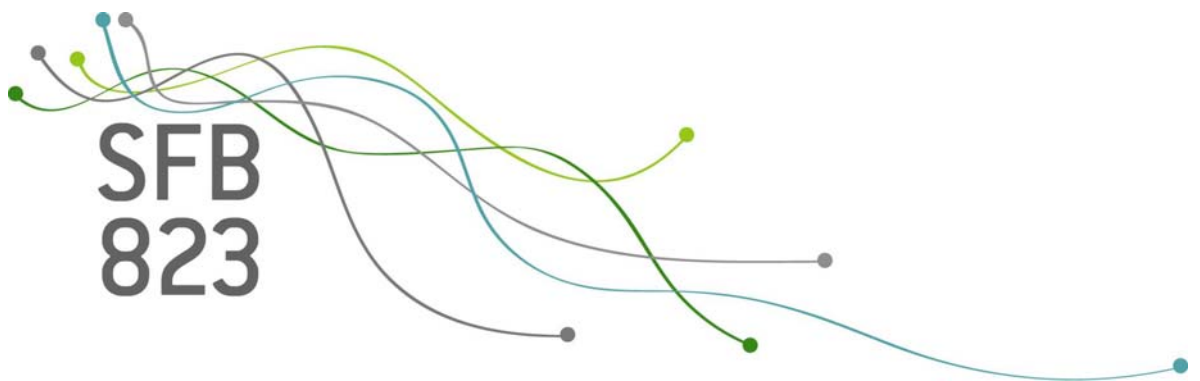# Change point analysis of second order characteristics in non-stationary time series

Holger Dette, Weichi Wu, Zhou Zhou

Change point analysis of second order characteristics in non-stationary time series

Holger Dette

Ruhr-Universität Bochum

Fakultät für Mathematik

44780 Bochum

Germany

email: holger.dette@ruhr-uni-bochum.de

Weichi Wu

University of Toronto

100 St. George St.

Toronto ON M5S 3G3

Canada

email: weichi.wu@utoronto.ca

Zhou Zhou

University of Toronto

100 St. George St.

Toronto ON M5S 3G3

Canada

email: zhou@utstat.toronto.edu

March 27, 2015

Abstract

A restrictive assumption in the work on testing for structural breaks in time series consists in the fact that the model is formulated such that the stochastic process under the null hypothesis of "no change-point" is stationary. This assumption is crucial to derive (asymptotic) critical values for the corresponding testing procedures using an elegant and powerful mathematical theory, but it might be not very realistic from a practical point of view. For example, if change point analysis for a particular parameter of the process (such as the variance) is performed, it is not necessary clear why other parameters (such as the mean or higher order moments) have to stay constant under the hypothesis that there is no change point in the parameter of interest.

This paper develops change point analysis under less restrictive assumptions and deals with

the problem of detecting change points in the marginal variance and correlation structures of a non-stationary time series. A CUSUM approach is proposed, which is used to test the "classical" hypothesis of the form $H_0 : \theta_1 = \theta_2$ vs. $H_1 : \theta_1 \neq \theta_2$, where θ_1 and θ_2 denote second order parameters (such as the variance or the lag k-correlation) of the process before and after a change point. The asymptotic distribution of the CUSUM test statistic is derived under the null hypothesis. This distribution depends in a complicated way on the dependency structure of the nonlinear non-stationary time series and a bootstrap approach is developed to generate critical values. The results are then extended to test the hypothesis of a *non relevant change point*, i.e. $H_0 : |\theta_1 - \theta_2| \leq \delta$, which reflects the fact that inference should not be changed, if the difference between the parameters before and after the change-point is small.

In contrast to previous work, our approach does neither require the mean to be constant nor - in the case of testing for lag k-correlation - that the mean, variance and fourth order joint cumulants are constant under the null hypothesis. In particular, we allow that the variance has a change point at a different location than the auto-covariance. The results are illustrated by means of a simulation study, which shows that the new procedures have nice finite sample properties. The central England monthly temperature series are analyzed and significant change points in the variance and lag 1-correlation are found in the winter monthly temperature at the late 19th century.

1   Introduction

Change point analysis is a well studied subject in the statistical literature. Since the seminal work on detecting structural breaks in the mean of Page (1954) a powerful methodology has been developed to detect various types of change points in time series [see for example Aue and Horváth (2013) and Jandhyala et al. (2013) for recent reviews of the literature]. Several authors have argued that in applications besides the mean the detection of changes in the variance or the correlation structure of a time series is of importance as well. Typical examples include the discrimination between stages of high and low asset volatility or the detection of changes in the parameters of an $AR(p)$ model in order to obtain superior forecasting procedures. Wichern et al. (1976) studied the change point problem for the variance in a first order autoregressive model. These authors pointed out that - even if log-return data exhibits a stationary behavior in the mean - the variability is often not constant and as a consequence any conclusions based on the

assumption of homoscedasticity could be misleading. Abraham and Wei (1984) and Baufays and Rasson (1985) used a Bayesian and an ML approach to find change points in AR-models. Inclán and Tiao (1994) proposed a nonparametric CUSUM-type test for changes in the variance of an independent identically distributed sequence and Lee and Park (2001) derived corresponding results applicable to linear processes [see also Chen and Gupta (1997) who used the Schwarz information criterion]. Recently Galeano and Peña (2007) and Aue et al. (2009) suggested non-parametric tests for structural breaks in the variance matrix of a multivariate time series, while Davis et al. (2006) and Preuss et al. (2014) proposed methods for detecting multiple breaks in piecewise stationary processes.

This list of references is by no means complete but an important feature of the cited references and most of the literature on testing for structural breaks consists in the fact that the model is formulated such that the stochastic process under the null hypothesis of "no change-point" is stationary. This assumption is crucial to derive (asymptotic) critical values for the corresponding testing procedures using an elegant and powerful mathematical theory such as strong approximations or invariance principles. On the other hand this assumption drastically restricts the applicability of the methodology. For example, Inclán and Tiao (1994) and Aue et al. (2009) assume for the construction of a testing procedure for the hypothesis

$$H_0: \ \sigma_i^2 = \sigma_j^2 \quad \text{for all } i, j = 1, \ldots, n \ \text{ versus } \ H_1: \ \sigma_i^2 \neq \sigma_j^2 \text{ for some } i \neq j. \tag{1.1}$$

of a constant variance of a time series that the mean of the sequence under consideration does not change in time (as the variance under the null hypothesis). A similar assumption was made by Wied et al. (2012) in the context of testing for a constant correlation, where the authors suggested a CUSUM-type statistic for a change in the correlation of a stationary time series if at the same the means and variances do not change. However, from a practical point of view, assumptions of this type are very restrictive and there might be many situations where one is interested in a change of the variance (or the correlation) even if the mean (or the means and the variances) change gradually in time. In this case the classical approach is not applicable any more. Recently, Zhou (2013) investigated such a problem, in the context of testing for a constant mean, and demonstrated that the classical CUSUM approach yields to severe biased testing results if the assumption of (weak) stationarity (under the null hypothesis) is not satisfied. The situation gets even more complicated if one is interested in more sophisticated hypotheses such as *precise hypotheses* [see Berger and Delampady (1987)]. Here (in the simplest case) one assumes the existence of a change point $k \in \{1, \ldots, n\}$ such that

$$v_1 = \sigma_1^2 = \ldots = \sigma_k^2 \ \neq \ v_2 = \sigma_{k+1}^2 = \ldots = \sigma_n^2 \,, \tag{1.2}$$

and is interested in hypotheses of the form

$$H_0 : \Delta := |v_2 - v_1| \leq \delta \ \ \text{versus} \ \ H_1 : \Delta := |v_2 - v_1| > \delta \tag{1.3}$$

for some pre-specified constant $\delta > 0$. Throughout this paper we call hypotheses of the form (1.1) *"classical"* in order to distinguish these from the precise hypotheses of the form (1.3). Although hypotheses of the form (1.3) have been discussed in other fields [see Chow and Liu (1992) and Mcbride (1999)] the problem of testing precise hypotheses has only recently been considered by Dette and Wied (2014) in the context of change point analysis. These authors point out that in many cases a modification of the statistical analysis might not be necessary if a change point has been identified but the difference between the parameters before and after the change-point is rather small. In particular, inference might be robust under "small" changes of the parameters and changing decisions (such as trading strategies or modifying a manufacturing process) might be very expensive and should therefore only be performed if changes would have serious consequences. Testing hypothesis of the form (1.3) to detect a structural break also avoids the consistency problem mentioned in Berkson (1938), that is: any test will detect negligible changes in the parameter if the sample size is sufficiently large. Dette and Wied (2014) call the hypotheses of the form (1.3) hypotheses of a *non relevant* (null hypothesis) and *relevant change point* (alternative), and according to their argumentation only relevant change points should be detected, because one has to distinguish scientific from statistical significance.

Although the formulation of the testing problem in the form (1.3) is appealing, the construction of corresponding tests faces several mathematical challenges. In particular, one has to deal with the problem of non-stationarity (even under the null hypothesis of a *non relevant change point*). For example, Dette and Wied (2014) developed a CUSUM-type test for the hypotheses in (1.3), which is only applicable under the assumption that the time series before and after the change point is strictly stationary. From a practical point this assumption seems to be very strong and not very realistic.

The present paper is devoted to the construction of change point tests for the second-order characteristics of a non-stationary time series, in particular changes in the variance and the lag k-correlation. We consider piecewise locally stationary processes as discussed by Zhou (2013), which are introduced in Section 2. Section 3 is devoted to the "classical" change point problem for the variance or lag k-correlation of a piecewise locally stationary process. We propose a CUSUM approach based on nonparametric residuals and prove weak convergence of the corresponding CUSUM statistic. It turns out that the limiting distribution depends in a complicated way on the dependence structure of the piecewise locally stationary process, and for this reason a wild bootstrap approach is developed and its consistency proved. The methodology is very general

and applicable in many situations where the assumptions of classical tests are not satisfied. For example, in the problem of testing the "classical" hypothesis of a change in the lag k-correlation we do neither assume that the mean, variance or higher order joint cumulants of the nonstationary sequence are constant nor that the change in the variance and the lag k-correlation occur at the same location. Furthermore, we discover in this paper that the stochastic errors produced in the nonparametric estimation of the mean and variance function are asymptotically negligible in the second-order CUSUM statistic. The result is of particular interest and highly non-trivial because the order of the latter nonparametric errors are larger than the $1/\sqrt{n}$ convergence rate of the CUSUM test.

Section 4 is devoted to the problem of testing the hypothesis of a non relevant change in the variance or lag k-correlation. We use the CUSUM approach proposed in Dette and Wied (2014) to obtain a test for the hypothesis (1.3) and its analogue in the case of lag k-correlations. Asymptotic normality of a corresponding $\mathcal{L}_2$-type statistic is established and a wild bootstrap method is developed, which addresses the particular structure of the hypotheses in relevant change point analysis. To our best knowledge resampling procedures for this type of change point analysis in non-stationary nonparametric problems have not been considered in the literature so far. The finite sample properties of the new procedures are investigated by means of a simulation study in Section 5. In Section 6, we analyze the central England monthly temperature series and illustrate the usefulness of the proposed methodology in identifying second order change points in climate data. Finally, all proofs and technical details are deferred to an appendix and an online supplement, respectively.

2 Piecewise locally stationary processes

We start introducing some notations, which we frequently use throughout this paper. For a (real valued) random variable X and $p \geq 1$ we denote by $\|X\|_p = (\mathbb{E}|X|^p)^{1/p}$ the $\mathcal{L}_p$ norm of X. The symbol $\overset{\mathcal{D}}{\longrightarrow}$ means weak convergence of real valued random variables (convergence in distribution). For any interval $\mathcal{I} \subset \mathbb{R}$ and nonnegative integer q define $\mathcal{C}^q(\mathcal{I})$ as the set of q times continuously differentiable functions $f : \mathcal{I} \to \mathbb{R}$ and $\mathcal{C}(\mathcal{I}) = \mathcal{C}^0(\mathcal{I})$. Let $\{\varepsilon_i\}_{i \in \mathbb{Z}}$ denote a sequence of independent identically distributed (i.i.d.) random variables and denote by $\mathcal{F}_i = \sigma(..., \varepsilon_0, ..., \varepsilon_{i-1}, \varepsilon_i)$ the sigma field generated by $\{\varepsilon_j | j \leq i\}$. We define the sigma field $\mathcal{F}_i^{(j)} = \sigma(..., \varepsilon_{j-1}, \varepsilon_j', \varepsilon_{j+1}..., \varepsilon_i)$, where $\{\varepsilon_i'\}_{i \in \mathbb{Z}}$ is an independent copy of $\{\varepsilon_i\}_{i \in \mathbb{Z}}$, and $\mathcal{F}_i^* = \mathcal{F}_i^{(0)}$ for short. In the following discussion we will also make frequent use of the projection operator $\mathcal{P}_j(\cdot) = \mathbb{E}(\cdot | \mathcal{F}_j) - \mathbb{E}(\cdot | \mathcal{F}_{j-1})$.

In this paper, we consider the model

$$Y_i = \mu(t_i) + e_i, \qquad i = 1, \ldots, n, \qquad (2.1)$$

where (for the sake of simplicity) $t_i = i/n$ $(i = 1, \ldots, n)$ and μ is a smooth function. Note that formally $\{Y_i\}_{i=1}^n$ is a triangular array of random variables but we do not reflect this fact in our notation. Change point problems for this model have found considerable attention in the recent literature, where most of the work refers to problems of detecting a gradual change of the mean in the situation of zero mean and independent identically distributed (i.i.d.) errors (even assumed to be Gaussian in some cases) [see Müller (1992) for an early reference and Mallik et al. (2011) and Mallik et al. (2013) for more recent references]. Recently Vogt and Dette (2015) proposed a generalized CUSUM approach to detect gradual changes in model (2.1) using a different concept of local stationarity [see Vogt (2012)].

In the present paper we consider non-stationary processes of the form (2.1) and are interested in identifying abrupt changes in the second order properties such as the variance or the correlation at a given lag. More precisely we consider an error process $\{e_i\}_{i=1}^n$ in (2.1), which is piecewise locally stationary (PLS) with r breaks for some $r \in \mathbb{N}$. Formally, we use the following definition for a *PLS* process and the concept of "*physical dependence measure for PLS*", which is given in Zhou (2013).

Definition 2.1.
(1) The sequence $\{e_i\}_{i=1}^n$ is called PLS with r break points if there exist constants $0 = b_0 < b_1 < \ldots < b_r < b_{r+1} = 1$ and nonlinear filters $G_0, G_1, \ldots, G_r$, such that

$$e_i = G_j(t_i, \mathcal{F}_i), \;\; if \; b_j < t_i \leq b_{j+1},$$

where $\mathcal{F}_i = \sigma(\ldots, \varepsilon_0, \ldots, \varepsilon_{i-1}, \varepsilon_i)$, and $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ is a sequence of i.i.d. random variables.
(2) Assume that $\max_{1\leq i\leq n} \|e_i\|_p < \infty$ for some $p \geq 1$. Then for $k > 0$, define the kth physical dependence measure in $\mathcal{L}_p$-norm as

$$\delta_p(k) = \max_{0\leq i\leq r} \sup_{b_i < t \leq b_{i+1}} \|G_i(t, \mathcal{F}_k) - G_i(t, \mathcal{F}_k^*)\|_p,$$

where $\delta_p(k) = 0$ if $k < 0$.

For the asymptotic analysis presented later in this paper we list the following conditions:

(A1) The process $\{e_i\}_{i=1}^n$ is PLS and *piecewise stochastic Lipschitz continuous*. This means that there exists a constant $C > 0$, such that for all $i \in \{0, \ldots, r\}$ and all $t, s \in [b_i, b_{i+1}]$

the condition

$$\|G_i(t, \mathcal{F}_0) - G_i(s, \mathcal{F}_0)\|_\iota / (t-s) \leq C(t-s)$$

holds, where $\iota \geq 8$ and C denotes a positive constant. In addition, $\mathbb{E}[e_i] = 0$ for all $1 \leq i \leq n$, and we assume the existence of a strictly positive *variance function* $\sigma^2 : [0,1] \to \mathbb{R}^+$, such that $\sigma_i^2 := \sigma^2(t_i) = \operatorname{Var}(e_i)$ $(i = 1, \ldots, n)$.

(A2) The second derivative $\ddot{\mu}$ of the function μ in model (2.1) exists and is Lipschitz continuous on the interval $[0, 1]$.

(A3) $\max_{0 \leq i \leq r} \sup_{t \in (b_i, b_{i+1}]} \|G_i(t, \mathcal{F}_0)\|_\iota < \infty$ for some $\iota \geq 8$.

(A4) $\delta_\iota(k) = O(\chi^k)$ for some $\chi \in (0, 1)$ and some $\iota \geq 8$.

