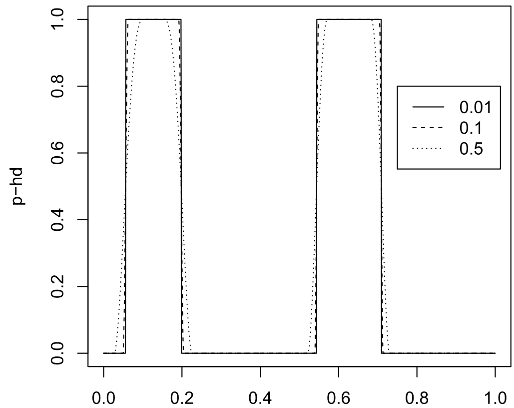

FIG. 2. *Smooth approximation p_{h_d} of the step function $q = \mathbf{1}_{\mathcal{M}_c^+}$ for different choices of the bandwidth h_d.*

as $n, N \to \infty$, $h_d \to 0$. In Figure 2 we display the functions

$$p_{h_d} : t \to \frac{1}{h_d} \int_c^\infty K_d\left(\frac{\mu(t) - \mu(0) - u}{h_d}\right) du \quad \text{and} \quad q : t \to \mathbf{1}\big(\mu(t) - \mu(0) \geq c\big)$$

and visualize that p_{h_d} is a smooth approximation of the indicator function for decreasing h_d (for the function considered in Figure 1). This smoothing is introduced to derive the asymptotic properties of the statistic $\hat{T}_{N,c}^+$ and to construct a valid bootstrap procedure without estimating the critical roots and derivatives of the regression function. Thus, intuitively (rigorous arguments will be given in the following sections) the statistic $\hat{T}_{N,c}^+$ is a consistent estimator of T_c^+ and a similar argument for T_c^- will provide a consistent estimator of the quantity T_c defined in (1.5). The null hypothesis is finally rejected for large values of this estimate.

In order to make these heuristic arguments more rigorous, we make the following basic assumptions for the model (1.3).

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 $I_{1,\delta}, \dots, I_{k_\delta,\delta}$, such that $\bigcup_{i=1}^{k_\delta} I_{i,\delta}$ is a decomposition of

$$\{t \in [0, 1] : |\mu(t) - \mu(0) - c| \leq \delta\} \cup \{t \in [0, 1] : |\mu(t) - \mu(0) + c| \leq \delta\},$$

where the number of intervals k_δ satisfies $\sup_{0 \leq \delta \leq \epsilon_0} k_\delta \leq M$ 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 $|\mu(1) - \mu(0)| \neq c$.

It is worthwhile to mention that all results presented in the paper remain true if the regression function is Lipschitz continuous on the interval [0, 1] and the assumptions regarding its differentiability (such as Assumption 2.1) hold in a neighbourhood of the critical roots. Our first result makes the approximation of T_c^+ by its deterministic counterpart

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

in (2.3) rigorous. 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),*

$$T_{N,c}^+ - T_c^+ = O\big(\max\{h_d^\iota, N^{-1}\}\big)$$

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

PROOF. By elementary calculations it follows that

$$\begin{aligned}
&\int_c^\infty \frac{1}{h_d}K_d\Big(\frac{\mu(i/N)-\mu(0)-u}{h_d}\Big)\,du - \mathbf{1}\big(\mu(i/N)-\mu(0) > c\big)\\
&\quad= \mathbf{1}\big(\big\{\big|c-\big(\mu(i/N)-\mu(0)\big)\big| \le h_d\big\}\big)\int_{\frac{c-\mu(i/N)+\mu(0)}{h_d}}^\infty K_d(x)\,dx\\
&\qquad - \mathbf{1}\big(\big\{\mu(i/N)-\mu(0)-h_d < c \le \mu(i/N)-\mu(0)\big\}\big).
\end{aligned}$$

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

$$\begin{aligned}
&\big|T_{N,c}^+ - T_c^+\big|\\
&\quad= \bigg|\frac{1}{N}\sum_{i=1}^{N}\int_c^\infty \frac{1}{h_d}K_d\Big(\frac{\mu(\frac{i}{N})-\mu(0)-u}{h_d}\Big)\,du - \mathbf{1}\Big(\mu\Big(\frac{i}{N}\Big)-\mu(0) > c\Big)\bigg|\\
&\qquad + O\Big(\frac{1}{N}\Big)\\
&\quad\le \frac{2}{N}\sum_{i=1}^{N}\mathbf{1}\big(\big|\mu(i/N)-\mu(0)-c\big| \le h_d\big) + O\big(N^{-1}\big)\\
&\quad= 2m_{c+\mu(0),h_d}(\mu) + O\big(N^{-1}\big) = O\Big(\max\Big\{h_d^\iota, \frac{1}{N}\Big\}\Big)
\end{aligned}$$

as $N \to \infty$, $h_d \to 0$. □

2.2. *Alternatives measures of mass excess.* In this section, we briefly mention several alternative measures of mass excess, which might be of interest in applications and for which similar results as stated in this paper can be derived. For the sake of brevity, we do not state these results in full detail in this paper and only describe the measures with corresponding estimates.

2.2.1. *Deviations from an average trend.* In applications, one might be also interested if there exist relevant deviations of the sequence $(\mu(i/n))_{i=\lfloor nt_0\rfloor+1,\dots,n}$ from an average trend formed from the previous period $(\mu(i/n))_{i=1,\dots,\lfloor nt_0\rfloor}$. This question can be addressed by estimating the quantity

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

Using similar arguments as given in this paper (and the Supplementary Material [22]) one can prove consistency and derive the asymptotic distribution of the estimate

$$\frac{1}{N} \sum_{i=\lfloor Nt_0\rfloor}^{N} \int_c^\infty \frac{1}{h_d} K_d\bigg(\frac{\hat{\mu}_{b_n}(i/N) - \int_0^{t_0} \hat{\mu}_{b_n}(s)\,ds - u}{h_d}\bigg)\,du,$$

where $\hat{\mu}_{b_n}$ is local linear estimator of μ (in Section 4 we will use a bias corrected version of $\hat{\mu}_{b_n}$).

2.2.2. *Relative deviations.* If $\mu(0) \neq 0$ an alternative measure of excess can be defined by

$$\int_0^1 \mathbf{1}\bigg(\bigg|\frac{\mu(t) - \mu(0)}{\mu(0)}\bigg| > c\bigg)\,dt = \lambda\bigg(\bigg\{t \in [0, 1] \colon \bigg|\frac{\mu(t) - \mu(0)}{\mu(0)}\bigg| > c\bigg\}\bigg). \tag{2.5}$$

This measure of excess allows to define a relevant change in the mean relative to its initial value and makes the choice of the constant c easier in applications. For example, if one chooses $c = 0.1$, one is interested in relevant deviation from the initial value by more than 10%. The quantity in equation (2.5) can be estimated in a similar way as described in the previous paragraph and the details are omitted for the sake of brevity.