Remark 2.1.

a) We emphasize that the bound of $\max_{1 \leq i \leq n} \|e_i\|_p$ in Definition 2.1 does not depend on n. This assumption is made in order to simplify the assumptions and the proofs in the subsequent discussion. It is also possible to develop corresponding results for an n-dependent bound with an additional complication in the technical arguments of the proofs and in the assumptions.

b) Note that the process $\{e_i^2\}_{i=1}^n$ of squared errors is also PLS. Simple calculations show that $\{e_i^2\}_{i=n}^n$ satisfies the assumptions (A1), (A3), (A4) with $\iota \geq 4$.

3 Tests for changes in the second order structure

Suppose that we observe data $\{Y_i\}_{i=1}^n$ according to model (2.1), where the process $\{e_i\}_{i=1}^n$ is PLS and μ is an unknown deterministic trend. We are interested in testing nonparametrically the "classical" hypothesis of a change point in the variance or the lag k-correlation. The important difference to previous work on this subject [see for example Inclán and Tiao (1994) or Aue et al. (2009)] is that in general the process is NOT assumed to be stationary under the null hypothesis of no change point. This means - for example - that the approach proposed here can be used to test the hypotheses (1.1), where the mean is not constant. The price for this type of flexibility is that critical values of the asymptotic distribution of the CUSUM statistic are not directly available. For the solution of this problem we will develop a bootstrap CUSUM-type test for the "classical" hypotheses of a change point in the variance or lag k-correlation, which is based on residuals from a local linear fit. For the definition of the local linear estimator we assume throughout this paper that the corresponding kernel function, say K, is symmetric with support $[-1, 1]$ satisfying $\int K(x)dx = 1$, and define for $b > 0$ the function $K_b(\cdot) = K(\frac{\cdot}{b})$. The moments of

K and K^2 are denoted by $\mu_l = \int_{\mathbb{R}} x^l K(x) dx$ and $\phi_l = \int_{\mathbb{R}} x^l K^2(x) dx$, respectively $(l = 0, 1, \dots)$. We also assume that $K \in \mathcal{C}^{(2)}([-1, 1])$.

3.1 Change point tests for the variance

Our first goal is to investigate the stability of the variances $\sigma_i^2 = \sigma^2(t_i) = \text{Var}(Y_i)$ $(i = 1, \dots, n)$ testing nonparametrically the "classical" hypotheses (1.1). For this purpose we consider the CUSUM statistic

$$\hat{T}_n = \max_{1 \le i \le n} \left| \hat{S}_i - \frac{i}{n} \hat{S}_n \right|, \tag{3.1}$$

where $\hat{S}_i = \sum_{j=1}^{i} \hat{e}_j^2$ denotes the sum of squared nonparametric residuals $\hat{e}_i = Y_i - \hat{\mu}_{b_n}(t_i)$, and $\hat{\mu}_{b_n}(\cdot)$ is the local linear estimator of the function μ with bandwidth b_n, that is

$$(\hat{\mu}_{b_n}(t), \hat{\dot{\mu}}_{b_n}(t)) = \operatorname*{argmin}_{b_0, b_1} \sum_{i=1}^{n} \left(Y_i - b_0 - b_1(t_i - t) \right)^2 K_{b_n}(t_i - t) \tag{3.2}$$

[see Fan and Gijbels (1996)]. Weak convergence of the statistic $\hat{T}_n/\sqrt{n}$ follows under the additional assumption

(A5) The long run variance function

$$\kappa_1^2(t) = \sum_{k=-\infty}^{\infty} \text{cov}(G_i^2(t, \mathcal{F}_k), G_i^2(t, \mathcal{F}_0)) \qquad \text{if } t \in (b_i, b_{i+1}], \tag{3.3}$$

exists, $\kappa_1^2(0) := \lim_{t\downarrow 0} \kappa_1^2(t)$ exists and $\inf_{t \in [0,1]} \kappa_1^2(t) > 0$.

The following result provides the asymptotic distribution of $\hat{T}_n$. Its proof is complicated and therefore deferred to Section 7.1.1 in the Appendix.

Theorem 3.1. *If assumptions (A1)-(A5) are satisfied and $nb_n^6 \to 0$, $nb_n^3 \to \infty$, then under the null hypothesis of no change in the variance we have*

$$\frac{1}{\sqrt{n}} \hat{T}_n \xrightarrow{\mathcal{D}} \mathcal{K}_1 := \sup_{t \in (0,1)} |U_1(t) - tU_1(1)|, \tag{3.4}$$

where $\{U_1(t)\}_{t \in [0,1]}$ is a zero mean Gaussian process with covariance function

$$\gamma(t, s) = \int_0^{\min(t,s)} \kappa_1^2(r) dr. \tag{3.5}$$

It follows from the proof of Theorem 3.1 that the statistic $\frac{1}{\sqrt{n}}\hat{T}_n$ has the same limit distribution as the statistic which is obtained if the nonparametric residuals $\hat{e}_i$ are replaced by the "true" errors e_i from model (2.1). This observation is remarkable and highly non-trivial because the error from the nonparametric estimation is larger than $1/\sqrt{n}$.

As a consequence of Theorem 3.1, we obtain - in principle - an asymptotic level α test for the hypothesis (1.1) by rejecting H_0, whenever $\frac{1}{\sqrt{n}}\hat{T}_n > q_{1-\alpha}$, where $q_{1-\alpha}$ is the $(1-\alpha)$-quantile of the distribution of the random variable $\mathcal{K}_1$ in (3.4). However, under non-stationarity (more precisely under the PLS assumption), the function κ_1^2 defined in (3.3) and, as a consequence, the covariance structure of the Gaussian process $\{U_1(t) - tU_1(1)\}_{t\in[0,1]}$ involves the complex dependency structure of the data generating process. Therefore it is very difficult to estimate the critical value $q_{1-\alpha}$ of the asymptotic distribution of the CUSUM test statistic directly. As an alternative, a data-driven critical value will be derived in the following discussion using a wild bootstrap method to mimic the distributional properties of the Gaussian process $U_1(\cdot)$. Following Zhou (2013) we define for a fixed window size, say m, the quantity

$$\hat{\Phi}_{i,m} = \frac{1}{\sqrt{m(n-m+1)}}\sum_{j=1}^{i}\left(\hat{S}_{j,m} - \frac{m}{n}\hat{S}_n\right)R_j, \qquad i = 1, ..., n-m+1, \qquad (3.6)$$

where $\hat{S}_{j,m} = \sum_{r=j}^{j+m-1}\hat{e}_r^2$ and $\{R_i\}_{i\in\mathbb{Z}}$ is a sequence of i.i.d standard normal distributed random variables, which is independent of $\{\varepsilon_i\}_{i\in\mathbb{Z}}$.

Theorem 3.2. *Suppose that the conditions of Theorem 3.1 are satisfied. In addition, assume that $m \to \infty$, $m/n \to 0$, $mb_n^4\log^2 n \to 0$ and $\frac{m\log^2 n}{nb_n^{3/2}} \to 0$. Then*

$$M_n = \max_{m+1\leq i\leq n-m+1}\left|\hat{\Phi}_{i,m} - \frac{i}{n-m+1}\hat{\Phi}_{n-m+1,m}\right| \xrightarrow{\mathcal{D}} \mathcal{K}_1$$

conditional on $\mathcal{F}_n$, where the random variable $\mathcal{K}_1$ is defined in (3.4).

Theorem 3.2 provides an asymptotic level α test for the hypothesis of a constant variance in model (2.1), where the critical values are obtained by resampling. The details are summarized in the following algorithm. Some illustrations of this method are given in Section 5 and 6.

Algorithm 3.1.
[1] Calculate the statistic $\hat{T}_n$ defined in (3.1).
[2] Generate B conditionally *i.i.d* copies $\{\hat{\Phi}_{i,m}^{(r)}\}_{i=1}^{n-m+1}$ $(r = 1, \ldots, B)$ of the random variables $\{\hat{\Phi}_{i,m}\}_{i=1}^{n-m+1}$ defined in (3.6) and calculate

$$M_r = \max_{m+1\leq i\leq n-m+1}\left|\hat{\Phi}_{i,m}^{(r)} - \frac{i}{n-m+1}\hat{\Phi}_{n-m+1,m}^{(r)}\right|.$$

9

[3] Let $M_{(1)} \leq M_{(2)} \leq \ldots \leq M_{(B)}$ denote the order statistics of $M_1, \ldots, M_B$. The null hypothesis of constant variance is rejected at level α, whenever

$$\hat{T}_n/\sqrt{n} > M_{\lfloor B(1-\alpha) \rfloor}. \tag{3.7}$$

The p-value of this test is given by $1 - \frac{B^*}{B}$, where $B^* = \max\{r : M_{(r)} \leq \hat{T}_n/\sqrt{n}\}$.

Remark 3.1. it follows by similar arguments as given in the proof of Theorem 2, Proposition 3 of Zhou (2013), and Lemma 8.1 and Lemma 8.2 in the appendix of this paper that the bootstrap test (3.7) is consistent. In fact the test is able to detect local alternatives of the form $\sigma^2(\cdot) = \sigma^2 + n^{-1/2} f(\cdot)$, where $f(\cdot)$ is a nonconstant piecewise Lipschitz continuous function.

3.2 Changes in the correlation

In this section we consider the problem of testing whether there are changes in the correlation $\rho_{i,k} := corr(Y_i, Y_{i+k})$ for some pre-specified lag k. Namely, we are interested in testing the hypotheses

$$H_0 : \ \rho_{i,k} = \rho_{j,k} \ \text{ for all } i, j = 1, \ldots, n \ \text{ versus } H_1 : \ \rho_{i,k} \neq \rho_{j,k} \text{ for some } i \neq j. \tag{3.8}$$

A test for the classical hypothesis $H_0 : \rho_1 = \rho_2$ in stationary processes can be derived by similar arguments as given in Wied et al. (2012) under the additional assumption that the mean and variance are not changing. However, statistical inference regarding changes the correlation structure in a general local stationary framework (including non constant mean or variance) requires estimates of the covariances and variances. For this purpose we consider two different estimators. First, let $\hat{\sigma}^2(t_i) = \hat{\sigma}^2_{c_n, b_n}(t_i)$ denote the local linear estimates of $\sigma_i^2 = \sigma^2(t_i) = \text{Var}(Y_i)$, which is defined as

$$(\hat{\sigma}^2_{c_n, b_n}(t), \widehat{\dot{\sigma}^2}_{c_n, b_n}(t)) = \underset{b_0, b_1}{\text{argmin}} \sum_{i=1}^{n} (\hat{e}_i^2 - b_0 - b_1(t_i - t)) K_{c_n}(t_i - t), \tag{3.9}$$

where $\hat{e}_i = Y_i - \hat{\mu}_{b_n}(t_i)$ denote the residuals obtained from a fit of the local linear estimate with bandwidth b_n. This estimate is appropriate if there is no structural break in the variance.
If this situation cannot be excluded, a more refined estimate for the function σ^2 is required. To be precise, we also allow the variance to have a structural break at a point, say $\tilde{t}_v$, which does not necessarily coincide with the location of the change point in the lag k-correlation. We assume that $\ddot{\sigma}^2$ is Lipschitz continuous on the intervals $(0, \tilde{t}_v)$ and $(\tilde{t}_v, 1)$ and that there exists a constant $\zeta > 0$, such that $\tilde{t}_v \in [\zeta, 1 - \zeta]$. We define an estimator, say t_n^*, of the change point $\tilde{t}_v$

10

in the variance by

$$t_n^* = \operatorname*{argmax}_{\lfloor n\zeta \rfloor \leq i \leq n - \lfloor n\zeta \rfloor + 1} |\mathcal{M}(i)|, \tag{3.10}$$

where

$$\mathcal{M}(i) = \frac{1}{L}\Big(\sum_{j=i-L+1}^{i} \hat{e}_j^2 - \sum_{j=i}^{i+L-1} \hat{e}_j^2 \Big) \tag{3.11}$$

and $L \in \mathbb{N}$ is a regularization parameter, which increases with n. Note that the maximum in (3.10) is not taken over the full range $1 \leq i \leq n$ as recommended in Andrews (1993) [see also Qu (2008)]. Finally, the second estimator of the variance function is defined by

$$\hat{\sigma}_n^{2*}(s) = \tilde{\sigma}_1^2(s)I(s \leq t_n^*) + \tilde{\sigma}_2^2(s)I(s > t_n^*) , \tag{3.12}$$

where $\tilde{\sigma}_1^2$ and $\tilde{\sigma}_2^2$ are the local linear estimator of the variance function from the samples $Y_1, \ldots, Y_{\lfloor nt_n^* \rfloor}$ and $Y_{\lfloor nt_n^* \rfloor + 1}, \ldots, Y_n$, respectively. The following result shows that t_n^* is a consistent estimate of $\tilde{t}_v$. A proof can be found in Section 7.1.

Lemma 3.1. *Assume that $nb_n^6 \to 0$, $nb_n^3 \to \infty$ and that Assumptions (A1) - (A4) are satisfied with $\iota > 8$. Suppose that the variance function is twice differentiable on the intervals $(0, \tilde{t}_v)$ and $(\tilde{t}_v, 1)$, such that the second derivative $\ddot{\sigma}^2$ is Lipschitz continuous (here $\tilde{t}_v$ is the location of the change point of the variance, which is defined as $\tilde{t}_v = 1$ if there exists no jump). Let $L = \lfloor n^\alpha \rfloor$ for $4/\iota < \alpha < 1/2$, then the estimator t_n^* defined in (3.10) satisfies $t_n^* - \tilde{t}_v = O_p(n^{-(1-\alpha)})$ if $\tilde{t}_v < 1$.*

Remark 3.2. Observe that the lower bound for the parameter α in Lemma 3.1 converges to 0 as $\iota \to \infty$. Consequently, if $0 < \tilde{t}_v < 1$, the rate of convergence of the estimator t_n^* is arbitrarily close to the optimal rate n^{-1} if Assumptions (A1) and (A4) hold for any $\iota > 0$.

The lag k-correlation at the point t_i is estimated by a local average of the quantities of the form

$$\hat{W}_i^k = \frac{\hat{e}_i \hat{e}_{i+k}}{\hat{\sigma}^2(t_i)}, \tag{3.13}$$

where $\hat{\sigma}^2$ is either the estimate $\hat{\sigma}_{c_n, b_n}^2$ defined in (3.9) (if a change point in the variance can be excluded) or the estimate $\hat{\sigma}_n^{2*}$ defined in (3.12). For convenience, we set $\hat{e}_i = 0$, whenever $i > n$, in the following discussion. The corresponding partial sum is denoted by $\hat{S}_i^W = \sum_{j=1}^{i} \hat{W}_j^k$, and

11

we consider the CUSUM statistic

$$\hat{T}_n^c = \max_{1 \le i \le n} \left| \hat{S}_i^W - \frac{i}{n} \hat{S}_n^W \right|. \tag{3.14}$$

In order to specify the necessary assumptions for the asymptotic theory, recall the definition of the random variable $\hat{W}_i^k$ in (3.13) and define an unobservable analogue by

$$W_i^k = \frac{e_i e_{i+k}}{\sigma(t_i)\sigma(t_{i+k})}. \tag{3.15}$$

It is easy to see that W_i^k is $\mathcal{F}_{i+k}$ measurable and that the process $(W_i^k)_{i=0}^{n-k}$ is PLS. Define q as the number of break points, $0 = c_0 < c_1 < ... < c_q = 1$ as the corresponding locations of the breaks and $H_0, H_1, \ldots, H_q$ as the corresponding nonlinear filters, that is $W_i^k = H_j(t_i, \mathcal{F}_{i+k})$ if $c_j < t_i \le c_{j+1}$. For the asymptotic analysis of the CUSUM-test we require the following additional assumption:

(A6) The long run variance function

$$\kappa_2^2(t) = \sum_{k=-\infty}^{\infty} \text{cov}(H_i(t, \mathcal{F}_k), H_i(t, \mathcal{F}_0)) \qquad \text{if } t \in (c_i, c_{i+1}],$$

exists, the limit $\kappa_2^2(0) = \lim_{t \downarrow 0} \kappa_2^2(t)$ exists and $\inf_{t \in [0,1]} \kappa_2^2(t) > 0$.

Theorem 3.3. *Assume that $b_n \to 0$, $c_n/b_n \to 0$, $c_n b_n^{-2} \to \infty$, $nc_n^4 \to 0$, $nb_n^6 c_n^{-1/2} \to 0$, $nb_n^4 c_n^{1/2} \to \infty$ and suppose that Assumptions (A1) - (A4) and (A6) are satisfied with $\iota \ge 8$. Assume that $\hat{\sigma}^2$ in (3.13) is either the estimate $\hat{\sigma}_{c_n,b_n}^2$ defined in (3.9), if there is no change point, or the estimate $\hat{\sigma}_n^{2*}$ defined in (3.12) if there exists one change point, say $\tilde{t}_v$, in the variance function. In the latter case let σ^2 be strictly positive and twice differentiable on the intervals $(0, \tilde{t}_v)$ and $(\tilde{t}_v, 1)$, such that the second derivative $\ddot{\sigma}^2$ is Lipschitz continuous. Then under the null hypothesis of no change point in the lag k-correlation we have*

$$\frac{1}{\sqrt{n}} \hat{T}_n^c \xrightarrow{\mathcal{D}} \mathcal{K}_2 := \sup_{t \in (0,1)} |U_2(t) - tU_2(1)|,$$

where $\{U_2(t) \mid t \in [0,1]\}$ is a centered Gaussian process with covariance kernel

$$\gamma(t,s) = \int_0^{\min(t,s)} \kappa_2^2(r)dr. \tag{3.16}$$

In order to develop a consistent bootstrap test recall the notation (3.4), define $S_{j,m}^W = \sum_{r=j}^{j+m-1} \hat{W}_r^k$

and $\hat{\Phi}_{i,m}^W$ as in (3.6) where $\hat{S}_n$ and $\hat{S}_{j,m}$ are replaced by $\hat{S}_n^W$ and $\hat{S}_{j,m}^W$, respectively.

Theorem 3.4. *Assume that the conditions of Theorem 3.3 are satisfied. If $m \to \infty$, $m/n \to 0$, $\sqrt{m}\left(c_n^2 + \left(\frac{1}{\sqrt{nc_n}} + b_n^2 + \frac{1}{\sqrt{nb_n}}\right)c_n^{-1/4}\right)\log n \to 0$, then*

$$M_n^W = \max_{m+1 \le i \le n-m+1}\left|\hat{\Phi}_{i,m}^W - \frac{i}{n-m+1}\hat{\Phi}_{n-m+1,m}^W\right| \xrightarrow{\mathcal{D}} \mathcal{K}_2 \qquad (3.17)$$

conditional on $\mathcal{F}_n$, where the random variable $\mathcal{K}_2$ is defined in Theorem 3.3.

Theorem 3.3 and 3.4 yield the consistency of the following bootstrap test for the hypothesis of a constant correlation with asymptotically correct type I error rate. The null hypothesis (3.8) is rejected whenever

$$\hat{T}_n^c/\sqrt{n} > M_{\lfloor B(1-\alpha)\rfloor}^W. \qquad (3.18)$$

Here $M_{\lfloor B(1-\alpha)\rfloor}^W$ is the $(1-\alpha)$-quantile of bootstrap sample of the distribution of the statistic M_n^W defined in (3.17), which is generated in the same way as described in Algorithm 3.1.

Remark 3.3.
(1) Assume that under the alternative hypothesis the variance has at most one structural break at the point $\tilde{t}_v$ and that the second derivative $\ddot{\sigma}^2$ of the variance function is Lipschitz continuous on the intervals $(0, \tilde{t}_v)$ and $(\tilde{t}_v, 1)$. Then it can be shown by similar arguments as indicated in Remark 3.1 that the bootstrap test (3.18) is able to detect local alternatives converging to the null hypothesis at a rate $n^{-1/2}$.
(2) In principle one could always work with the estimator $\hat{\sigma}_n^{2*}$ defined in (3.12). However, a numerical study indicates that in cases where there is in fact no change point in the variance the estimator $\hat{\sigma}_{c_n,b_n}^2$ defined in (3.9) has a better finite sample performance. Therefore we strictly recommend its use, if a structural change in the variance can be excluded.

Remark 3.4. Although the method described so far refers to the problem of detecting a change point in a particular lag k-correlation, it can easily be extended to the problem of testing for a change in one (or more) correlations simultaneously. We illustrate extensions of this type exemplarily for the problem of testing the hypotheses

$$H_0 : \rho_{i,k} = \rho_{j,k} = \rho_k \quad \text{for all } i, j = 1, \dots, n, k = 1, \dots, q \qquad (3.19)$$

$$H_1 : \text{there exists } 1 \le k \le q \text{ and } i \ne j \text{ such that } \rho_{i,k} \ne \rho_{j,k}. \qquad (3.20)$$

For this purpose recall the notations (3.15) and (3.13) and define the vectors $\mathbf{W}_i = (W_i^1, \dots, W_i^q)^T$

and $\hat{\mathbf{W}}_i = (\hat{W}_i^1, \ldots, \hat{W}_i^q)^T$. For some norm $\| \cdot \|$ on $\mathbb{R}^q$ we consider the statistic

$$\hat{\mathbf{T}}_n^c = \max_{1 \le j \le n-q} \left\| \sum_{i=1}^{j} \hat{\mathbf{W}}_i - \frac{1}{n} \sum_{i=1}^{n-q} \hat{\mathbf{W}}_i \right\|.$$

Observe that $\{\mathbf{W}_i\}_{i \in \mathbb{N}}$ is a PLS process, with nonlinear filter function, say $\tilde{\mathbf{W}}$, and (unknown) ℓ break points $0 = d_0 < d_1 < \ldots, < d_\ell < d_{\ell+1} = 1$. Assume that the conditions of Theorem 3.3 are satisfied, where assumption (A6) is now replaced by the condition

(A6*) The long run variance function

$$\tilde{\kappa}_2^2(t) = \sum_{k=-\infty}^{\infty} \mathrm{Cov}\ (\tilde{\mathbf{W}}_i(t, \mathcal{F}_k), \tilde{\mathbf{W}}_i(t, \mathcal{F}_0)) \in \mathbb{R}^{q \times q} \qquad \text{if } t \in (d_i, d_{i+1}],$$

exists. Let $\lambda(t) = \lambda_{\min}(\tilde{\kappa}_2^2(t))$ denote the smallest eigenvalue of the matric $\tilde{\kappa}^2(t)$, then $\lim_{t \downarrow 0} \lambda(t)$ exists and $\inf_{t \in [0,1]} \lambda(t) > 0$.

Under these assumptions it can be shown by similar arguments as given in the proof of Theorem 3.3 that

$$\frac{1}{\sqrt{n}} \hat{\mathbf{T}}_n^c \xrightarrow{\mathcal{D}} \sup_{t \in (0,1)} \|\mathbf{U}_2(t) - t\mathbf{U}_2(1)\|,$$

where $\{\mathbf{U}_2(t) | t \in [0,1]\}$ is a q-dimensional centered Gaussian process covariance kernel $\gamma(t, s) = \int_0^{\min(t,s)} \tilde{\kappa}_2^2(r) dr \in \mathbb{R}^{q \times q}$.

A similar result can also be derived for the new bootstrap procedure. To be precise define the vectors

$$\hat{\boldsymbol{\Phi}}_{i,m} = \frac{1}{\sqrt{m(n-m-2q+1)}} \sum_{j=q+1}^{i} \left(\hat{\mathbf{S}}_{j,m} - \frac{m}{n} \hat{\mathbf{S}}_{n-q} \right) R_j, \qquad i = q+1, \ldots, n-m-q+1,$$

(3.21)

where $\hat{\mathbf{S}}_{j,m} = \sum_{r=j}^{j+m-1} \hat{\mathbf{W}}_r$, $\hat{\mathbf{S}}_{n-q} = \sum_{r=1}^{n-q} \hat{\mathbf{W}}_r$ and $\{R_i\}_{i \in \mathbb{Z}}$ is a sequence of i.i.d standard normal distributed random variables, which is independent of $\{\varepsilon_i\}_{i \in \mathbb{Z}}$. If the conditions of Theorem 3.4 hold [where assumption (A6) is again replaced by (A6*)], then we have (conditional on $\mathcal{F}_n$)

$$M_n = \max_{q+1 \le i \le n-m-q+1} \left\| \hat{\boldsymbol{\Phi}}_{i,m} - \frac{i}{n-m-2q+1} \hat{\boldsymbol{\Phi}}_{n-m-q+1,m} \right\| \xrightarrow{\mathcal{D}} \sup_{t \in (0,1)} \|\mathbf{U}_2(t) - t\mathbf{U}_2(1)\|.$$

These results show that the bootstrap test (3.18) can easily be extended to discriminate between

the hypotheses (3.19) and (3.20).

4 Relevant changes of second order characteristics

After a change point has been detected and localized a modification of the statistical analysis is necessary, which addresses the different features of the data generating process before and after the change point. Dette and Wied (2014) pointed out that in many cases such a modification might not be necessary if the difference between the parameters before and after the change point is rather small. On the one hand, inference might be robust with respect to small changes of the variance or correlation structure. On the other hand, changing decisions (such as trading strategies or modifying a manufacturing process) might be very expensive and only be performed if changes would have serious consequences. For these reasons Dette and Wied (2014) proposed to investigate the hypothesis (1.3) of a *non relevant change point*, which will be discussed in this section for the variance (Section 4.1) and lag k-correlation (Section 4.2) in a general non-stationary context (more precisely under the assumption of PLS).

4.1 Relevant changes in the variance

First we investigate the problem of testing for a non relevant change in the variance of a time series. Recall the definition of model (2.1) and that $\sigma_i^2 = \sigma^2(t_i) = \mathrm{Var}(Y_i)$ is the variance of the response at the point t_i. Throughout this section we assume the existence of some fixed but unknown point $\tilde{t}_v \in [0, 1]$ such that the variance function in Assumption (A1) is constant on the intervals $(0, \tilde{t}_v)$ and $(\tilde{t}_v, 1)$ and that the variances σ_i^2 satisfy (1.2) with $k = \lfloor n\tilde{t}_v \rfloor$. For some pre-specified $\delta > 0$, we are interested in testing the hypothesis (1.3) of a non relevant change in the variance. Problems of this type have recently been discussed in Dette and Wied (2014) under the assumption that the process before and after the change point is stationary and that additionally the mean of the process is constant (even if there is a change of small order in the variance). From a practical point of view, assumptions of this type are not very satisfactory and in this section we will propose a procedure for detecting relevant changes which does not require such strong assumptions.

It turns out that a test for the hypothesis (1.3) needs an estimator of the change point, and we could use the estimator t_n^* defined in (3.10) for this purpose. This estimator requires the choice of the regularization parameter L. However, in the present context such a complicated estimator is in fact not necessary, because the variance before and after the change point is assumed to be

15

constant. Therefore we introduce the alternative estimator

$$\tilde{t}_n = \underset{1 \leq m \leq n}{\operatorname{argmax}} \left(\hat{S}_m - \frac{m}{n} \hat{S}_n \right)^2, \tag{4.1}$$

where $\hat{S}_m = \sum_{i=1}^m \hat{e}_i^2$ denotes the mth partial sum of the squared residuals $\hat{e}_i = Y_i - \hat{\mu}_{b_n}(t_i)$ obtained from a local linear fit. Estimators, maximizing the CUSUM-type statistics have been widely studied in the situation of stationary processes [see Jandhyala et al. (2013)], and it turns out that the statistic $\tilde{t}_n$ has better finite sample properties than the estimator t_n^* defined in (3.10), if the variance function before and after the break point $\tilde{t}_v$ is in fact constant (which is our basic assumption throughout this section).

Let $\Delta = v_2 - v_1$ denote the "true" difference before and after the change point. Our first result establishes the asymptotic properties of the estimator $\tilde{t}_n$ under the PLS assumption and is proved in Section 7.2.

Lemma 4.1. *Assume that $b_n \to 0$, $nb_n^3 \to \infty$, $nb_n^6 \to 0$ and that Assumptions (A1) - (A5) are satisfied. Suppose that the variance function is strictly positive and constant on the intervals $(0, \tilde{t}_v)$ and $(\tilde{t}_v, 1)$ (here $\tilde{t}_v$ is the location of the change point, which is defined as $\tilde{t}_v = 1$ if there is no jump). The estimate $\tilde{t}_n$ defined in (4.1) has the following properties:*

(i) If $\Delta = 0$, then $\tilde{t}_n$ converges weakly to a $[0, 1]$-valued random variable.

(ii) If $\Delta \neq 0$, then $|\tilde{t}_n - \tilde{t}_v| = O_p(n^{-\alpha})$ for some $\alpha > 1/2$.

We now use the statistic (4.1) to define estimates of the variance before and after the change point, that is

$$\tilde{\Delta}_{n,1} = \frac{1}{\lfloor n\tilde{t}_n \rfloor} \sum_{j=1}^{\lfloor n\tilde{t}_n \rfloor} \hat{e}_j^2, \quad \tilde{\Delta}_{n,2} = \frac{1}{n - \lfloor n\tilde{t}_n \rfloor} \sum_{j=\lfloor n\tilde{t}_n \rfloor + 1}^{n} \hat{e}_j^2,$$

and denote by $\tilde{\Delta}_n = \tilde{\Delta}_{n,2} - \tilde{\Delta}_{n,1}$ an estimator of the difference Δ. Using similar arguments as given in the proof of Lemma 4.3 in Section 7.2 it follows that $\tilde{\Delta}_n - \Delta = O_p(\frac{\log n}{\sqrt{n}})$. In order to construct a test for the hypothesis (1.3) we now consider the statistic

$$\check{T}_n = \frac{3}{\tilde{t}_n^2(1 - \tilde{t}_n)^2} \int_0^1 \hat{U}_n^2(s) ds, \tag{4.2}$$

where the process $\hat{U}_n$ is the CUSUM process defined by

$$\hat{U}_n(s) = \frac{1}{n} \sum_{j=1}^{\lfloor ns \rfloor} \hat{e}_j^2 - \frac{s}{n} \sum_{j=1}^{n} \hat{e}_j^2.$$

16

The following result establishes the asymptotic properties of the statistic $\check{T}_n$. The proof is omitted because it follows by similar but easier arguments as given in the proof of Theorem 4.3 below, where we investigate the problem of testing for non relevant changes in the lag k-correlation.

Theorem 4.1. *Suppose that the conditions of Theorem 3.1 hold.*

(i) *If $\Delta \neq 0$, then*

$$\sqrt{n}(\check{T}_n - \Delta^2) \xrightarrow{\mathcal{D}} \mathcal{Z}_1(\Delta) := \frac{6}{\tilde{t}_v^2(1 - \tilde{t}_v)^2} \int_0^1 (U_1(s) - sU_1(1))[s\tilde{t}_v - s \wedge \tilde{t}_v]|\Delta|ds, \quad (4.3)$$

where $\{U_1(s)\}_{s \in [0,1]}$ denotes the Gaussian process defined in Theorem 3.1.

(ii) *If $\Delta = 0$, then $\check{T}_n = O_P(1/n)$.*

It follows from Theorem 4.1 that an asymptotic level α test for the hypothesis (1.3) of a non relevant change in the variance is obtained by rejecting the null hypothesis, whenever

$$\check{T}_n > \delta^2 + \frac{v_{1-\alpha}}{\sqrt{n}}, \quad (4.4)$$

where $v_{1-\alpha}$ denotes the $(1-\alpha)$-quantile of the distribution of the random variable $\mathcal{Z}_1(\delta)$ defined in (4.3). We note that this distribution is a centered normal distribution with a variance depending on the data generating process in a complicated way, in particular on the long run variance defined in (3.3). In order to circumvent this problem we will develop a resampling procedure to obtain critical values, where we have to address the particular structure of the hypothesis (1.3) of a non relevant change point in the variance. To be precise, define

$$\hat{V}_j = \hat{e}_j^2 - \hat{\Delta}_n I(j \geq \lfloor n\tilde{t}_n \rfloor). \quad (4.5)$$

Let $\{R_j\}_{j \in \mathbb{Z}}$ denote a sequence of i.i.d. standard normal distributed random variables, which is independent of $\{\mathcal{F}_i\}_{i \in \mathbb{Z}}$, define $\hat{S}_{j,m}^V = \sum_{r=j}^{j+m-1} \hat{V}_r$, $\hat{S}_n^V = \sum_{r=1}^n \hat{V}_r$ and consider the random variables

$$\hat{\Phi}_{i,m}^V = \frac{1}{\sqrt{m(n-m+1)}} \sum_{j=1}^{n-m+1} \left(\hat{S}_{j,m}^V - \frac{m}{n} \hat{S}_n^V \right) R_j. \quad (4.6)$$

The following result provides a bootstrap approximation for the distribution of the random variable $\mathcal{Z}_1(1)$. The proof follows by similar but easier arguments as given in Section 7, where a corresponding statement is proved for change point tests in the correlation structure.

Theorem 4.2. *Assume that the conditions of Theorem 4.1 are satisfied. In addition, assume that $m \to \infty$, $m/n \to 0$, $mb_n^4 \log^2 n \to 0$ and $\frac{m \log^2 n}{n b_n^{3/2}} \to 0$, then*

$$M_n = \frac{1}{n} \frac{6}{\tilde{t}_n^2(1 - \tilde{t}_n)^2} \sum_{m+1 \le i \le n-m+1} \left(\hat{\Phi}_{i,m}^V - \frac{i}{n-m+1} \hat{\Phi}_{n-m+1,m}^V \right) \left(\frac{i\tilde{t}_n}{n} - \frac{i}{n} \wedge \tilde{t}_n \right) \Rightarrow \mathcal{Z}_1(1)$$

conditionally on $\mathcal{F}_n$, where $\mathcal{Z}_1(1)$ denotes the random variable defined in (4.3).