2.3. *Locally stationary processes.* In Sections 3 and 4, we will establish the asymptotic properties of the statistic $\hat{T}^+_{N,c}$ as an estimator of T^+_c and derive a bootstrap approximation to derive critical values. Since we are interested in a procedure for nonstationary processes we require several technical assumptions on the error process in model (1.3). The less experienced reader can easily skip this paragraph and consider an independent identically distributed array of centered random variables $\epsilon_{i,n}$ in model (1.3) with variance σ^2. The main challenge in the proofs is neither the dependence structure nor the nonstationarity of the error process but consists in the fact that definition (2.2) defines a complicated map from the class of estimators to the Lebesgue measure of random sets of the form $\{t : |\hat{\mu}_{b_n}(t) - \hat{\mu}_{b_n}(0)| > c\}$. Thus, although a standardized version of the local linear estimator $\hat{\mu}_{b_n}$ is asymptotically normally distributed (under suitable conditions), a rigorous analysis of this mapping is required to derive the distributional properties of the statistic $\hat{T}^{+}_{N,c}$. These depend sensitively on the local behaviour of the function μ at points satisfying the equation $|\mu(t) - \mu(0)| = c$ and the corresponding analysis represents the most important part of the work, which is independent of the error structure in model (1.3).

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

Let $\eta = (\eta_i)_{i\in\mathbb{Z}}$ be a sequence of independent identically distributed random variables, $\mathcal{F}_i = \{\eta_s : s \leq i\}$, denote by $\eta' = (\eta'_i)_{i\in\mathbb{Z}}$ an independent copy of η and define $\mathcal{F}^*_i = (\dots, \eta_{-2}, \eta_{-1}, \eta'_0, \eta_1, \dots, \eta_i)$. 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

$$\delta_q(G, k) = \sup_{t\in[0,1]} \big\|G(t, \mathcal{F}_k) - G\big(t, \mathcal{F}^*_k\big)\big\|_q. \tag{2.6}$$

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

$$\sup_{0\leq s<t\leq 1} \big\|G(t, \mathcal{F}_i) - G(s, \mathcal{F}_i)\big\|_q / |t - s| < \infty. \tag{2.7}$$

The filter G is used to model nonstationarity. 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 (2.7) means that the data generating mechanism G is varying smoothly in time. We refer to [47] for more details, in particular for examples of locally stationary linear and nonlinear time series, calculations of the dependence measure (2.6) and for the verification of (2.7). With this notation, we make the following assumptions regarding the error process in model (1.3).

ASSUMPTION 2.2. The error process $(\epsilon_{i,n})_{i=1,\dots,n}$ in model (1.3) 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) 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* $\sigma^2(t) := \sum_{i=-\infty}^{\infty} \operatorname{cov}(G(t, \mathcal{F}_i), G(t, \mathcal{F}_0))$, $t \in [0, 1]$ of the filter G is Lipschitz continuous and *nondegenerate*, 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 (1.3) 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 Gaussian approximation of the locally stationary time series; see, for example, [48].

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 $\mathcal{C}$ do not vanish. These assumptions are quite common in the literature (see, e.g., condition (B.ii) in [27] or assumption (A1) in [37]). We discuss this case separately because of (at least) two reasons. First, the results and required assumptions are slightly simpler here. Second, and more importantly, 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 (1.6), (1.7) and (1.8) difficult. The general case is postponed to Section 4, where we solve the bias problem and also introduce a bootstrap procedure. 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 Theorems 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 $t_1^+, \dots, t_{k^+}^+$ of the equation $\mu(t) - \mu(0) = c$ satisfying $\dot{\mu}(t_j^+) \neq 0$ for $1 \le j \le k^+$, and define*

$$\bar{R}_{1,n} = \frac{n^{1/4} \log^2 n}{nb_n}, \qquad \bar{R}_{2,n} = \left(\frac{1}{Nb_n} + \frac{1}{Nh_d}\right)(b_n \wedge h_d),$$

$$\bar{\chi}_n = \left(b_n^4 + \frac{1}{nb_n}\right) h_d^{-1}.$$

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

$$\sqrt{nb_n}\left(\hat{T}^+_{N,c} - T^+_c - \sum_{j=1}^{k^+} \mu_{2,K} b_n^2 \frac{\ddot{\mu}(t_j^+)}{|\dot{\mu}(t_j^+)|} + \frac{b_n^2 c_{2,K}\ddot{\mu}(0)}{2c_{0,K}} \frac{1}{|\dot{\mu}(t_j^+)|}\right) \overset{\mathcal{D}}{\Rightarrow} \mathcal{N}(0, \tau^{2,+}),$$

where $\tau^{2,+} = \tau_1^{2,+} + \tau_2^{2,+}$,

$$\tau_1^{2,+} = \sum_{s=1}^{k^+} \frac{\sigma^2(t_s^+)}{\dot{\mu}(t_s^+)^2} \int K^2(x)\,dx,$$

$$\tau_2^{2,+} = \frac{\sigma^2(0)}{c_{0,K}^2}\left(\sum_{j=1}^{k^+} \frac{1}{|\dot{\mu}(t_j^+)|}\right)^2 \int_0^1 (\mu_{2,K} - t\mu_{1,K})^2 K^2(t)\,dt,$$

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

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

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 $t \in \mathcal{C}^+ = \{t \in [0,1] : \mu(t) - \mu(0) = c\}$. This condition guarantees that the mean function μ is strictly monotone in a neighbourhood of the roots. Moreover, Assumption 2.1(b), Assumptions 2.2 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 (1.8).

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* $\dot{\mu}(t_j^-) \neq 0$ $(j = 1, \ldots, k^-)$, *then, as* $n \to \infty$,

$$\sqrt{nb_n}\big(\hat{T}^+_{N,c} - T^+_c - \beta_c^+, \hat{T}^-_{N,c} - T^-_c - \beta_c^-\big)^T \overset{\mathcal{D}}{\Longrightarrow} \mathcal{N}(0, \tilde{\Sigma}),$$

where

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

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*

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

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 [37]. 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 [27]).

REMARK 3.2. The most general assumptions under which the results of our paper hold are the following.

(a) The mean trend is a piecewise Lipschitz continuous function, with a bounded number of jump points. If $D^+(t_0)$ and $D^-(t_0)$ denote the limit of the function $|\mu(\cdot) - \mu(0)|$ from the left an right at the jump point t_0, then $(D^+(t_0) - c)(D^-(t_0) - c) > 0$. In other words, at any jump, the function $|\mu(\cdot) - \mu(0)|$ does not "cross" the level c.

(b) There is a finite number of critical roots and the mean trend function has a Lipschitz continuous second derivative in a neighbourhood of each critical root.