We summarize the bootstrap test for the hypothesis (1.3) of a non relevant change in the variance structure in the following algorithm.

Algorithm 4.1.
[1] Calculate the statistic $\check{T}_n$ defined in (4.2).
[2] Generate B conditionally *i.i.d* copies $\{\hat{\Phi}_{i,m}^{(r)}\}_{i=1}^{n-m+1}$ $(r = 1, 2, ..., B)$ of the sequence $\{\hat{\Phi}_{i,m}^V\}_{i=1}^{n-m+1}$ defined in (4.6) and calculate

$$M_r^V = \frac{1}{n} \frac{6}{\tilde{t}_n^2(1 - \tilde{t}_n)^2} \sum_{m+1 \le i \le n-m+1} \left(\hat{\Phi}_{i,m}^{(r)V} - \frac{i}{n-m+1} \hat{\Phi}_{n-m+1,m}^{(r)V} \right) \left(\frac{i\tilde{t}_n}{n} - \frac{i}{n} \wedge \tilde{t}_n \right).$$

[3] Let $M_{(1)}^V \le M_{(2)}^V \le ... \le M_{(B)}^V$ denote the order statistics of $M_1^V, \ldots, M_B^V$. Reject the null hypothesis (1.3) of a non relevant change in the variance at level α if

$$\check{T}_n > \delta^2 + M_{(\lfloor B(1-\alpha) \rfloor)}^V \delta / \sqrt{n}. \tag{4.7}$$

The p-value of this test is given by $1 - \frac{B^*}{B}$, where $B^* = \max\{r : \delta^2 + \frac{M_{(r)}^V \delta}{\sqrt{n}} \le \check{T}_n\}$.

Remark 4.1. It is of interest to investigate the power of the tests (4.4) and (4.7). For this purpose note that it follows for $\Delta \neq 0$ from (4.3)

$$\beta_n(\delta, \Delta) = \mathbb{P}\left(\check{T}_n > \delta^2 + \frac{v_{1-\alpha}\delta}{\sqrt{n}} \right) = \mathbb{P}\left(\sqrt{n} \frac{\check{T}_n - \Delta^2}{|\Delta|} > \sqrt{n} \frac{\delta^2 - \Delta^2}{|\Delta|} + \frac{v_{1-\alpha}\delta}{|\Delta|} \right)$$

$$\approx 1 - \Psi\left(\sqrt{n} \frac{\delta^2 - \Delta^2}{|\Delta|} + \frac{v_{1-\alpha}\delta}{|\Delta|} \right), \tag{4.8}$$

where Ψ is the distribution function of the random variable $\mathcal{Z}_1(1)$ (in fact a centered normal distribution). Therefore, under the alternative of a relevant change $\Delta^2 > \delta^2$, we have $\beta_n(\delta, \Delta) \to 1$ as $n \to \infty$, which provides the consistency of the test (4.4). On the other hand under the null

hypothesis $0 < \Delta^2 \leq \delta^2$ we have

$$\lim_{n\to\infty} \beta_n(\delta, \Delta) = \begin{cases} 0 & \text{if } 0 < \Delta^2 < \delta^2 \\ \alpha & \text{if } \Delta^2 = \delta^2 \end{cases}$$

If $\Delta = 0$, then $\check{T}_n = O_P(1/n)$ and $\lim_{n\to\infty} \beta_n(\delta, \Delta) = 0$, which means that the test (4.4) has in fact asymptotic level α.

We can also use (4.8) to investigate the power as a function of the parameter δ in the hypothesis (1.3). For example, we can see that for sufficiently large n the power $\beta_n(\delta, \Delta)$ is approximately 1 if $\delta \to 0$, and $\beta_n(\delta, \Delta)$ is approximately 0 if $\delta \to \infty$. Moreover, it is easy to see that all statements mentioned in this remark hold also for the bootstrap test defined by (4.7).

4.2 Relevant changes in correlation

Consider model (2.1), denote by $\rho_{i,k} = \text{corr}(Y_i, Y_{i+k})$ the correlation at lag k and suppose that for some unknown $t \in (0, 1)$

$$\rho_1 = \rho_{1,k} = ... = \rho_{\lfloor nt \rfloor, k}, \qquad \rho_2 = \rho_{\lfloor nt \rfloor + 1, k} = ... = \rho_{n-k,k}. \tag{4.9}$$

In this section we are interested in the problem of testing the hypothesis of a non relevant change in the correlation at lag k, that is

$$H_0 : |\rho_1 - \rho_2| \leq \delta \text{ versus } H_1 : |\rho_1 - \rho_2| > \delta \tag{4.10}$$

for some pre-specified $\delta > 0$. Dette and Wied (2014) provided a method for testing the hypothesis (4.10) under this and the additional assumption that the process before and after the change point exhibits a stationary behaviour. However, in general local stationary framework the construction of a test is more difficult and will be explained in the following paragraphs.

We denote by $\Delta = \rho_2 - \rho_1$ the (unknown) difference before and after the change point and assume throughout this section that under the null hypothesis of a non relevant change in the correlation the variance function σ^2 has either no jumps or has a jump at a point, say $\tilde{t}_v$, which does not necessarily coincide with the change point t in the correlation structure. In order to estimate the correlation consistently before and after the change point we recall the definition of the variance estimator (3.12), which addresses the problem that the variance function before

and after the change point is not constant. We define the CUSUM process

$$\hat{U}_n^{[k]}(s) = \frac{1}{n}\sum_{j=1}^{\lfloor ns\rfloor} \frac{\hat{e}_j\hat{e}_{j+k}}{\hat{\sigma}_n^{2*}(t_j)} - \frac{s}{n}\sum_{j=1}^{n-k} \frac{\hat{e}_j\hat{e}_{j+k}}{\hat{\sigma}_n^{2*}(t_j)}, \tag{4.11}$$

where $\hat{e}_i = Y_i - \hat{\mu}_{b_n}(t_i)$ denotes the nonparametric residuals from the local linear fit and we use the convention that $\hat{e}_i = 0$ for $i \geq n$. The estimator for the change point of the correlation structure is finally defined by

$$\hat{t}_n = \operatorname*{argmax}_{1\leq m\leq n} \left(\hat{U}_n^{[k]}(m/n)\right)^2. \tag{4.12}$$

Note that the statistic $\hat{t}_n$ depends on the estimator $\hat{t}_n^*$ for the change point in the variance, which is defined in (3.10). The first result of this section establishes consistency of this estimate.

Lemma 4.2. *Suppose that the conditions of Lemma 3.1 are satisfied, and that the conditions for the bandwidths b_n and c_n of Theorem 3.3 hold. The estimate $\hat{t}_n$ of the change point in the correlation structure defined by (4.12) satisfies*

$$\hat{t}_n \xrightarrow{\mathcal{D}} T_{\max}, \ \ if \ |\Delta| = 0, \tag{4.13}$$

$$|\hat{t}_n - t| = O_p(n^{-\alpha}), \ \ if \ |\Delta| > 0, \tag{4.14}$$

for some $\alpha > 1/2$, where $T_{\max}$ is a $[0,1]$-valued random variable.

The test for the hypothesis of a non relevant change will be based on the statistic

$$\hat{T}_n^r = \frac{3}{\hat{t}_n^2(1-\hat{t}_n)^2}\int_0^1 (\hat{U}_n^{[k]}(s))^2 ds, \tag{4.15}$$

where the the process $\hat{U}_n^{[k]}$ is defined in (4.11). The following theorem shows that $\hat{T}_n^r$ is a consistent estimator of Δ^2 and also provides its asymptotic distribution.

Theorem 4.3. *Assume that the conditions for the bandwidths b_n and c_n of Theorem 3.3 hold, and that Assumptions (A1) - (A4) are satisfied with $\iota \geq 16$. Suppose further that the variance function is strictly positive, twice differentiable on the intervals $(0, \tilde{t}_v)$ and $(\tilde{t}_v, 1)$, such that the second derivative $\ddot{\sigma}^2$ is Lipschitz continuous (here $\tilde{t}_v$ is the location of the change point, which is defined as $\tilde{t}_v = 1$ if there is no jump). If there exists a break, then we assume additionally that $\tilde{t}_v \in [\zeta, 1 - \zeta]$ for some constant $\zeta > 0$.*

20

i) If $\Delta \neq 0$, then

$$\sqrt{n}(\hat{T}_n^r - \Delta^2) \xrightarrow{\mathcal{D}} \mathcal{Z}_2(\Delta) := \frac{6}{t^2(1-t)^2} \int_0^1 [U_2(s) - sU_2(1)][st - s \wedge t]|\Delta|ds, \qquad (4.16)$$

where the Gaussian process $\{U_2(s)\}_{s \in [0,1]}$ is defined in Theorem 3.3.

ii) If $\Delta = 0$, then $\hat{T}_n^r = O_P(1/n)$.

A careful inspection of the proof of Theorem 4.3 shows that the statement (4.16) remains correct for any estimator of the change point in the correlation structure, which satisfies (4.13) and (4.14). Moreover, Theorem 4.3 yields an asymptotic level α test for the hypothesis (4.10) of a non relevant change in the correlation structure by rejecting H_0, whenever $\hat{T}_n^r > \delta^2 + \frac{\bar{v}_{1-\alpha}}{\sqrt{n}}$, where $\bar{v}_{1-\alpha}$ denotes the $(1-\alpha)$-quantile of the distribution of the random variable $\mathcal{Z}_2(\delta)$ defined in (4.16). We note again that this distribution is a centered normal distribution with a variance depending on the data generating process in a complicated way. In order to provide a consistent bootstrap approximation of the distribution of the random variable $\mathcal{Z}_2(1)$, recall the definition of the estimator $\hat{t}_n$ of the change point in the correlation structure in (4.12). We consider the statistics $\hat{\Delta}_{n,1} = \frac{1}{\lfloor n\hat{t}_n \rfloor} \sum_{j=1}^{\lfloor n\hat{t}_n \rfloor} \frac{\hat{e}_j \hat{e}_{j+k}}{\hat{\sigma}_n^{2*}(t_j)}$, $\hat{\Delta}_{n,2} = \frac{1}{n-\lfloor n\hat{t}_n \rfloor} \sum_{j=\lfloor n\hat{t}_n \rfloor+1}^{n-k} \frac{\hat{e}_j \hat{e}_{j+k}}{\hat{\sigma}_n^{2*}(t_j)}$ and define

$$\hat{\Delta}_n = \hat{\Delta}_{n,2} - \hat{\Delta}_{n,1} \qquad (4.17)$$

as an estimator of the difference $\Delta = \rho_2 - \rho_1$. The next lemma provides consistency of $\hat{\Delta}_n$ and is proved in Section 7.2.

Lemma 4.3. *Suppose that the conditions of Theorem 4.3 and Assumption (A6) are satisfied, then $\hat{\Delta}_n - \Delta = O_p\left(\frac{\log n}{\sqrt{n}}\right)$.*

Define

$$\hat{A}_j = \frac{\hat{e}_j \hat{e}_{j+k}}{\hat{\sigma}_n^{2*}(t_j)} - \hat{\Delta}_n I(j \geq \lfloor n\hat{t}_n \rfloor), \qquad (4.18)$$

where the variance estimator is given by (3.12), and let $\{R_j\}_{j \in \mathbb{Z}}$ be a sequence of i.i.d. standard normal distributed random variables, which is independent of $\{\mathcal{F}_i\}_{i \in \mathbb{Z}}$. We introduce the partial sums $\hat{S}_{j,m}^A = \sum_{r=j}^{j+m-1} \hat{A}_r$, $\hat{S}_n^A = \sum_{r=1}^n \hat{A}_r$ and define

$$\hat{\Phi}_{i,m}^A = \frac{1}{\sqrt{m(n-m+1)}} \sum_{j=1}^{n-m+1} \left(\hat{S}_{j,m}^A - \frac{m}{n}\hat{S}_n^A\right)R_j, \qquad (4.19)$$

then the following result is proved in Section 7.3.

Theorem 4.4. *Suppose the conditions of Theorem 4.3 hold and that $m \to \infty$, $m/n \to 0$, $\sqrt{m}\big(c_n^2 + (\frac{1}{\sqrt{nc_n}} + b_n^2 + \frac{1}{\sqrt{nb_n}})c_n^{-1/4}\big)\log n \to 0$, then*

$$M_n^r = \frac{1}{n}\frac{6}{\hat{t}_n^2(1-\hat{t}_n)^2} \sum_{m+1 \leq i \leq n-m+1} \left(\hat{\Phi}_{i,m}^A - \frac{i}{n-m+1}\hat{\Phi}_{n-m+1,m}^A\right)\left(\frac{i\hat{t}_n}{n} - \frac{i}{n}\wedge\hat{t}_n\right) \xrightarrow{\mathcal{D}} \mathcal{Z}_2(1) \quad (4.20)$$

conditional on $\mathcal{F}_n$, where the random variable $\mathcal{Z}_2(1)$ is defined in Theorem 4.3.

Theorem 4.4 provides a consistent asymptotic level α bootstrap test for the hypothesis of a non relevant change in the correlation structure. The hypothesis (4.10) is rejected, whenever

$$\hat{T}_n^r > M_{(\lfloor B(1-\alpha)\rfloor)}\delta/\sqrt{n} + \delta^2. \quad (4.21)$$

Here $M_{\lfloor B(1-\alpha)\rfloor}^r$ is the $(1-\alpha)$-quantile of bootstrap sample of the distribution of the statistic M_n^r defined in (4.20), which is generated in the same way as described in Algorithm 4.1. The consistency and the properties of the power function of the bootstrap test follow by similar arguments as given in of Remark 4.1 for the test of non relevant change in variance. The details are omitted for the sake of brevity.

Remark 4.2. The proposed method can easily be generalized to address the problem of testing for a relevant change in several correlations simultaneously. Exemplarily we illustrate such a generalization in the situation, where one is interested in detecting a relevant change in any of the lag 1- to lag q-correlations. Consider model (2.1) and suppose that there exist time points $t_k \in (0,1)$ such that

$$\rho_1(k) = \rho_{1,k} = ... = \rho_{\lfloor nt_k \rfloor,k}, \qquad \rho_2(k) = \rho_{\lfloor nt_k \rfloor+1,k} = ... = \rho_{n-k,k}$$

$(1 \leq k \leq q)$. We are interested in testing the hypotheses

$$H_0: \ |\rho_1(k) - \rho_2(k)| \leq \delta_k \text{ for all } k = 1,\ldots,q$$
$$H_1: \text{there exists a lag } k \in \{1,\ldots,q\} \text{ such that } |\rho_1(k) - \rho_2(k)| > \delta_k,$$

where $\delta_1,\ldots,\delta_q$ are given thresholds. For each lag k, let $\hat{t}_{n,k}$ denote the estimator for t_k defined in (4.12) and define

$$\hat{T}_{n,k}^r = \frac{3}{\hat{t}_{n,k}^2(1-\hat{t}_{n,k})^2}\int_0^1 \left(\hat{U}_n^{[k]}(s)\right)^2 ds, \quad (4.22)$$

22

where $\hat{U}_n^{[k]}$ is given by (4.11). Recall the definition of the q-dimensional process $\mathbf{U}_2(t) = (U_{2,1}, \ldots, U_{2,q})^T$ in Remark 3.4, and suppose that the conditions of Theorem 4.3 and (A6*) hold. Then it can be shown

$$\max_{1 \leq i \leq q} |\sqrt{n}(\hat{T}_{n,i}^r - \Delta_i^2)| \xrightarrow{\mathcal{D}} \mathcal{Z}^* = \max_{1 \leq i \leq q} \left| \frac{6}{t_i^2(1-t_i)^2} \int_0^1 [U_{2,i}(s) - sU_{2,i}(1)][st_i - s \wedge t_i] |\Delta_i| ds \right|,$$

where $\Delta_k = \rho_1(k) - \rho_2(k)$ $(1 \leq k \leq q)$. Similarly, bootstrap methodology can be developed considering consistent estimates, say $\hat{\Delta}_{k,n}$, of Δ_k. Let

$$\hat{A}_j^{[k]} = \frac{\hat{e}_j \hat{e}_{j+k}}{\hat{\sigma}_n^{2*}(t_j)} - \hat{\Delta}_{k,n} I(j \geq \lfloor n\hat{t}_k \rfloor), \tag{4.23}$$

and consider the vector $\hat{\mathbf{A}}_j = (\hat{A}_1^{[1]}, \ldots, \hat{A}_q^{[k]})^T$. Further, define $\hat{\mathbf{S}}_{j,m}^A = \sum_{r=j}^{j+m-1} \hat{\mathbf{A}}_r$, $\hat{\mathbf{S}}_{n-q}^A = \sum_{r=1}^{n-q} \hat{\mathbf{A}}_r$ and

$$\hat{\boldsymbol{\Phi}}_{i,m}^A = \frac{1}{\sqrt{m(n-2q-m+1)}} \sum_{j=q+1}^{n-m+1} \left(\hat{\mathbf{S}}_{j,m}^A - \frac{m}{n} \hat{\mathbf{S}}_{n-q}^A \right) R_j \tag{4.24}$$

$(i = q+1, ..., n-m-q+1)$, where $\{R_j\}_{j \in \mathbb{Z}}$ is a sequence of i.i.d. standard normal distributed random variables, which is independent of $\{\mathcal{F}_i\}_{i \in \mathbb{Z}}$. Let $\hat{\boldsymbol{\Phi}}_{i,m}^{A[k]}$ be the k_{th} entry of $\hat{\boldsymbol{\Phi}}_{i,m}^A$ $(1 \leq k \leq q)$ and define

$$M_{n,k}^r = \frac{1}{n} \frac{6}{\hat{t}_{n,k}^2(1-\hat{t}_{n,k})^2} \sum_{m+1 \leq i \leq n-m+1} \left(\hat{\boldsymbol{\Phi}}_{i,m}^{A[k]} - \frac{i}{n-m+1} \hat{\boldsymbol{\Phi}}_{i,m}^{A[k]} \right) \left(\frac{i}{n} \hat{t}_{n,k} - \frac{i}{n} \wedge \hat{t}_{n,k} \right). \tag{4.25}$$

Then (conditional on $\mathcal{F}_n$) $\max_{1 \leq k \leq q} |\hat{\Delta}_k M_{n,k}^r| \xrightarrow{\mathcal{D}} \mathcal{Z}^*$, which provides the consistency of a corresponding bootstrap test.