In particular, we exclude the case where jumps occur at critical roots, but there might be jumps at other points in the interval [0, 1]. In this case, the local linear estimator $\hat{\mu}_{b_n}$ has to be modified to address for these jumps (see [35] or [23] among others). For the sake of a transparent representation and for the sake of brevity, we state our results under Assumptions 2.1 and 2.2.

Theorems 3.1 and 3.2 can be used to construct tests for the hypotheses (1.7) and (1.8). 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 (1.6). However, such tests would either require undersmoothing or estimation of the bias β_c^+ and β_c^- in (3.1), which is not an easy task. We address this problem by a jackknife method in the following section where we also develop a bootstrap test to avoid the estimation of the critical roots.

4. Bias correction and bootstrap. In this section, we will address the bias problem mentioned in the previous section adopting the jackknife bias reduction technique proposed by [38]. In a second step, we will use these results to construct a bootstrap procedure. Moreover, we also relax the main assumption in Section 3 that the derivative of the mean function does not vanish at critical roots $t \in \mathcal{C}$.

4.1. *Bias correction.* 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 [46] 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 [48] showed that the estimate $\tilde{\mu}_{b_n}$ is asymptotically equivalent to a local linear estimate with kernel

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

In order to use these bias corrected estimators for the construction of tests for the hypotheses defined in (1.6)–(1.8), 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 C.2 in the Supplementary Material [22] we obtain the stochastic expansion

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

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

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

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

$$\tilde{T}_{N,c}^+ = \frac{1}{N}\sum_{i=1}^{N}\int_c^{\infty} \frac{1}{h_d} K_d\left(\frac{\tilde{\mu}_{b_n}(i/N) - \tilde{\mu}_{b_n}(0) - u}{h_d}\right) du. \tag{4.2}$$

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 $\mu^{(1)}$ does not vanish at t. A point $t \in \mathcal{C}$ is called a *critical point of μ of order $k \ge 1$* if the first k derivatives of μ at t vanish while the $(k+1)$st derivative of μ at t is nonzero, that is $\mu^{(s)}(t) = 0$ for $1 \le s \le k$ and $\mu^{(k+1)}(t) \ne 0$. 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 v_s^- and v_s^+, respectively ($1 \le s \le k^-$, $1 \le s \le k^+$). 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 $0 = s_0 < s_1 < \cdots < s_q < s_{q+1} = 1$ such that the mean function μ is strictly monotone on each interval $(s_i, s_{i+1}]$ ($0 \le i \le q$).

It is shown in Lemma C.1 of the Supplementary Material [22] that under the assumptions made so far the set $\{t : |\mu(t) - c| \le h_n, t \in [0, 1]\}$ can be decomposed as a union of disjoint "small" intervals around the critical roots t_i^+ and t_i^-, whose Lebesgue measure is of order $h_n^{1/(v_i^+ + 1)}$ and $h_n^{1/(v_i^- + 1)}$, respectively, and therefore depends on the order of the corresponding root. In the Supplementary Material [22] we prove the following result, which clarifies the distributional properties of the estimator $\tilde{T}_{N,c}^+$ defined in (4.2) if the sample size converges to infinity.

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

$$\chi_n^+ = \left(b_n^6 + \frac{1}{nb_n}\right) h_d^{-2} h_d^{\frac{1}{v^+ + 1}}, \qquad R_{1,n}^+ = h_d^{-\frac{v^+}{v^+ + 1}} \left(b_n^3 + \frac{1}{nb_n}\right),$$

$$R_{2,n}^+ = \frac{n^{1/4} \log^2 n}{nb_n} h_d^{-\frac{v^+}{v^+ + 1}}, \qquad R_{3,n}^+ = \left(\frac{1}{Nb_n} + \frac{1}{Nh_d}\right)\left(b_n \wedge h_d^{\frac{1}{v^+ + 1}}\right).$$

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

$$\pi_n := \left(b_n^3 + (nb_n)^{-1/2} \log n\right) \log n,$$

then we have the following results:

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

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

where

$$\sigma_1^{2,+} = \left(\int K_d\left(z^{v^++1}\right)dz\right)^2 \left((v^++1)!\right)^{\frac{2}{v^++1}} \times \sum_{\{t_l^+ : v_l^+ = v^+\}} \frac{\sigma^2(t_l^+)}{|\mu^{(v^++1)}(t_l^+)|^{\frac{2}{v^++1}}} \int \left(K^*(x)\right)^2 dx, \tag{4.3}$$

$$\sigma_2^{2,+} = \sigma^2(0)\left((v^++1)!\right)^{\frac{2}{v^++1}} \int \left(\bar{K}^*(t)\right)^2 dt \times \left(\sum_{\{t_l^+ : v_l^+ = v^+\}} \left|\mu^{(v^++1)}(t_l^+)\right|^{\frac{-1}{v^++1}} \int K_d\left(z^{v^++1}\right)dz\right)^2. \tag{4.4}$$

(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*

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

where

$$\begin{aligned}\rho_1^{2,+} = {} & \left|(v^++1)!\right|^{\frac{1}{v^++1}} \sum_{\{t_l^+ : v_l^+ = v^+\}} \frac{\sigma^2(t_l^+)}{|\mu^{(v^++1)}(t_l^+)|^{\frac{2}{v^++1}}} \\ & \times \int K^*(u)K^*(v)K_d\left(z^{v^++1}\right) \\ & \times K_d\left(\left(z + r\left|\frac{(v^++1)!}{\mu^{(v^++1)}(t_l^+)}\right|^{\frac{-1}{v^++1}}(v-u)\right)^{v^++1}\right) du\,dv\,dz\end{aligned}$$

and $\rho_2^{2,+} = r^{-1}\sigma_2^{2,+}$, *where* $\sigma_2^{2,+}$ *is defined in* (4.4).

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 $b_n/h_d^{\frac{1}{v^++1}} = r \in [0,\infty)$, Theorem 4.1 and Proposition 2.1 give $\sqrt{nh_d}h_d^{v^+/2(v^++1)}(\tilde{T}_{N,c}^+ - T_c^+) - R_n \stackrel{\mathcal{D}}{\Longrightarrow} \mathcal{N}(0, \rho_1^{2,+} + \rho_2^{2,+})$, 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 $O(\sqrt{nh_d}h_d^{(v^++2)/2(v^++1)})$, 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 $nh_d^3 = o(1)$ 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 (4.2) such that condition (a) of Theorem 4.1 holds, that is $h_d = o(b_n^{v^++1})$. 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

$$\tilde{T}_{N,c}^- = \frac{1}{N}\sum_{i=1}^{N}\int_{-\infty}^{-c}\frac{1}{h_d}K_d\left(\frac{\tilde{\mu}_{b_n}(i/N) - \tilde{\mu}_{b_n}(0) - u}{h_d}\right)du. \tag{4.5}$$

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*

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

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

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

and $\sigma_1^{2,-}$, $\sigma_2^{2,-}$ are defined similarly as $\sigma_1^{2,+}$, $\sigma_2^{2,+}$ in (4.3), (4.4), *respectively.*

The continuous mapping theorem and Theorem 4.2 imply the weak convergence of the estimator $\tilde{T}_{N,c}$ of T_c, that is $\sqrt{nb_n}h_d^{\frac{v}{v+1}}(\tilde{T}_{N,c} - T_c) \to N(0, \sigma^2)$, where $v = \max\{v^+, v^-\}$ and the asymptotic variance is given by $\sigma^2 = \Sigma_{11}\mathbf{1}(v^+ \geq v^-) + \Sigma_{22}\mathbf{1}(v^+ \leq v^-) + 2\Sigma_{12}\mathbf{1}(v^+ = v^-)$.

4.2. *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 (1.6). 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). As an alternative, 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 the stochastic approximation $\tilde{T}^+_{N,c} - T^+_c = I' + o_p((\sqrt{nb_n}h_d^{v^+/(v^++1)})^{-1})$, where the statistic I' is defined as

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

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

ALGORITHM 4.1. (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 $\bigcup_{k=1}^{v^+}\mathcal{U}_\varepsilon(t_k^+)$ 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 (4.2), respectively.

(3) Calculate

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

(4) Let $q^+_{1-\alpha}$ denote the the $1-\alpha$ quantile of a centered normal distribution with variance $\bar{V}$, then the null hypothesis in (1.7) is rejected, whenever

$$nNb_nh_d(\tilde{T}^+_{N,c} - \Delta) > q^+_{1-\alpha}. \tag{4.6}$$

THEOREM 4.3. *Assume that the conditions of Theorem* 4.1(a) *are satisfied, then the test* (4.6) *defines a consistent and asymptotic level* α *test for the hypotheses* (1.7).

REMARK 4.1. (a) It follows from the proof of Theorem 4.3 in the Supplementary Material [22] that

$$\mathbb{P}(\text{test (4.6) rejects}) \longrightarrow \begin{cases} 1 & \text{if } T_c^+ > \Delta, \\ \alpha & \text{if } T_c^+ = \Delta, \\ 0 & \text{if } T_c^+ < \Delta. \end{cases}$$

Moreover, these arguments also show that the power of the test (4.6) depends on the "signal to noise ratio" $(\Delta - T_c^+)/\sqrt{\sigma_1^{2,+} + \sigma_2^{2,+}}$ and that it is able to detect local alternatives converging to the null at a rate $O((nb_n)^{-1/2} h_d^{-v^+/(v^+ +1)})$. 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 (B.2) in the Supplementary Material [22]).

(b) It is also of interest to discuss some uniformity properties in this context. For this purpose we consider the situation in Theorem 4.3, assume that f is a potential mean function in (1.3) and denote by v_f^+ and q_f the corresponding quantities in Assumptions 4.1 and 4.2 for $\mu = f$. For given numbers $\tilde{q}, \tilde{v} < \infty$ let $\mathcal{F}$ denote the class of all $3 \vee (\tilde{v} + 1) + 1$ times differentiable functions f on the interval [0, 1] satisfying $\sup_{f \in \mathcal{F}} v_f^+ \le \tilde{v}$ and $\sup_{f \in \mathcal{F}} q_f \le \tilde{q}$. Consider a sequence $(\Delta_n)_{n \in \mathbb{N}}$ satisfying $\sqrt{nb_n} h_d^{\tilde{v}/(\tilde{v}+1)} (\Delta - \Delta_n) \to -\infty$ and define for a given level $c > 0$, constants M, L, $\eta, \iota > 0$ the set $\mathcal{F}_c(M, \eta, \iota, \tilde{q}, \tilde{v}, L, \Delta_n)$ as the class of all functions $f \in \mathcal{F}$ with the properties:

(i) The cardinality of the set $\mathcal{E}_c^+(f) = \{t \in [0, 1] : f(t) - f(0) = c\}$ is at most M.

(ii) $\min\{|t_1 - t_2| : t_1, t_2 \in \mathcal{E}_c^+(f); t_1 \neq t_2\} \ge \eta$; $\min\{t_1 : t_1 \in \mathcal{E}_c^+(f)\} \ge \eta$; $\max\{t_1 : t_1 \in \mathcal{E}_c^+(f)\} \le 1 - \eta$.

(iii) $\sup_{t \in [0,1]} (f(t) - f(0)) \ge c + \iota$.

(iv) $\sup_{t \in [0,1]} \max_{1 \le s \le 3 \vee (\tilde{v}+1)+1} |f^{(s)}(t)| \le L$.

(v) $T_{f,c}^+ := \int \mathbf{1}(f(t) - f(0) > c)\, dt \ge \Delta_n$.

If $\mathbb{P}_f$ denotes the distribution of the process $(X_{i,n})_{i=1,\dots,n}$ in model (1.3) with $\mu = f$, then it follows by a careful inspection of the proof of Theorem 4.3 in the Supplementary Material [22] that

$$\lim_{n \to \infty} \inf_{f \in \mathcal{F}_c(M, \eta, \iota, \tilde{q}, \tilde{v}, L, \Delta_n)} \mathbb{P}_f(\text{test (4.6) rejects}) = 1.$$

(c) The bootstrap procedure can easily be modified to test the hypothesis (1.6) referring to the quantity T_c. In step (2), we additionally calculate the statistic $\tilde{T}_{N,c}^-$ defined in (4.5), $\tilde{T}_{N,c} = \tilde{T}_{N,c}^+ + \tilde{T}_{N,c}^-$ and the quantity

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

$$\times\left(K^*\left(\frac{i/N-j/n}{b_n}\right)-\bar{K}^*\left(\frac{j}{nb_n}\right)\right)\right)^2,$$

where

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

Finally, the null hypothesis (1.6) is rejected if $nNb_nh_d(\tilde{T}_{N,c}-\Delta)>q_{1-\alpha}$, 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 $t\in[m/n,1-m/n]$

$$\hat{\sigma}^2(t)=\sum_{j=1}^n\frac{m\Delta_j^2}{2}\omega(t,j), \tag{4.7}$$

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

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

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 (4.7) does not involve estimated residuals. The following result shows that $\hat{\sigma}^2$ is consistent and can be used in Algorithm 4.1.

THEOREM 4.4. *Let Assumptions 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 (4.7) satisfies*

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

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

$$\hat{\sigma}^2(t)-\sigma^2(t)=O_p\left(\sqrt{\frac{m}{n\tau_n}}+\frac{1}{m}+\tau_n^2+m^{5/2}/n\right) \tag{4.8}$$

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

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

Note that error term $\sqrt{\frac{m}{n\tau_n}} + \frac{1}{m} + \tau_n^2$ in (4.8) 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 (4.7) achieves a better rate than the long-run variance estimator proposed in [48] (see Theorem 5 in this reference).