5   Finite sample properties

In this section we investigate the finite sample properties of the proposed tests by means of a simulation study. In all examples considered we used the function $\mu(t) = 8(-(t-0.5)^2 + 0.25)$ as mean function and a sequence of independent identically normal distributed random variables $\{\varepsilon_j\}_{j \in \mathbb{Z}}$ in the definition of the errors $e_i = G_j(i/n, \mathcal{F}_i)$ in model (2.1), where $\mathcal{F}_i = \sigma(\ldots, \varepsilon_0, \ldots, \varepsilon_i)$. The dependency structures differ by different choices for the nonlinear filter G_j. The sample size is $n = 500$ and all results are based on 2000 simulation runs. In each run,
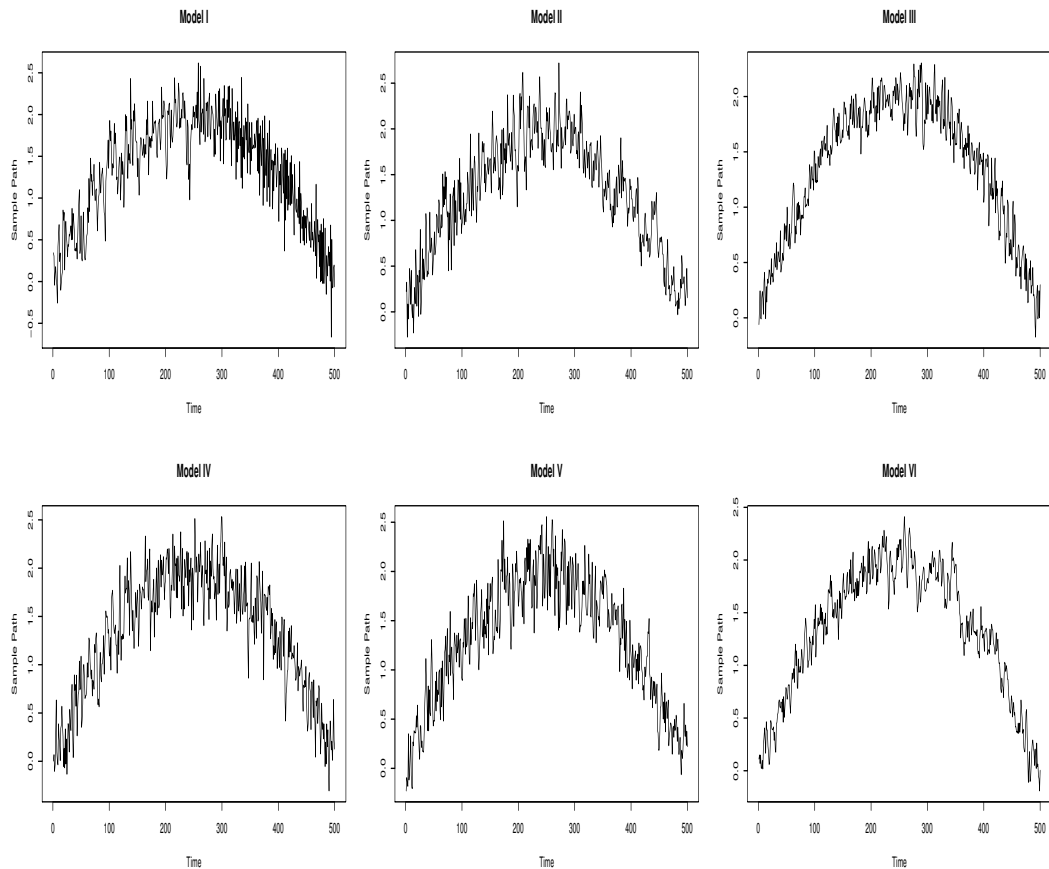
Figure 1: *Typical sample paths of the processes corresponding to model I-VI.*

the critical values are generated by $B = 2000$ bootstrap replications.

5.1 Change point tests for the variance

In this section we investigate the finite sample properties of the bootstrap tests for the "classical" hypothesis (1.1) of a constant variance and for the null hypothesis (1.3) of a non relevant change in the variance. It turned out that for change point analysis of the variance the *Minimal Volatility (MV) method* for the selection of the bandwidth provided slightly better results than the commonly used cross validation. The MV-method has been advocated in Politis et al. (1999). To be precise consider a sufficiently wide interval $(a, b) \subset (0, 1)$ and define the points $d_i := a + (i-1)(b-a)/l \in (a, b)$, $i = 1, 2, \ldots, l$ as potential bandwidths. For each d_i, we calculate the statistic $\hat{T}_n(i)$, which is defined as the test statistic $\hat{T}_n$ in (3.1) with bandwidth $b_n = d_i$. We then calculate for $i = 4, \ldots, l-3$ the standard errors $SD(i) := sd(\hat{T}_n(j), i-3 \leq j \leq i+3)$ of different subsamples of these statistics. Finally, the bandwidth b_n in the interval (a, b) is chosen

as $b_n = d_{i^*}$, where $i^* = \text{argmin}_{4 \leq i \leq l-3} SD(i)$.

In the following we discuss three models for the filters of the innovations in model (2.1).

(I) $G(t, \mathcal{F}_i) = H(t, \mathcal{F}_i)/4$, where $H(t, \mathcal{F}_i) = 0.5H(t, \mathcal{F}_{i-1}) + \varepsilon_i$ for $t \leq 0.5$, and $H(t, \mathcal{F}_i) = -0.5H(t, \mathcal{F}_{i-1}) + \varepsilon_i$ for $t > 0.5$.

(II) $G(t, \mathcal{F}_i) = H(t, \mathcal{F}_i)\sqrt{1 - a^2(t)}/4$, where $H(t, \mathcal{F}_i) = \sum_{j=0}^{\infty} a^j(t)\varepsilon_{i-j}$ and $a(t) = 1/4 + t/2$.

(III) $G(t, \mathcal{F}_i) = H(t, \mathcal{F}_i)\sqrt{1 - a^2(t)}/8$ for $t \leq 0.5$, and $G(t, \mathcal{F}_i) = H(t, \mathcal{F}_i)\sqrt{2(1 - b^2(t))}/8$ for $t > 0.5$. Here $H(t, \mathcal{F}_i) = \sum_{j=0}^{\infty} a^j(t)\varepsilon_{i-j}$ for $t \leq 0.5$, and $H(t, \mathcal{F}_i) = \sum_{j=0}^{\infty} b^j(t)\varepsilon_{i-j}$ for $t > 0.5$, $a(t) = 1/4 + t/2$ and $b(t) = 0.5 - (t - 0.5)^2$.

Model (I) is a piecewise stationary process. The correlation has a structural break at the point $t = 0.5$. Model (II) is a locally stationary process. The MA coefficient is smoothly time-varying. By our construction, the variance of model (I) and model (II) remains constant, which means that both models correspond to the "classical" null hypothesis in (1.1). Model (III) is a piecewise locally stationary process. The correlation of model (III) has a jump at the point $t = 0.5$. The MA coefficients before and after the point $t = 0.5$ are smoothly varying. The variance of model (III) also has a jump at $t = 0.5$ and remains constant before and after the jump, respectively. This model corresponds to the null hypothesis in (1.3) of a non relevant change point in the variance. Typical trajectories of the processes corresponding to model (I) - (III) are displayed in the upper part of Figure 1.

For model (I) and (II) we investigate approximation of the nominal level of the bootstrap test (3.7) for the classical hypothesis (1.1) of a change in the variance over the interval $[0, 1]$. The corresponding rejection probabilities are displayed in the left and middle column of Table 1. We display the simulated type I error using different bandwidths in the interval $(0.025, 0.3)$ and the bandwidth calculated by the MV-method (last line). We observe that the results are rather stable with respect to the choice of b_n. Only if $b_n \leq 0.1$ the level is overestimated. In particular the bandwidth calculated by the MV-method yields good results for both models.

For model (III), we are interested in testing the hypothesis (1.3) of non relevant changes in variance, where the threshold is given by $\delta = 1/64$. The corresponding rejection probabilities are shown in the right column of Table 1 for the case that $\Delta = \delta = 1/64$. Note that this choice corresponds to the boundary of the null hypothesis, and according to Remark 4.1 the nominal level is smaller in the interior of the null hypothesis, i.e. $\Delta < \delta$ (these results are not displayed for the sake of brevity). Interestingly - compared to the problem of testing the "classical" hypothesis (1.1) - the method is more sensitive with respect to the choice of the bandwidth. However, the MV-method yields a rather accurate approximation of the nominal level.

Table 1: *Simulated Type I error of the tests* (3.7) *and* (4.7) *for a change in the variance for various bandwidths and the bandwidth calculated by the MV-method (last line). Left and middle column: test for the hypothesis* (1.1) *(model (I) and (II)). Right column: test for the hypothesis* (1.3) *(model (III)).*

model	I		II		III	
b_n/α	5%	10%	5%	10%	5%	10%
0.025	19.15	31.55	30.45	44.9	2.45	4.95
0.05	8.8	16.9	15.2	25.9	5.4	8.7
0.075	5.7	13.1	10.25	18.75	7.55	11.9
0.1	6.05	12.8	8.85	16.3	8.15	12.7
0.125	4.75	10.95	7.95	15.2	7.4	12.15
0.15	5.55	11.9	7.3	14.5	8.75	14.05
0.175	4.1	10.2	6.5	14.5	8	13.05
0.2	4.6	10.7	4.9	11.15	7.55	11.8
0.225	3.55	8.85	4.75	10.3	7.3	12.85
0.25	3	8	5.15	10.8	8.35	14.75
0.275	3.05	9.3	4.75	10.7	7.8	14.15
0.3	4.15	9.55	5.6	10.35	16.35	26.65
MV	4.7	11.4	6.75	13.9	6.75	12.3

5.2 Change point tests for a lag k-correlation

We now investigate the same properties of the tests for changes in the lag 1-correlation. For this purpose we consider the following models.

(IV) $G(t, \mathcal{F}_i) = H(t, \mathcal{F}_i)\sqrt{1 - (t - 0.5)^2}/4$, where $H(t, \mathcal{F}_i) = 0.3H(t, \mathcal{F}_i) + \varepsilon_i$.

(V) $G(t, \mathcal{F}_i) = H(t, \mathcal{F}_i)\sqrt{c(t)}/4$ for $t \leq 0.5$, and $G(t, \mathcal{F}_i) = H(t, \mathcal{F}_i)\sqrt{d(t)}/4$ for $t > 0.5$, where $c(t) = 1 - (t - 0.5)^2$, $d(t) = 1 - \frac{1}{2}\sin t$ and $H(t, \mathcal{F}_i) = 0.3H(t, \mathcal{F}_i) + \varepsilon_i$.

(VI) $G(t, \mathcal{F}_i) = H(t, \mathcal{F}_i)\sqrt{1 - (t - 0.5)^2}/8$, where $H(t, \mathcal{F}_i) = 0.5H(t, \mathcal{F}_{i-1}) + \varepsilon_i$ for $t \leq 0.5$, and $H(t, \mathcal{F}_i) = 0.7H(t, \mathcal{F}_{i-1}) + \varepsilon_i$ for $t > 0.5$.

Model (IV) is a locally stationary processes. The variance of the process is time-varying, but the correlation remains constant. Model (V) and model (VI) are piecewise locally stationary processes, where the variance has a change point. Before and after the jump, the variance varies smoothly. The correlation of model (IV) and (V) is constant, while the correlation of model (VI) has a break at $t = 0.5$. Typical trajectories corresponding to these processes are depicted in the lower part of Figure 1

Note that the change point analysis by the tests proposed in Section 4.1 and 4.2 requires the choice of two bandwidths in the local linear estimates of the mean and variance. We use a generalized cross validation method introduced by Zhou and Wu (2010) to select the bandwidth for estimating the mean function. Then we apply this cross validation procedure again to select the bandwidth for estimating the variance function. The parameters L and ζ in the estimator (3.10) are chosen as $L = \lfloor n^{1/3} \rfloor$ and $\zeta = 0.016$, respectively.

The corresponding rejection probabilities are displayed in Table 2 for various bandwidths b_n in the interval $(0.025, 0.3)$. At each fixed b_n, the bandwidth for variance c_n is calculated by cross validation. The results for the bandwidths calculated by cross validation are displayed in the last row of the table. The left and middle column correspond to the "classical" hypothesis whether the correlation remains constant (model (IV) and (V)). We observe a very accurate approximation of the nominal level if the bandwidth is chosen such that $b_n < 0.1$. If $b_n \in (0.1, 0.3)$ the nominal level is slightly underestimated. For both models the generalized cross validation proposed by Zhou and Wu (2010) yields a rather accurate approximation of the nominal level,

The right column of Table 2 shows the simulated type I error of the test (4.10) for a non relevant change in correlation with $\delta = 0.2$. Again the case $\delta = \Delta = 0.2$ is displayed in model (VI) corresponding to the boundary of the null hypothesis, and the simulated rejection probabilities are smaller if $|\Delta| < \delta$. Compared to the test for a non relevant change in the variance (see Table 1) the test for a non relevant change in the correlation is rather stable with respect to the choice of the bandwidth. Also the proposed cross validation methodology performs reasonably well, which is reported in the last line of Table 2.

Finally, we display in Figure 2 the simulated rejection probabilities of the tests for the hypotheses (1.3) and (4.10) for a non relevant change in the variance or correlation, respectively, as a function of the parameter $\delta \in [0, 2\Delta]$. The significance level is chosen as 0.1. As expected the probability of rejection decreases with δ (see also the discussion in Remark 4.1).

5.3 Power properties

In this section we investigate the power of the tests proposed in this paper considering the following four scenarios.

(I') $G(t, \mathcal{F}_i) = (\mathbf{1}(t \leq 0.5) + \mathbf{1}(t > 0.5)\sqrt{1 + \lambda})H(t, \mathcal{F}_i)/4$, where $\lambda > -1$, $H(t, \mathcal{F}_i) = 0.5H(t, \mathcal{F}_{i-1}) + \varepsilon_i$ for $t \leq 0.5$, and $H(t, \mathcal{F}_i) = -0.5H(t, \mathcal{F}_{i-1}) + \varepsilon_i$ for $t > 0.5$.