5. Simulation study. 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 (4.6) for the hypotheses (1.7). Similar results can be obtained for the corresponding tests for the hypotheses (1.6) and (1.8). The code used to obtain the presented results is available from the second author on request.

Throughout this section, all kernels are chosen as Epanechnikov kernel. 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 $\tilde{\mathbf{e}}_b = (\tilde{e}_{1,b}, \dots, \tilde{e}_{n,b})^T$. Throughout this section, we use the bandwidth

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

where $\hat{\Gamma}_n$ is an estimator of the covariance matrix $\Gamma_n := \{\mathbb{E}(\epsilon_{i,n}\epsilon_{j,n})\}_{1\le i,j\le n}$, which is obtained by the banding techniques as described in [45].

It turns out that Algorithm 4.1 is not very sensitive with respect to the choice of the bandwidth h_d as long as it is chosen sufficiently small. Similarly, the number N of knots used in the Riemann approximation (2.2) has a negligible influence on the test, provided it has been chosen sufficiently large. 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, and investigate the influence of other choices below. The number of knots is always given by $N = n$. 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 4.2]. For the data analysis in Section A of the Supplementary Material [22], we suggest a data-driven procedure and use a slight modification of the minimal volatility method as proposed by [48]. To be precise—in order to avoid choosing too large values for m and τ—we penalize the quantity

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

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 (4.7) of the long-run variance with parameters m_h and τ_j and IS is the average of the quantities $ISE_{h,j}$.

All simulation results presented in this section are based on 2000 simulation runs. We consider the model (1.3) 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 normally distributed random variables. For the mean trend we consider the following two cases:

(a) $\mu(t) = 8(-(t - 0.5)^2 + 0.25)$;
(b) $\mu(t) = \sin(2|t - 0.6|\pi)(1 + 0.4t)$.

Note that the mean trend (b) is not differentiable at the point 0.6. However, using similar but more complicated arguments as given in the Supplementary Material [22], 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 $t_1^+, \dots, t_{k^+}^+, t_1^-, \dots, t_{k^-}^-$.

We begin illustrating the finite sample properties of the (uncorrected) estimator $\hat{T}^+_{N,c}$ in (2.2) and its bias correction $\tilde{T}^+_{N,c}$ in (4.2) for the quantity T_c^+, where $c = 1.8$. The corresponding values of T_c^+ are $T^+_{1.8} = 0.3163$ and $T^+_{1.8} = 0.1406$ in models (a) and (b), respectively. In Table 1 we display the bias and standard deviation of the two estimators. We observe a substantial reduction of the bias by a factor between 5 and 75, while there is a slight increase in standard deviation. Except for one case the bias corrected estimate $\tilde{T}^+_{N,c}$ has a smaller mean squared error than the uncorrected estimate.

Next, we investigate the finite sample properties of the bootstrap test (4.6) for the hypotheses (1.7), where the threshold is given by $\Delta = 0.3$ and $\Delta = 0.15$. Following the discussion in Remark 4.1(a), we display in Table 2 the simulated type 1 error at the boundary of the null hypothesis in (1.7), that is $T_c^+ = \Delta$. A good

TABLE 1
Simulated bias and standard deviation of the estimators $\hat{T}^+_{N,c}$ and its bias correction $\tilde{T}^+_{N,c}$, where $c = 1.8$. The sample size is $n = 500$ and the bandwidth has been chosen by GCV

	Model							
	(a, I)		(a, II)		(b, I)		(b, II)	
Accuracy	Bias	sd	Bias	sd	Bias	sd	Bias	sd
$\hat{T}^+_{N,1.8}$	−0.105	0.063	−0.122	0.077	−0.077	0.055	−0.054	0.060
$\tilde{T}^+_{N,1.8}$	−0.008	0.065	−0.011	0.069	−0.001	0.076	0.010	0.085

TABLE 2

Simulated level of the test (4.6) *at the boundary of the null hypothesis* (1.7). *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* $b_{cv}^- = b_{cv} - 0.05$, $b_{cv}^+ = b_{cv} + 0.05$

			Model							
			(a, I)		(b, I)		(a, II)		(b, II)	
n	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

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 $c = 1.78$ for 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.

We also briefly address the problem of the sensitivity of the procedure with respect to the choice of the bandwidth h_d. For this purpose, we consider the same scenarios as in Table 2. For the sake of brevity, we restrict ourselves to the case $n = 500$ and the data driven bandwidth b_{cv}. The results are shown in Table 3 for the bandwidths $h_d = n^{-1/2}/2 = 0.0224$, $h_d = 0.0112$ and $h_d = 0.0056$ and show that the procedure is very stable with respect to the choice h_d as long as h_d is chosen sufficiently small.

In Figure 3, we investigate the properties of the test (4.6) as a function of the threshold Δ and level c, where we restrict ourselves to the scenario ((I), (a)). For

TABLE 3
Simulated level of the test (4.6) at the boundary of the null hypothesis (1.7) for different choices of the bandwidth h_d. The sample size is $n = 500$. The bandwidth b_{cv} is chosen by GCV

			Model							
			(a, I)		(b, I)		(a, II)		(b, II)	
n	Δ	h_d	5%	10%	5%	10%	5%	10%	5%	10%
500	0.3	0.0224	4.6	9.55	5.45	9.85	5.65	9.25	6	10.1
		0.0112	5.3	9.5	6.75	11.01	4.95	8.85	4.6	8.15
		0.0056	4.9	9.5	6.7	11.25	5.2	9.3	5.25	9.5
500	0.15	0.0224	6.55	11.25	5.1	9.15	7.75	12.2	5.15	9.25
		0.0112	6.1	10.25	5.7	9.35	6.4	10.95	5.45	8.75
		0.0056	7.45	12.15	6.25	10.25	7.55	11.95	6.9	11.8

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 3 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 (4.6) for the hypotheses (1.7) with $c = 1.82$ and $\Delta = 0.3$, where the bandwidth is chosen as $b_n = 0.2$. The model is

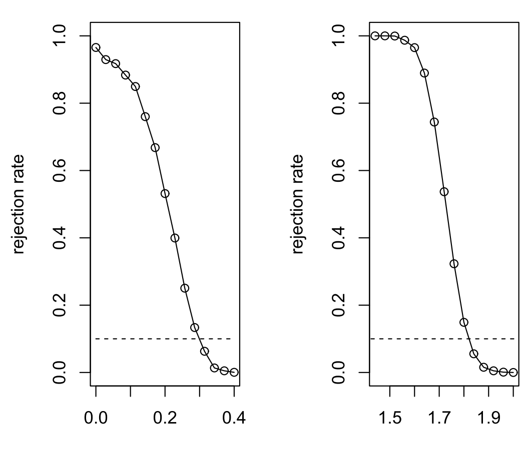

FIG. 3. *Simulated rejection probabilities of the test (4.6) in model (1.3) for varying values of c and Δ. Left: $c = 1.82$, $\Delta \in [0, 0.4]$ (the case $\Delta = 0.3$ corresponds to the boundary of the null hypothesis). Right: $\Delta = 0.3$, $c \in [1.44, 2]$ (the case $c = 1.82$ corresponds to the boundary of the null hypothesis). The dashed horizontal line represents the nominal level 10%.*

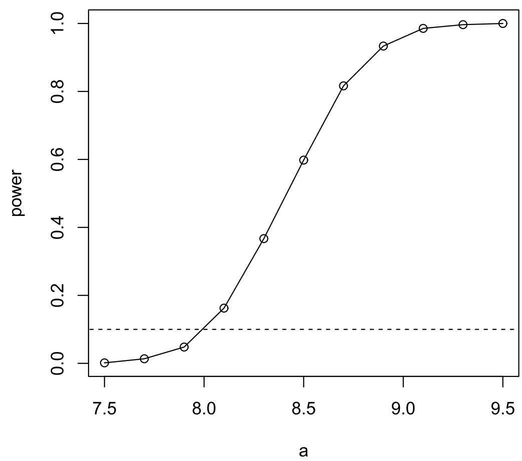

FIG. 4. *Simulated power of the test* (4.6) *in model* (1.3) *for the hypothesis* (1.7) *with* $c = 1.82$ *and* $\Delta = 0.3$. *The mean functions are given by* (5.1) *and the case* $a = 8$ *corresponds to the boundary of the null hypothesis. The dashed horizontal line represents the nominal level* 10%.

given by (1.3) with error (I) and different mean functions

$$\mu(t) = a\bigl(-(t-0.5)^2 + 0.25\bigr), \quad a \in [7.5, 9.5] \tag{5.1}$$

are considered (here the case $a = 8$ corresponds to the boundary of the hypotheses). The results are presented in Figure 4, which demonstrate that the test (4.6) has decent power.

Although hypotheses of the form (1.6) have not been investigated in the literature so far it was pointed out by a referee that it might be of interest to see a comparison with tests for similar hypotheses. The method most similar in spirit to our approach is the test of [21] for the hypotheses (1.2). Note that the procedure of these authors assumes a constant mean before and after the (relevant) change point, while we investigate if a (inhomogeneous) process deviates from its initial mean substantially over a sufficiently long period. Thus—strictly speaking—none of the procedures is applicable to the other testing problem. On the other hand both tests address the problem of relevant changes under different perspectives and it might therefore be of interest to see their performance in the respective alternative testing problems. For this purpose we consider model (1.3) with the mean functions:

(III) $\mu(t) = 2.5\sin(\pi t)$,
(IV) $\mu(t) = 0$ for $t \in [0, 1/3)$ and $\mu(t) = 2.5$ for $t \in [2/3, 1]$,

and an independent error process $\epsilon_{i,n} \sim N(0,1)/4$. Note that model (III) corresponds to the situation considered in this paper (i.e., a continuously varying mean function), while model (IV) reflects the situation investigated in [21]. In Figure 5, we display the rejection probabilities of both tests if the level c varies from 0.5 to 2.75 (thus the curves are decreasing with increasing c). The significance level is

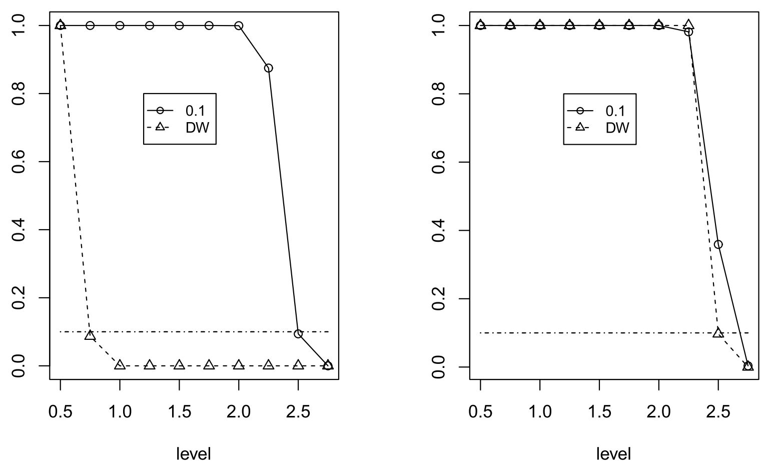

FIG. 5. *Rejection rates of the test of* [21] (*dashed line*) *and the bootstrap test* (4.6) *with* $\Delta = 0.1$ (*solid line*) *for various values of the level* c. *Left panel*: *regression function* (III); *right panel*: *regression function* (IV). *The nominal level is* 10%.

given by 10%, which means the value of c where the curve is 10% should be close to 2.5. For the hypotheses (1.6), we fixed Δ as 0.1, because for a comparison with the test of [21] it is irrelevant how long the threshold is exceeded and the power of the test (4.6) decreases for increasing values of Δ (see Figure 3).

We observe in the left panel of Figure 5 that the test of [21] performs poorly in model (III), where the mean is not constant and the conditions for its applications are not satisfied. On the other hand, the bootstrap test (4.6) shows a reasonable performance in model (IV) although the assumptions for its application are not satisfied. In particular, this test shows a similar performance as the test of [21] for small values of Δ, which is particularly designed for the hypotheses (1.2) (see the right panel of Figure 5).

6. Further discussion. We conclude this paper with a brief discussion of the extension of the proposed concept to the multivariate case and its relation to the concept of sojourn times in probability theory.

6.1. *Multivariate data.* The results of this paper can be extended to multivariate time series of the form $\mathbf{X}_{i,n} = \boldsymbol{\mu}(i/n) + \mathbf{e}_{i,n}$, where $\mathbf{X}_{i,n} = (X^1_{i,n}, \dots, X^m_{i,n})^T$, $\boldsymbol{\mu}(i/n) = (\mu^1(i/n), \dots, \mu^m(i/n))^T$ its corresponding expectation and $(\mathbf{e}_{i,n})_{i=1,\dots,n}$ is an m-dimensional time series such that $\mathbf{e}_{i,n} = \boldsymbol{G}(i/n, \mathcal{F}_i)$, where $\mathbf{G}(t, \mathcal{F}_i) = (G_1(t, \mathcal{F}_i), \dots, G_m(t, \mathcal{F}_i))^T$ is an m-dimensional filter. Assume that $\Sigma(t) = \sum_{i=-\infty}^{\infty} \operatorname{cov}(\boldsymbol{G}(t, \mathcal{F}_i), \boldsymbol{G}(t, \mathcal{F}_0))$ of the error process is strictly positive and let $\|\mathbf{v}\|$ denote the Euclidean norm of an m-dimensional vector $\mathbf{v}$. The excess mass for the m-dimensional mean function is then defined as $\boldsymbol{T}_c :=$

$\int_0^1 \mathbf{1}(\|\boldsymbol{\mu}(t) - \boldsymbol{\mu}(0)\| > c)\,dt$ and a test for the hypotheses $H_0 : \boldsymbol{T}_c \le \Delta$ versus $H_1 : \boldsymbol{T}_c > \Delta$ can be developed by estimating this quantity by

$$\hat{\boldsymbol{T}}_{N,c} = \frac{1}{N}\sum_{i=1}^{N}\int_{c^2}^{\infty}\frac{1}{h_d}K_d\Big(\frac{\|\hat{\boldsymbol{\mu}}(i/N) - \hat{\boldsymbol{\mu}}(0)\|^2 - u}{h_d}\Big)\,du,$$

where $\hat{\boldsymbol{\mu}}$ denote the vector of componentwise bias-corrected jackknife estimates of the vector of regression functions.