(II') $G(t, \mathcal{F}_i) = H(t, \mathcal{F}_i)\big[\mathbf{1}(t \leq 0.5)\sqrt{1 - a^2(t)} + \mathbf{1}(t > 0.5)\sqrt{(2 + \lambda)(1 - b^2(t))}\big]/8$, where

27

Table 2: *Simulated Type I error of the tests for a change in the lag 1-correlation for various bandwidths and the bandwidth calculated by generalized cross validation (last line). Left and middle column: test for the hypothesis* (3.8) *(model (IV) and (V)). Right column: test for the hypothesis* (4.10) *(model (VI)).*

model	IV		V		VI	
b_n/α	5%	10%	5%	10%	5%	10%
0.025	4.7	10.4	4.9	10.3	6	8.95
0.05	3.75	8.5	4.5	9.1	6	9.25
0.075	4.15	9.7	4.4	9.6	7.05	9.9
0.1	3.4	8.1	3.4	8.1	6.05	9.55
0.125	3.85	9.1	4.05	8.25	6.4	9.35
0.15	3.35	8.35	3.2	8.25	5.3	7.7
0.175	2.85	8.15	3.5	8.3	4.9	6.75
0.2	2.6	7.6	3.95	8.9	4.5	6.95
0.225	2.9	7.75	3.1	8.05	4	6
0.25	3.1	8	3.45	8.25	5.15	7.6
0.275	2.8	8.75	3.5	8.9	6.3	8.9
0.3	3.05	8.55	3.15	8.35	5.9	8.8
CV	5	10.15	4.75	9.65	6.75	10.45
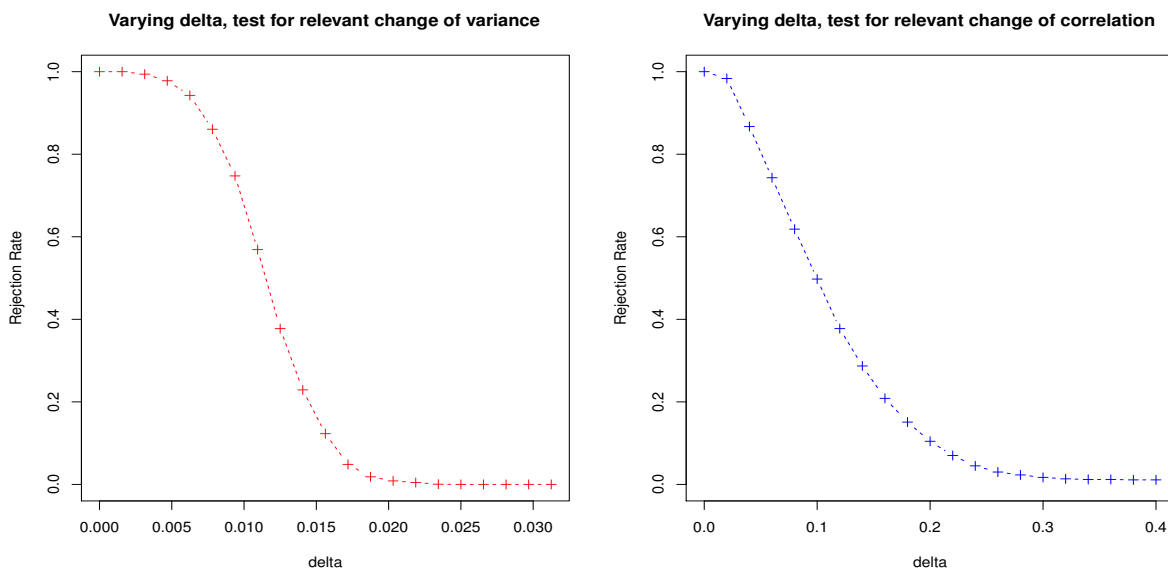
Figure 2: *Simulated rejection probabilities of the test* (4.10) *for the hypothesis of a non relevant change in the variance (left panel) and lag 1-correlation (right panel) as a function of the threshold $\delta \in [0, 2\Delta]$ in the hypothesis* (1.3).

$H(t, \mathcal{F}_i) = \sum_{j=0}^{\infty} a^j(t) \varepsilon_{i-j}$ for $t \leq 0.5$, and $H(t, \mathcal{F}_i) = \sum_{j=0}^{\infty} b^j(t) \varepsilon_{i-j}$ for $t > 0.5$ and the function a and b are defined by $a(t) = 1/4 + t/2$ and $b(t) = 0.5 - (t - 0.5)^2$, respectively.

(III') $G(t, \mathcal{F}_i) = \sqrt{1 - (t - 0.5)^2}/4 H(t, \mathcal{F}_i)$, where $H(t, \mathcal{F}_i) = 0.3 H(t, \mathcal{F}_i) + \varepsilon$ for $t \leq 0.5$, and $H(t, \mathcal{F}_i) = (0.3 - \lambda) H(t, \mathcal{F}_i) + \varepsilon$ for $t > 0.5$.

(IV') $G(t, \mathcal{F}_i) = H(t, \mathcal{F}_i) \sqrt{1 - (t - 0.5)^2}/8$, where $H(t, \mathcal{F}_i) = (0.5 - \lambda) H(t, \mathcal{F}_{i-1}) + \varepsilon_i$ for $t \leq 0.5$, and $H(t, \mathcal{F}_i) = 0.7 H(t, \mathcal{F}_{i-1}) + \varepsilon_i$ for $t > 0.5$,

Model (I') is used to study the power of the test (3.7) where the case $\lambda = 0$ corresponds to the null hypothesis of a constant variance. The power properties of the test (4.7) for a non relevant change in the variance is investigated in model (II'). Here we test the hypotheses $H_0 : \Delta \leq 1/64$ versus $H_1 : \Delta > 1/64$, where the case $-2 < \lambda \leq 0$ corresponds to the null hypothesis. Similarly, the power of the test for a constant lag 1-correlation (3.18) is studied in model (III') (again the case $\lambda = 0$ corresponds to the null hypothesis of a constant correlation) and the corresponding hypotheses $H_0 : \Delta \leq 0.2$ versus $H_1 : \Delta > 0.2$ of a non relevant change in the lag 1-correlation are investigated in model (IV') (here the case $-0.4 \leq \lambda \leq 0$ corresponds to the null hypothesis). The rejection probabilities for various values of λ are displayed in Figure 3. We observe that the proposed methodology can detect (relevant) changes in the variance or correlation with reasonable size.

6 Data Analysis

Various scientific data show that since the late 1800s, the mean global temperature starts to increase significantly. For example, during 1906–2005, the Earth's average surface temperature rose by $0.74 \pm 0.18\,°C$, with the rate of warming also increasing with time. On the other hand, however, there are much fewer studies on possible changes in the second order characteristics (especially the correlations) of temperature time series, both globally and regionally. In this section, we are interested in identifying possible changes in the second order structures of regional temperature in the recent three centuries. To this end, we analyse the Hadley Centre Central England Temperature (HadCET) data from 1659–2015. These data can be downloaded from http://www.metoffice.gov.uk/hadobs/hadcet/. We present the analysis results of monthly temperature series in January and July as representatives of the winter and summer monthly temperature patterns in central England. The time series are shown in Figure 4. There are apparent increasing trends in both time series.
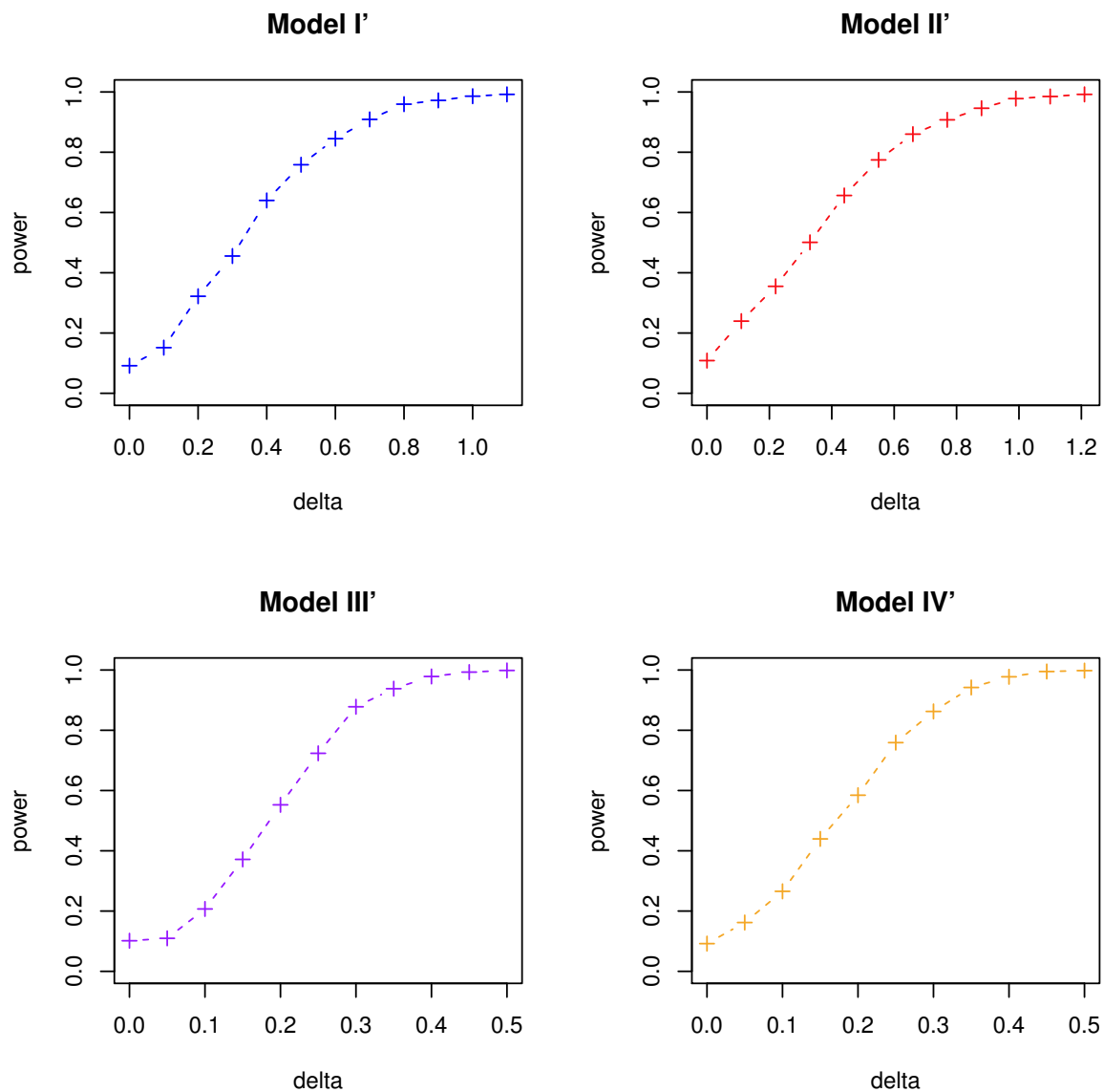
Figure 3: *Simulated power of tests for a change in the variance or lag 1-correlation. Left upper panel: test for a constant variance defined in (3.7) (model (I')). Right upper panel: test for a non relevant change in the variance defined in (4.7) (model (II')). Left lower panel: test for a constant lag 1-correlation defined in (3.18) (model (III')). Right lower panel: test for a non relevant change in the lag 1-correlation defined in (4.21) (model (IV')).*

We first test the constancy in the variance and lag 1-correlation for the January data. In our analysis the critical values are generated by 8000 bootstrap replications. The bandwidths were chosen as described in Section 5. The results are summarized in Table 3. The test (3.7) rejects the null hypothesis of no change point in the variance at 5% level. We then use the statistic (4.1) to estimate the change point and obtain $\tilde{t}_n = 226$, which corresponds to the year of 1884. Next we apply the test (3.7) again to the periods before and after the identified change point and conclude that there are no further structural breaks in the variance during the two periods. The estimates of the variance before and after the year 1884 are given by 4.05 and 2.85, respectively. Next we apply (3.18) to testing the constancy in the lag 1-correlation, where we use the statistic (3.10) to estimate the change points in the variance with $\zeta = 0.14$ and $L = 38$. We identify $t_n^* = 242$, which corresponds to the year of 1900. The result is close to the one which is obtained by the estimator (4.1). The null hypothesis of no change points in the lag 1-correlation is rejected at 5% level [see Table 3]. Next we use the statistic (4.12) to identify the location of the change point of the first order correlation and obtain $\hat{t}_n = 213$, which corresponds to the year of 1871. Again we investigate the existence of further changes in the lag 1-correlation before and after the year 1871 and conclude that there are no further structural breaks in the lag 1-correlation during the two periods. The estimates of the lag 1-correlation before and after the break point are equal to -0.108 and 0.231, respectively.

The results from Section 4 enable us to perform tests for relevant changes in the variance and lag 1-correlation for the January data. Figure 5 displays the p-values of the tests for a relevant change in the variance and log 1-correlation for different values of the threshold δ. At the 5% significance level, we conclude that there exists a relevant change with size $\delta = 0.645$ in the variance and a relevant change with size $\delta = 0.313$ in the lag 1-correlation.

For comparison, we also analyse the July data. For the variance, we choose the bandwidths $b_n = 0.205$ and $m = 33$. The test statistic (3.1) is 1.90, together with the simulated 90% critical value 1.83 and 95% critical value 2.04. Hence we cannot reject the null hypothesis of no change point in the variance at the 5% significance level. For the correlation, the bandwidths are chosen as $b_n = 0.26$, $c_n = 0.06$ and $m = 25$. The test statistic (3.14) is 0.84, and the simulated 90% critical value is 1.11, and the 95% critical value is 1.25. Hence, again we cannot reject the null hypothesis of no change points in lag 1-correlation at the 10% significance level.

In conclusion, our data analysis suggests that besides the mean trend, there exists strong evidence indicating structural changes in the second order structures of monthly temperatures in central England. Further, the latter changes are inhomogeneous among different seasons, in the sense that changes in the variance and correlation are more significant in the winter than in the summer. This implies that winter temperatures in central England have become more unstable

and more difficult to predict since the late 19th century. Finally, we also locate the change point in the second order structure of the HadCET temperate data through all three ways we proposed in our paper. The three change points, 1871, 1884 and 1900, are quite close. Our findings suggest that the time of change in the second order structure of our data coincides with that of the mean global temperature identified in various previous studies.
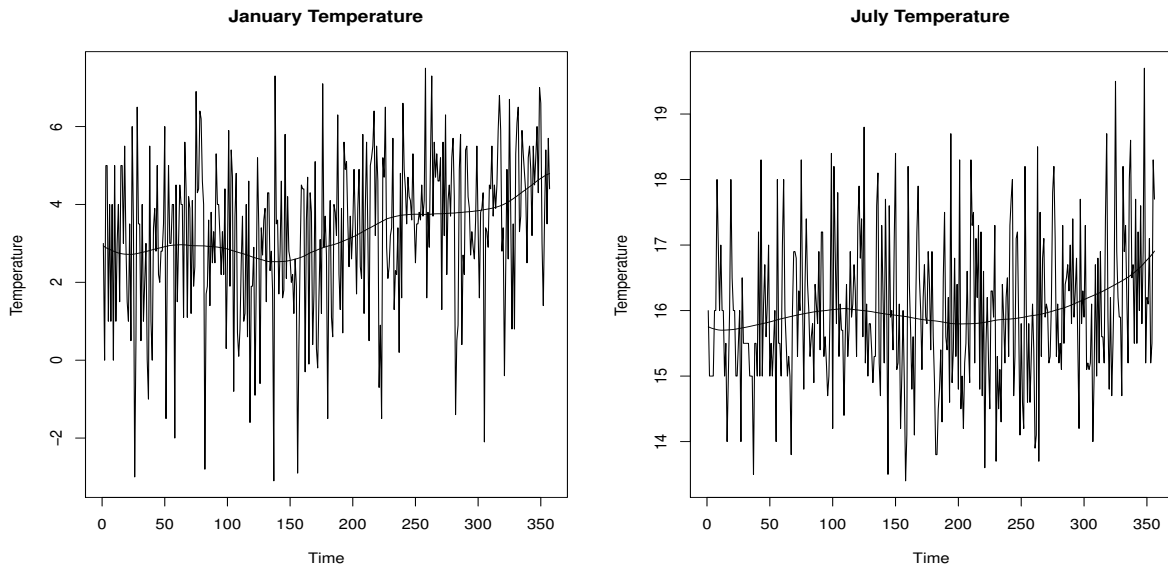

Figure 4: *Temperature in °C of UK from 1659–2015 in January (left panel) and July (right panel). The lines are the fitted trends of the means by local linear regression.*

Table 3: *Tests for the existence of a change point in the variance and lag 1-correlation in the HadCET data. v_α^* denotes the critical values obtained by the bootstrap procedure. "Whole" represents the whole period, "Before" and "After" represent the period before and after the detected change point, respectively.*

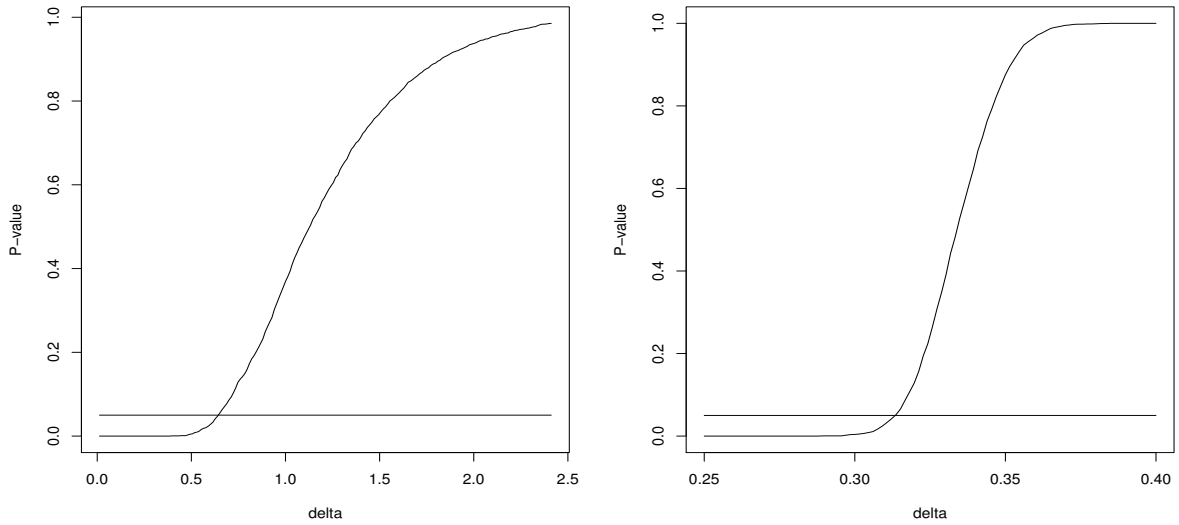|	Variance			lag 1-Correlation		
	Whole	Before	After	Whole	Before	After
Test Stat.	5.29**	2.82	3.34	1.53**	0.64	0.71
$v_{90\%}^*$	4.56	4.66	5.1	1.31	0.76	0.95
$v_{95\%}^*$	5.11	5.22	5.67	1.45	0.85	1.07
b_n	0.155	0.26	0.26	0.23	0.19	0.21
m	40	30	18	19	32	11
c_n	–	–	–	0.05	0.06	0.059

Figure 5: *p-values of the bootstrap test for a relevant change in the variance (left panel) and lag 1-correlation (right panel) for different values of the threshold δ. The horizontal line marks the significance level* 0.05.