The corresponding bootstrap test is now obtained by rejecting the null hypothesis at level α, whenever $nNb_nh_d\hat{\boldsymbol{T}}_{N,c} - \Delta > q_{1-\alpha}$, where $q_{1-\alpha}$ is the $(1-\alpha)$-quantile of the random variable

$$\sum_{i,j} K_d\Big(\frac{\hat{\boldsymbol{g}}(i/N) - c^2}{h_d}\Big)\Big(K^*\Big(\frac{j/n - i/N}{b_n}\Big) - \bar{K}^*\Big(\frac{j}{nb_n}\Big)\Big)$$
$$\times\, \big(\nabla\hat{\boldsymbol{g}}(i/N)\big)^T\,\hat{\Sigma}^{1/2}(j/n)\boldsymbol{V}_j,$$

$\nabla\hat{\boldsymbol{g}}(u)$ is the gradient of the function $\hat{\boldsymbol{g}}(u) = \|\hat{\boldsymbol{\mu}}(u) - \hat{\boldsymbol{\mu}}(0)\|^2$, $\boldsymbol{V}_1, \boldsymbol{V}_2, \dots$ are independent standard normally distributed m-dimensional random vectors and $\hat{\Sigma}(t)$ is an analogue of the long run variance matrix estimator defined in (4.7).

Under similar conditions as stated in Assumptions 2.1, 2.2, 4.1, 4.2 and in Theorem 4.1(a), an analogue of Theorem 4.3 can be proved, that is, the bootstrap test has asymptotic level α and is consistent.

6.2. *Estimates of excess measures related to sojourn times.* The excess measures (1.10) and (1.11) based on sojourn times can easily be estimated under the assumption that the process $\{\epsilon(t) - \epsilon(0)\}_{t\in[0,1]}$ is stationary with density f. In this case, the quantities e_c and $p_{c,\Delta}$ can be expressed as

$$e_c = \mathbb{E}(S_c) = \iint_0^1 \mathbf{1}\big(\big|\mu(t) - \mu(0) + x\big| > c\big) f(x)\,dt\,dx,$$
$$p_{c,\Delta} = \mathbb{P}(S_c > \Delta) = \mathbb{E}\big(\mathbb{E}\big(\mathbf{1}(S_c > \Delta)|\epsilon(t) - \epsilon(0) = x\big)\big)$$

and corresponding estimators are given by

$$\hat{e}_c = \frac{1}{Nnh_d}\sum_{i=1}^{n}\sum_{s=1}^{N}\int_c^{\infty} K_d\Big(\frac{|\hat{\mu}(s/N) - \hat{\mu}(0) + \hat{Z}(i/n)| - u}{h_d}\Big)\,du,$$

$$\hat{p}_{c,\Delta} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\Big(\frac{1}{Nh_d}\sum_{s=1}^{N}\int_c^{\infty} K_d\Big(\frac{|\hat{\mu}(s/N) - \hat{\mu}(0) + \hat{Z}(i/n)| - u}{h_d}\Big)\,du > \Delta\Big),$$

respectively, where $\hat{\mu}(t) - \hat{\mu}(0)$ is a consistent estimator (say a local linear) of $\mu(t) - \mu(0)$ and $\hat{Z}(t) = \hat{\epsilon}(t) - \hat{\epsilon}(0)$ denotes the corresponding residual. Statistical analysis can then be developed along the lines of this paper.

However, in the case of a nonstationary error process as considered in this paper the situation is much more complicated and we leave the development of estimators and investigation of their (asymptotic) properties for future research.

Acknowledgements. The authors would like to thank Martina Stein who typed this manuscript with considerable technical expertise, V. Spokoiny for explaining his results to us and V. Golosnoy for some help with the literature on control charts. The authors are also grateful to four unknown reviewers for their constructive comments on an earlier version of this manuscript.

SUPPLEMENTARY MATERIAL

Supplement to "Detecting relevant changes in the mean of nonstationary processes—A mass excess approach" (DOI: 10.1214/19-AOS1811SUPP; .pdf). We provide proof for (B.17) and (B.18) for Theorem 4.1, proof of Theorem 4.4 and technical lemmas in the Supplementary Material.

REFERENCES

[1] ÁLVAREZ-ESTEBAN, P. C., DEL BARRIO, E., CUESTA-ALBERTOS, J. A. and MATRÁN, C. (2008). Trimmed comparison of distributions. *J. Amer. Statist. Assoc.* **103** 697–704. MR2435470

[2] ÁLVAREZ-ESTEBAN, P. C., DEL BARRIO, E., CUESTA-ALBERTOS, J. A. and MATRÁN, C. (2012). Similarity of samples and trimming. *Bernoulli* **18** 606–634. MR2922463

[3] ANDREWS, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica* **61** 821–856. MR1231678

[4] AUE, A., HÖRMANN, S., HORVÁTH, L. and REIMHERR, M. (2009). Break detection in the covariance structure of multivariate time series models. *Ann. Statist.* **37** 4046–4087. MR2572452

[5] AUE, A. and HORVÁTH, L. (2013). Structural breaks in time series. *J. Time Series Anal.* **34** 1–16. MR3008012

[6] BAI, J. and PERRON, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66** 47–78. MR1616121

[7] BAÍLLO, A. (2003). Total error in a plug-in estimator of level sets. *Statist. Probab. Lett.* **65** 411–417. MR2039885

[8] BERMAN, S. M. (1992). *Sojourns and Extremes of Stochastic Processes. The Wadsworth & Brooks/Cole Statistics/Probability Series.* Wadsworth & Brooks/Cole, Pacific Grove, CA. MR1126464

[9] BROWN, R. L., DURBIN, J. and EVANS, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *J. Roy. Statist. Soc. Ser. B* **37** 149–163. MR0378310

[10] CADRE, B. (2006). Kernel estimation of density level sets. *J. Multivariate Anal.* **97** 999–1023. MR2256570

[11] CHAMP, C. W. and WOODALL, W. H. (1987). Exact results for Shewhart control charts with supplementary runs rules. *Technometrics* **29** 393–399.

[12] CHANDLER, G. and POLONIK, W. (2006). Discrimination of locally stationary time series based on the excess mass functional. *J. Amer. Statist. Assoc.* **101** 240–253. MR2268042

[13] CHENG, M.-Y. and HALL, P. (1998). Calibrating the excess mass and dip tests of modality. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **60** 579–589. MR1625938

[14] CHERNOZHUKOV, V., FERNÁNDEZ-VAL, I. and GALICHON, A. (2009). Improving point and interval estimators of monotone functions by rearrangement. *Biometrika* **96** 559–575. MR2538757

[15] CHERNOZHUKOV, V., FERNÁNDEZ-VAL, I. and GALICHON, A. (2010). Quantile and probability curves without crossing. *Econometrica* **78** 1093–1125. MR2667913

[16] CHOW, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica* **28** 591–605. MR0141193

[17] CUEVAS, A., GONZÁLEZ-MANTEIGA, W. and RODRÍGUEZ-CASAL, A. (2006). Plug-in estimation of general level sets. *Aust. N. Z. J. Stat.* **48** 7–19. MR2234775

[18] DAHLHAUS, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25** 1–37. MR1429916

[19] DETTE, H., NEUMEYER, N. and PILZ, K. F. (2006). A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli* **12** 469–490. MR2232727

[20] DETTE, H. and VOLGUSHEV, S. (2008). Non-crossing non-parametric estimates of quantile curves. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 609–627. MR2420417

[21] DETTE, H. and WIED, D. (2016). Detecting relevant changes in time series models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **78** 371–394. MR3454201

[22] DETTE, H. and WU, W. (2019). Supplement to "Detecting relevant changes in the mean of nonstationary processes—A mass excess approach." DOI:10.1214/19-AOS1811SUPP.

[23] GIJBELS, I., LAMBERT, A. and QIU, P. (2007). Jump-preserving regression and smoothing using local linear fitting: A compromise. *Ann. Inst. Statist. Math.* **59** 235–272. MR2394168

[24] HARTIGAN, J. A. and HARTIGAN, P. M. (1985). The dip test of unimodality. *Ann. Statist.* **13** 70–84. MR0773153

[25] JANDHYALA, V., FOTOPOULOS, S., MACNEILL, I. and LIU, P. (2013). Inference for single and multiple change-points in time series. *J. Time Series Anal.* **34** 423–446. MR3070866

[26] KRÄMER, W., PLOBERGER, W. and ALT, R. (1988). Testing for structural change in dynamic models. *Econometrica* **56** 1355–1369. MR0973984

[27] MASON, D. M. and POLONIK, W. (2009). Asymptotic normality of plug-in level set estimates. *Ann. Appl. Probab.* **19** 1108–1142. MR2537201

[28] MERCURIO, D. and SPOKOINY, V. (2004). Statistical inference for time-inhomogeneous volatility models. *Ann. Statist.* **32** 577–602. MR2060170

[29] MÜLLER, D. W. and SAWITZKI, G. (1991). Excess mass estimates and tests for multimodality. *J. Amer. Statist. Assoc.* **86** 738–746. MR1147099

[30] NASON, G. P., VON SACHS, R. and KROISANDT, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **62** 271–292. MR1749539

[31] OMBAO, H., VON SACHS, R. and GUO, W. (2005). SLEX analysis of multivariate nonstationary time series. *J. Amer. Statist. Assoc.* **100** 519–531. MR2160556

[32] PAGE, E. S. (1954). Continuous inspection schemes. *Biometrika* **41** 100–115. MR0088850

[33] POLONIK, W. (1995). Measuring mass concentrations and estimating density contour clusters—An excess mass approach. *Ann. Statist.* **23** 855–881. MR1345204

[34] POLONIK, W. and WANG, Z. (2005). Estimation of regression contour clusters—An application of the excess mass approach to regression. *J. Multivariate Anal.* **94** 227–249. MR2167913

[35] QIU, P. (2003). A jump-preserving curve fitting procedure based on local piecewise-linear kernel estimation. *J. Nonparametr. Stat.* **15** 437–453. MR2017479

[36] RINALDO, A. and WASSERMAN, L. (2010). Generalized density clustering. *Ann. Statist.* **38** 2678–2722. MR2722453

[37] SAMWORTH, R. J. and WAND, M. P. (2010). Asymptotics and optimal bandwidth selection for highest density region estimation. *Ann. Statist.* **38** 1767–1792. MR2662359

[38] SCHUCANY, W. R. and SOMMERS, J. P. (1977). Improvement of kernel type density estimators. *J. Amer. Statist. Assoc.* **72** 420–423. MR0448691

[39] SPOKOINY, V. (2009). Multiscale local change point detection with applications to value-at-risk. *Ann. Statist.* **37** 1405–1436. MR2509078

[40] TAKÁCS, L. (1996). Sojourn times. *J. Appl. Math. Stoch. Anal.* **9** 415–426. MR1429264

[41] TSYBAKOV, A. B. (1997). On nonparametric estimation of density level sets. *Ann. Statist.* **25** 948–969. MR1447735

[42] VOGT, M. (2012). Nonparametric regression for locally stationary time series. *Ann. Statist.* **40** 2601–2633. MR3097614

[43] WELLEK, S. (2010). *Testing Statistical Hypotheses of Equivalence and Noninferiority*, 2nd ed. CRC Press, Boca Raton, FL. MR2676002

[44] WOODALL, W. H. and MONTGOMERY, D. C. (1999). Research issues and ideas in statistical process control. *J. Qual. Technol.* **31** 376–386.

[45] WU, W. B. and POURAHMADI, M. (2009). Banding sample autocovariance matrices of stationary processes. *Statist. Sinica* **19** 1755–1768. MR2589209

[46] WU, W. B. and ZHAO, Z. (2007). Inference of trends in time series. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 391–410. MR2323759

[47] ZHOU, Z. and WU, W. B. (2009). Local linear quantile estimation for nonstationary time series. *Ann. Statist.* **37** 2696–2729. MR2541444

[48] ZHOU, Z. and WU, W. B. (2010). Simultaneous inference of linear models with time varying coefficients. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **72** 513–531. MR2758526

DEPARTMENT OF MATHEMATICS
RUHR-UNIVERSITÄT BOCHUM
BUILDING IB 2/65
UNIVERSITÄTSSTRASSE 150
D-44801 BOCHUM
GERMANY
E-MAIL: holger.dette@ruhr-uni-bochum.de

CENTER FOR STATISTICAL SCIENCE
DEPARTMENT OF INDUSTRIAL ENGINEERING
TSINGHUA UNIVERSITY
BEIJING 100084
CHINA
E-MAIL: wuweichi@mail.tsinghua.edu.cn