References

Abraham, B. and Wei, W. W. S. (1984). Inferences about the parameters of a time series model with changing variance. *Metrika*, 31:183–194.

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(6):4046–4087.

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

Baufays, P. and Rasson, J. P. (1985). Variance changes in autoregressive models. In Anderson, D., editor, *Time Series Analysis: Theory and Practice 7*, pages 119–127. North-Holland, New York.

Berger, J. O. and Delampady, M. (1987). Testing precise hypotheses. *Statistical Science*, 2(3):317–335.

Berkson, J. (1938). Some difficulties of interpretation encountered in the application of the chi-square test. *Journal of the American Statistical Association*, 33(203):526–536.

Chen, J. and Gupta, A. K. (1997). Testing and locating variance changepoints with application to stock prices. *Journal of the American Statistical Association*, 92(438):739–747.

Chow, S.-C. and Liu, P.-J. (1992). *Design and Analysis of Bioavailability and Bioequivalence Studies*. Marcel Dekker, New York.

Davis, R. A., Lee, T. C. M., and Rodriguez-Yam, G. A. (2006). Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association*, 101(473):223–239.

Dette, H. and Wied, D. (2014). Detecting relevant changes in time series models. *Journal of the Royal Statistical Society, Ser., B., to appear, arxiv.org/abs/1403.8120* .

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. Chapman & Hall, London.

Galeano, P. and Peña, D. (2007). Covariance changes detection in multivariate time series. *Journal of Statistical Planning and Inference*, 137:194–211.

Inclán, C. and Tiao, G. C. (1994). Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association*, 89(427).

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.

Lee, S. and Park, S. (2001). The cusum of squares test for scale changes in infinite order moving average processes. *Scandinavian Journal of Statistics*, 28(4):625644.

Mallik, A., Banerjee, M., and Sen, B. (2013). Asymptotics for p-value based threshold estimation in regression settings. *Electronic Journal of Statistics*, 7:2477–2515.

Mallik, A., Sen, B., Banerjee, M., and Michailidis, G. (2011). Threshold estimation based on a p-value framework in dose-response and regression settings. *Biometrika*, 98:887–900.

McBride, G. B. (1999). Equivalence tests can enhance environmental science and management. *Australian & New Zealand Journal of Statistics*, 41:19–29.

Müller, H.-G. (1992). Change-points in nonparametric regression analysis. *Annals of Statistics*, 20:737–761.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, 41:100–115.

Politis, D. N., Romano, J. P., and Wolf, M. (1999). *Subsampling.* Springer, New York.

Preuss, P., Puchstein, R., and Dette, H. (2014). Detection of multiple structural breaks in multivariate time series. *Journal of the American Statistical Association, DOI: 10.1080/01621459.2014.920613.*

Qu, Z. (2008). Testing for structural change in regression quantiles. *Journal of Econometrics*, 148:170–184.

Solomon, S. (2007). *Climate change 2007-the physical science basis: Working group I contribution to the fourth assessment report of the IPCC*, volume 4. Cambridge University Press.

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

Vogt, M. and Dette, H. (2015). Detecting gradual changes in locally stationary processes. *Annals of Statistics*, 43(2):713–740.

Wichern, D. W., Miller, R. B., and Hsu, D.-A. (1976). Changes of variance in first-order autoregressive time series models - with an application. *Journal of the Royal Statistical Society. Ser. C (Applied Statistics)*, 25(3):248–256.

Wied, D., Krämer, W., and Dehling, H. (2012). Testing for a change in correlation at an unknown point in time using an extended functional delta method. *Econometric Theory*, 28(3):570–589.

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

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

Zhou, Z. (2014). Inference of weighted V-statistics for non-stationary time series and its applications.. *Annals of Statistics*, 42:87–114.

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

7 Proofs of main results

In this section we provide proofs of the main results, where some of the technical details are deferred to an online supplement. Throughout this section the symbol $\Rightarrow$ denotes weak convergence of a stochastic process in $\mathcal{C}(0,1)$ with the uniform topology.

7.1 Proof of Theorem 3.1, 3.2, Lemma 3.1, Theorem 3.3 and 3.4

7.1.1 Proof of Theorem 3.1

In order to study the asymptotic properties of the statistic $\hat{T}_n$ we introduce the random variable

$$T_n = \max_{1 \leq i \leq n} \left| S_i - \frac{i}{n} S_n \right|,$$

where S_i is the ith partial sum of the PLS $\{e_i^2\}_{i=1}^n$, that is $S_i = \sum_{j=1}^i e_j^2$. It follows from the results of Zhou (2013) that $\frac{1}{\sqrt{n}} T_n \xrightarrow{\mathcal{D}} \sup_{t \in (0,1)} |U_1(t) - t U_1(1)|$, where $\{U_1(t)\}_{t \in [0,1]}$ is a centered Gaussian process with covariance kernel (3.5).

We will show below that for the maximum deviation between $\hat{S}_i$ and S_i satisfies

$$\max_{1 \leq i \leq n} |S_i - \hat{S}_i| = O_p(\sqrt{nb_n} + nb_n^3 + b_n^{-1}). \tag{7.1}$$

From (7.1) we obtain $T_n - \hat{T}_n = O_p(\sqrt{nb_n} + nb_n^3 + b_n^{-1}) = o(\sqrt{n})$, where the last estimate follows from our choice of the bandwidth b_n. This yields the assertion of Theorem 3.1.

For a proof of the remaining estimate (7.1) we use the decomposition

$$S_i - \hat{S}_i = A_{n,i} + B_{n,i}, \tag{7.2}$$

where the quantities $A_{n,i}$ and $B_{n,i}$ are defined by

$$A_{n,i} = 2 \sum_{j=1}^{i} e_j \big(\hat{\mu}_{b_n}(t_j) - \mu(t_j) \big), \quad B_{n,i} = \sum_{j=1}^{i} \big(\mu(t_j) - \hat{\mu}_{b_n}(t_j) \big)^2.$$

Observing the estimate (8.4) in Section 8.1 of the technical appendix we have $0 \leq B_{n,i} \leq B_{n,n} = O_p(b_n^{-1} + nb_n^4)$ for all $1 \leq i \leq n$, which implies

$$\max_{1 \leq i \leq n} B_{n,i} = O_p(b_n^{-1} + nb_n^4). \tag{7.3}$$

By Lemma 8.1 (which is proved in Section 8.1) it follows that

$$\max_{\lfloor nb_n \rfloor \leq i \leq n - \lfloor nb_n \rfloor} \Big| A_{n,i} - 2 \sum_{j=\lfloor nb_n \rfloor+1}^{i} a_{n,j} - 2 \sum_{j=1}^{\lfloor nb_n \rfloor} e_j(\hat{\mu}_{b_n}(t_j) - \mu(t_j)) \Big| = O_p(n\chi_n),$$

$$\max_{n-\lfloor nb_n \rfloor+1 \leq i \leq n} \Big| A_{n,i} - 2 \sum_{j=\lfloor nb_n \rfloor}^{n-\lfloor nb_n \rfloor} a_{n,j} - 2 \sum_{j=1}^{\lfloor nb_n \rfloor - 1} e_j(\hat{\mu}_{b_n}(t_j) - \mu(t_j))$$

$$- 2 \sum_{j=n-\lfloor nb_n \rfloor+1}^{i} e_j(\hat{\mu}_{b_n}(t_j) - \mu(t_j)) \Big| = O_p(n\chi_n),$$

where $\chi_n = b_n^3 + \frac{b_n}{n}$, and

$$a_{n,j} = \frac{e_j}{nb_n} \sum_{s=1}^{n} K_{b_n}\Big(\frac{s-j}{n} \Big) e_s \qquad (j = 1, \ldots, n). \tag{7.4}$$

A further application of the estimate (8.4) in Section 8.1 and the Cauchy-Schwarz inequality gives

$$\Big\| \max_{1 \leq j \leq \lfloor nb_n \rfloor} \Big| \sum_{i=1}^{j} e_j \big(\mu(t_j) - \hat{\mu}_{b_n}(t_j) \big) \Big| \Big\|_2 \leq \sum_{i=1}^{\lfloor nb_n \rfloor} \|e_j\|_4 \|\mu(t_j) - \hat{\mu}_{b_n}(t_j)\|_4$$

$$= O(\sqrt{nb_n} + nb_n^3),$$

$$\Big\| \max_{n-\lfloor nb_n \rfloor+1 \leq j \leq n} \Big| \sum_{i=n-\lfloor nb_n \rfloor+1}^{j} e_j \big(\mu(t_j) - \hat{\mu}_{b_n}(t_j) \big) \Big| \Big\|_2 = O(\sqrt{nb_n} + nb_n^3).$$

37

This implies that

$$\max_{1 \leq i \leq n} |A_{n,i}| \leq \max_{\lfloor nb_n \rfloor \leq i \leq n - \lfloor nb_n \rfloor} |\bar{A}_{n,i}| + O_p(\sqrt{nb_n} + nb_n^3), \tag{7.5}$$

where $\bar{A}_{n,i} = 2 \sum_{j=\lfloor nb_n \rfloor}^{i} a_{n,j}$ and $a_{n,j}$ is defined in (7.4).

In the following we derive an estimate for the first term on the right-hand side of (7.5). For this purpose we consider the random variables $\tilde{e}_{s,m} = \mathbb{E}(e_s | \varepsilon_s, ..., \varepsilon_{s-m})$ and note that the sequence $(\tilde{e}_{s,m})_{s=1}^{n}$ is m-dependent. Now define $a_{n,j}^{(m)} = e_j \sum_{s=1}^{n} K_{b_n}\left(\frac{s-j}{n}\right) \tilde{e}_{s,m}/(nb_n)$ and

$$\bar{A}_{n,i}^{(m)} = 2 \sum_{j=\lfloor nb_n \rfloor}^{i} a_{n,j}^{(m)},$$

then a similar argument as given in the proof of Theorem 1 of Zhou (2014) shows that

$$\max_{1 \leq j \leq n} \left\| \sum_{s=1}^{n} K_{b_n}\left(\frac{s-j}{n}\right)(\tilde{e}_{s,m} - e_s) \right\|_4 \leq C\sqrt{nb_n} m \chi^m$$

for some constant $\chi \in (0,1)$. By the Cauchy-Schwartz inequality it follows that

$$\left\| \max_{\lfloor nb_n \rfloor \leq i \leq n - \lfloor nb_n \rfloor} |\bar{A}_{n,i} - \bar{A}_{n,i}^{(m)}| \right\|_2 \leq \left\| \frac{2}{nb_n} \sum_{j=\lfloor nb_n \rfloor}^{n - \lfloor nb_n \rfloor} |e_j| \sum_{s=1}^{n} K_{b_n}\left(\frac{s-j}{n}\right)(\tilde{e}_{s,m} - e_s)| \right\|_2$$
$$= O(\sqrt{n} m \chi^m b_n^{-1/2}). \tag{7.6}$$

With the notations $\tilde{a}_{n,j}^{(m)} = \tilde{e}_{j,m} \sum_{s=1}^{n} K_{b_n}\left(\frac{s-j}{n}\right) \tilde{e}_{s,m}/(nb_n)$ and $\tilde{A}_{n,i}^{(m)} = 2 \sum_{j=\lfloor nb_n \rfloor}^{i} \tilde{a}_{n,j}^{(m)}$ it is easy to see that

$$\left\| \max_{\lfloor nb_n \rfloor \leq i \leq n - \lfloor nb_n \rfloor} |\tilde{A}_{n,i}^{(m)} - \bar{A}_{n,i}^{(m)}| \right\|_2 \leq \frac{2}{nb_n} \sum_{j=1}^{n} \|e_j - \tilde{e}_{j,m}\|_4 \left\| \sum_{s=1}^{n} K_{b_n}\left(\frac{s-j}{n}\right) \tilde{e}_{s,m} \right\|_4. \tag{7.7}$$

Now an elementary calculation via Burkholder's inequality shows

$$\max_{1 \leq j \leq n} \left\| \frac{1}{nb_n} \sum_{s=1}^{n} K_{b_n}\left(\frac{s-j}{n}\right) \tilde{e}_{s,m} \right\|_4 = O\left(\frac{1}{\sqrt{nb_n}}\right),$$

and by a similar argument as given in the proof of Theorem 1 of Zhou (2014) we have for some constant $\chi \in (0,1)$ the estimate $\max_{1 \leq j \leq n} \|\tilde{e}_{j,m} - e_j\|_4 = O(\chi^m)$. This gives for the left-hand

side of (7.7)

$$\left\|\max_{\lfloor nb_n\rfloor \leq i \leq n-\lfloor nb_n\rfloor} |\bar{A}_{n,i}^{(m)} - \tilde{A}_{n,i}^{(m)}|\right\|_2 = O(\sqrt{n/b_n}\chi^m),$$

and an application of (7.6) yields

$$\left\|\max_{\lfloor nb_n\rfloor \leq i \leq n-\lfloor nb_n\rfloor} |\bar{A}_{n,i} - \tilde{A}_{n,i}^{(m)}|\right\|_2 = O(\sqrt{n/b_n}m\chi^m). \tag{7.8}$$

A tedious but straightforward calculation shows that $\mathcal{P}_{j-l}(\tilde{e}_{j,m}\tilde{e}_{i,m}) = 0$ for $l > 2m$. For example, if $i \geq j - m$, then by definition, $\tilde{e}_{j,m}\tilde{e}_{i,m}$ is $\sigma(\varepsilon_{j-2m}, \varepsilon_{j-2m+1}, ..., \varepsilon_i)$ measurable. Consequently, $\mathbb{E}(\tilde{e}_{j,m}\tilde{e}_{i,m}|\mathcal{F}_{j-l}) = \mathbb{E}(\tilde{e}_{j,m}\tilde{e}_{i,m}|\mathcal{F}_{j-l-1}) = \mathbb{E}(\tilde{e}_{j,m}\tilde{e}_{i,m})$ if $l > 2m$, which gives $\mathcal{P}_{j-l}(\tilde{e}_{j,m}\tilde{e}_{i,m}) = 0$. The other cases $i \leq j - l - 1$ and $j - l \leq i \leq j - m - 1$ are treated similarly, and details are omitted for the sake of brevity. Observing $\mathcal{P}_{j-l}(\tilde{e}_{j,m}\tilde{e}_{i,m}) = 0$ for $l > 2m$ we obtain

$$\left\|\max_{\lfloor nb_n\rfloor \leq i \leq n-\lfloor nb_n\rfloor} |\tilde{A}_{n,i}^{(m)} - \mathbb{E}\tilde{A}_{n,i}^{(m)}|\right\|_2 \leq 2\sum_{l=0}^{2m}\left\|\max_{\lfloor nb_n\rfloor \leq i \leq n-\lfloor nb_n\rfloor} |\sum_{j=\lfloor nb_n\rfloor}^{i}\mathcal{P}_{j-l}\tilde{a}_{n,j}^{(m)}|\right\|_2. \tag{7.9}$$

Similar arguments as given in the proof of Theorem 1 in Wu (2005) show

$$\|\mathcal{P}_{j-l}\tilde{a}_{n,j}^{(m)}\|_2 \leq \frac{M}{n}\left\|\tilde{e}_{j,m}\sum_{s=1}^{n}\tilde{e}_{s,m}K_{b_n}\left(\frac{s-j}{n}\right) - \tilde{e}_{j,m}^{(j-l)}\sum_{s=1}^{n}\tilde{e}_{s,m}^{(j-l)}K_{b_n}\left(\frac{s-j}{n}\right)\right\|_2,$$

and by the triangle inequality it follows that

$$\|\mathcal{P}_{j-l}\tilde{a}_{n,j}^{(m)}\|_2 \leq M(Z_{1,j} + Z_{2,j}),$$

where the terms $Z_{1,j}$ and $Z_{2,j}$ are defined by

$$Z_{1,j} = \frac{1}{nb_n}\left\|\tilde{e}_{j,m}\sum_{s=1}^{n}K_{b_n}\left(\frac{s-j}{n}\right)\left[\tilde{e}_{s,m}^{(j-l)} - \tilde{e}_{s,m}\right]\right\|_2,$$

$$Z_{2,j} = \frac{1}{nb_n}\left\|\left[\tilde{e}_{j,m}^{(j-l)} - \tilde{e}_{j,m}\right]\sum_{s=1}^{n}K_{b_n}\left(\frac{s-j}{n}\right)\tilde{e}_{s,m}^{(j-l)}\right\|_2,$$

$\tilde{e}_{s,m}^{(j)} = \mathbb{E}(e_s^{(j)}|\varepsilon_{s-m}, \ldots, \varepsilon_j', \ldots, \varepsilon_s)$ for $s - m \leq j \leq s$, $e_s^{(j)} = G_l(t_s, \mathcal{F}_s^{(j)})$ for $b_l < t_s \leq b_{l+1}$ and we use the convention $\tilde{e}_{s,m}^{(j)} = \tilde{e}_{s,m}$ for $j < s - m$ or $j > s$. Elementary calculations show that for

39

$l \geq 0$

$$\left\| \sum_{s=1}^{n} K_{b_n}\left(\frac{s-j}{n}\right)\tilde{e}_{s,m}^{(j-l)} \right\|_4 = O(\sqrt{nb_n}), \qquad 1 \leq j \leq n,$$

while by definition $\|\tilde{e}_{j,m}^{(j-l)} - \tilde{e}_{j,m}\|_4 = 0$ for $l > m$. On the other hand, if $1 \leq j \leq n$, $0 \leq l \leq m$, we have by Assumption (A4)

$$\|\tilde{e}_{j,m}^{(j-l)} - \tilde{e}_{j,m}\|_4 = \left\| \mathbb{E}(e_j - e_j^{(j-l)}|\varepsilon_{j-m}, ..., \varepsilon_{j-l}, \varepsilon'_{j-l}, \varepsilon_j) \right\|_4 \leq M\chi^l,$$

which gives $Z_{2,j} = O(\frac{\chi^l}{\sqrt{nb_n}})$. Observing that $\tilde{e}_{s,m}^{(j-l)} - \tilde{e}_{s,m} = 0$ if $s \geq j-l+m+1$ or $s \leq j-l-1$, it is easy to see that $Z_{1,j} = O(\frac{m}{nb_n})$. It now follows from Doob's inequality

$$\left\| \max_{\lfloor nb_n \rfloor \leq i \leq n - \lfloor nb_n \rfloor} \left| \sum_{j=\lfloor nb_n \rfloor}^{i} \mathcal{P}_{j-l}\tilde{a}_{n,j}^{(m)} \right| \right\|_2 = O\left(\sqrt{n}\left(\frac{\chi^l}{\sqrt{nb_n}} + \frac{m}{nb_n}\right)\right),$$

and we obtain from (7.9) that

$$\left\| \max_{\lfloor nb_n \rfloor \leq i \leq n - \lfloor nb_n \rfloor} |\tilde{A}_{n,i}^{(m)} - \mathbb{E}\tilde{A}_{n,i}^{(m)}| \right\|_2 = O\left(\frac{m^2}{n^{1/2}b_n} + (b_n)^{-1/2}\right). \tag{7.10}$$

Finally, similar arguments as given in the proof of Lemma 5 in Zhou and Wu (2010) show

$$\max_{\lfloor nb_n \rfloor \leq i \leq n - \lfloor nb_n \rfloor} \mathbb{E}[\tilde{A}_{i,m}] = O\left(\sum_{i=1}^{n}\sum_{j=1}^{n} \chi^{|i-j|}/(nb_n)\right) = O(b_n^{-1}).$$

Observing (7.5), (7.8) and (7.10) and taking $m = M\log n$ for a sufficiently large constant $M > 0$ yields $\max_{1 \leq i \leq n} |A_{n,i}| = O_p(\sqrt{nb_n} + nb_n^3 + b_n^{-1})$. Consequently, the assertion (7.1) follows from (7.2), (7.3) and this estimate.

7.1.2 Proof of Theorem 3.2

We recall the definition of $\hat{\Phi}_{i,n}$ in (3.6) and define on the interval $[0,1]$ the linear interpolation

$$\hat{\tilde{\Phi}}_{m,n}(t) = \hat{\Phi}_{\lfloor nt \rfloor,m} + (nt - \lfloor nt \rfloor)(\hat{\Phi}_{\lfloor nt \rfloor+1,m} - \hat{\Phi}_{\lfloor nt \rfloor,m}). \tag{7.11}$$

The assertion follows if the weak convergence

$$\{\hat{\tilde{\Phi}}_{m,n}(t)\}_{t\in[0,1]} \Rightarrow \{U_1(t)\}_{t\in[0,1]}$$

40

conditional on $\mathcal{F}_n$ can be established. For a proof of this statement define $\Phi_{i,m}$ and $\tilde{\Phi}_{m,n}(t)$ by replacing the nonparametric residuals $\{\hat{e}_i\}_{i=1}^n$ by the (non-observable) errors $\{e_i\}_{i=1}^n$ in the definition (3.6) and (7.11) of $\hat{\Phi}_{i,m}$ and $\hat{\tilde{\Phi}}_{m,n}(t)$, respectively. Note that similar arguments as given in the proof of Theorem 3 in Zhou (2013) show that $\{\tilde{\Phi}_{m,n}(t)\}_{t\in[0,1]} \Rightarrow \{U_1(t)\}_{t\in[0,1]}$. The assertion of Theorem 3.2 then follows from the estimate

$$\sup_{t\in[0,1]} \left|\tilde{\Phi}_{m,n}(t) - \hat{\tilde{\Phi}}_{m,n}(t)\right| = O_p\left(\left(\frac{m\log^2 n}{nb_n^{3/2}}\right)^{1/2} + \sqrt{m}b_n^2 \log n\right). \tag{7.12}$$

In order to prove (7.12) let C denote a sufficiently large constant, which may vary from line to line in the following calculations, and consider the event

$$A_n = \left\{ \sup_{t\in[0,1]} |\hat{\mu}_{b_n}(t) - \mu(t)| \le C\left(\frac{\log n}{\sqrt{nb_n}b_n^{1/4}} + b_n^2 \log n\right)\right\}.$$

By Lemma 8.3 of Section 8.1 in the technical appendix we have that $\lim_{n\to\infty}\mathbb{P}(A_n) = 1$. This yields for $1 \le j \le n - m + 1$ the estimate

$$\mathbb{E}[(S_{j,m} - \hat{S}_{j,m})^2 I(A_n)] \le \mathbb{E}\left[\sum_{r=j}^{j+m-1}(e_r - \hat{e}_r)^2 \sum_{r=j}^{j+m-1}(2e_r + \hat{e}_r - e_r)^2 I(A_n)\right]$$

$$\le C\left(\frac{m^2 \log^2 n}{nb_n^{3/2}} + m^2 b_n^4 \log^2 n\right).$$

Similarly, it follows that $\mathbb{E}[(S_n - \hat{S}_n)^2 \frac{m^2}{n^2} I(A_n)] \le C(\frac{m^2 \log^2 n}{nb_n^{3/2}} + m^2 b_n^4 \log^2 n)$, which gives

$$\|(\Phi_{n-m+1} - \hat{\Phi}_{n-m+1})I(A_n)\|_2^2 = \frac{1}{m(n-m+1)} \sum_{j=1}^{n-m+1} \mathbb{E}\left[\left(S_{j,m} - \hat{S}_{j,m} - \frac{m}{n}(S_n - \hat{S}_n)\right)^2 I(A_n)\right]$$

$$\le C\left(\frac{m\log^2 n}{nb_n^{3/2}} + mb_n^4 \log^2 n\right).$$

An application of Doob's inequality and Proposition 8.2 in Section 8.2 finally yields

$$\max_{1\le i\le n-m+1} |\Phi_{i,m} - \hat{\Phi}_{i,m}| = O_p\left(\left(\frac{m\log^2 n}{nb_n^{3/2}}\right)^{1/2} + \sqrt{m}b_n^2 \log n\right).$$

The estimate (7.12) now follows from this result and definition (7.11), which completes the proof of Theorem 3.2.

7.1.3 Proof of Lemma 3.1

Define $\mathcal{N}(i) = \frac{1}{L}\big(\sum_{j=i-L+1}^{i} e_j^2 - \sum_{j=i}^{i+L-1} e_j^2\big)$ and recall the definition of $\mathcal{M}(i)$ in (3.11). By similar arguments as given in the proof of Lemma 8.3 in the technical appendix (note that $\iota > 8$) we have $\|\mathcal{M}(i) - \mathcal{N}(i)\|_4 = b_n^2 + \frac{1}{\sqrt{nb_n}}$, and Proposition 8.1 yields

$$\max_{L \leq i \leq n-L+1} |\mathcal{M}(i) - \mathcal{N}(i)| = O_p\Big(n^{1/4}b_n^2 + \frac{1}{n^{1/4}b_n^{1/2}}\Big). \tag{7.13}$$

Consider the case that $i \in B := \{i : |t_i - \tilde{t}_v| > L\}$. Then by our assumption on the variance function, there exists a large constant C, such that $|\mathbb{E}\mathcal{N}(i)| \leq CL/n$ for $L \leq i \leq n-L+1$, $i \in B$. By Lemma 8.3 and Lemma 8.4 in the technical appendix it now follows $\|\mathcal{N}(i) - \mathbb{E}\mathcal{N}(i)\|_{\iota/2} \leq CL^{-1/2}$ ($L \leq i \leq n - L + 1$, $i \in B$), which gives

$$\max_{L \leq i \leq n-L+1, i \in B} |\mathcal{N}(i)| = O_p(L^{-1/2}n^{2/\iota} + L/n).$$

Combining this estimate with (7.13) yields

$$\max_{L \leq i \leq n-L+1, i \in B} |\mathcal{M}(i)| = O_p\Big(L^{-1/2}n^{2/\iota} + L/n + n^{1/4}b_n^2 + \frac{1}{n^{1/4}b_n^{1/2}}\Big).$$

Similarly, we can show that $\mathcal{M}(\lfloor n\tilde{t}_v \rfloor) = \sigma(t_v^+) - \sigma(t_v^-) + O_p\big(n^{1/4}b_n^2 + \frac{1}{n^{1/4}b_n^{1/2}} + L^{-1/2} + L/n\big)$. The choice of L implies that

$$\lim_{n\to\infty} \mathbb{P}\Big(|\mathcal{M}(\lfloor n\tilde{t}_v \rfloor)| > \max_{L \leq i \leq n-L+1, i \in B} |\mathcal{M}(i)|\Big) = 1,$$

which completes the proof of Lemma 3.1. □

7.1.4 Proof of Theorem 3.3

We restrict ourselves to the case of a variance function with no change point and the corresponding estimator (3.9). The statement for the estimator (3.12) follows by similar arguments as given in Theorem 4.3, where we deal with the problem of testing for relevant changes in the correlation.

Recall the definition (3.15), define $S_i^W = \sum_{j=1}^{i} W_j^k$ as the corresponding partial sum and consider the CUSUM statistic $T_n^c = \max_{1 \leq i \leq n} |S_i^W - \frac{i}{n}S_n^W|$, We will show the estimate

$$\max_{1 \leq i \leq n} |\hat{S}_i^W - S_i^W| = O_p(nc_n^2 + nb_n^3 c_n^{-1/4} + b_n^{-1}c_n^{-1}), \tag{7.14}$$

which implies $\hat{T}_n^c - T_n^c = O_p(nc_n^2 + nb_n^3 c_n^{-1/4} + b_n^{-1} c_n^{-1})$. It follows from Zhou (2013) that $T_n^c/\sqrt{n}$ converges weakly to the distribution of the random variable $\mathcal{K}_2$ defined in Theorem 3.3. By our choice of the bandwidth b_n we have $nc_n^2 + nb_n^3 c_n^{-1/4} + b_n^{-1} c_n^{-1} = o(\sqrt{n})$, and the assertion of Theorem 3.3 follows.

For the sake of simplicity we omit in the subscripts c_n, b_n in the variance estimator $\hat{\sigma}_{c_n, b_n}$ and the superscript k in the definition $\hat{W}_i^k, W_i^k$ the proof of the estimate (7.14). With the notation $\tilde{W}_i = \frac{e_i e_{i+k}}{\sigma^2(t_i)}$ we obtain

$$\max_{1 \le j \le n} \left| \sum_{i=1}^{j} \left(W_i - \tilde{W}_i \right) \right| \le \max_{1 \le j \le n} \sum_{i=1}^{j} \frac{|e_i e_{i+k}| \cdot |\sigma(t_i) - \sigma(t_{i+k})|}{\sigma^2(t_i)\sigma(t_{i+k})} = O_p(1), \tag{7.15}$$

where we have used the fact that the variance function is Lipschitz continuous. Let $\bar{W}_i = \frac{\hat{e}_i \hat{e}_{i+k}}{\sigma^2(t_i)}$ denote the analogue of $\hat{W}_i$, where the estimate $\hat{\sigma}^2(t_i)$ has been replaced by the "true" variance $\sigma^2(t_i)$. By a careful inspection of the proof of Theorem 3.1, it can be seen that

$$\max_{1 \le j \le n} \left| \sum_{i=1}^{j} \left(\bar{W}_i - \tilde{W}_i \right) \right| = O_p(\sqrt{nb_n} + nb_n^3 + b_n^{-1}). \tag{7.16}$$

Define

$$\Lambda_j := \sum_{i=1}^{j} (\hat{W}_i - \bar{W}_i) = \sum_{i=1}^{j} \frac{\hat{e}_i \hat{e}_{i+k}(-\hat{\sigma}^2(t_i) + \sigma^2(t_i))}{\hat{\sigma}^2(t_i)\sigma^2(t_i)},$$

then our next goal is to estimate $\max_{1 \le j \le n} |\Lambda_j|$. For this purpose we consider the random variable

$$\bar{\Lambda}_j := \sum_{i=1}^{j} \frac{\hat{e}_i \hat{e}_{i+k}(-\hat{\sigma}^2(t_i) + \sigma^2(t_i))}{\sigma^4(t_i)}$$

(here the estimator in the denominator has been replaced by the true variance function), and obtain

$$\max_{1 \le j \le n} |\Lambda_j - \bar{\Lambda}_j| \le \sum_{i=1}^{n} \frac{|\hat{e}_i \hat{e}_{i+k}|(\hat{\sigma}^2(t_i) - \sigma^2(t_i))^2}{\hat{\sigma}^2(t_i)\sigma^4(t_i)}. \tag{7.17}$$

For the expectation of the right-hand side it follows

$$\mathbb{E}\left[\sum_{i=1}^{n} \frac{|\hat{e}_i \hat{e}_{i+k}|(\hat{\sigma}^2(t_i) - \sigma^2(t_i))^2}{\hat{\sigma}^2(t_i)\sigma^4(t_i)} \right] \le C \sum_{i=1}^{n} \|\hat{e}_i\|_4 \|\hat{e}_{i+k}\|_4 \|(\hat{\sigma}^2(t_i) - \sigma^2(t_i))^2\|_2. \tag{7.18}$$

43

By Lemma 8.3 of Section 8.1 in the technical appendix we have that

$$\|\hat{\mu}_{b_n}(t) - \mu(t)\|_4 = O\Big(b_n^2 + \frac{1}{\sqrt{nb_n}}\Big), \tag{7.19}$$

which implies $\|\hat{e}_i\|_4 \leq C$. On the other hand, Corollary 8.1 in Section 8.1 shows

$$\max_{1\leq i\leq n} \|(\hat{\sigma}^2(t_i) - \sigma^2(t_i))^2\|_2 = O\Big(b_n^4 + \frac{1}{nb_n} + c_n^4 + \frac{1}{nc_n}\Big), \tag{7.20}$$

and we obtain from (7.17), (7.18) and Proposition 8.1 in Section 8.2 the estimate

$$\max_{1\leq j\leq n}|\Lambda_j| \leq \max_{1\leq j\leq n}|\bar{\Lambda}_j| + \max_{1\leq j\leq n}|\Lambda_j - \bar{\Lambda}_j| = \max_{1\leq j\leq n}|\bar{\Lambda}_j| + O_p(nb_n^4 + b_n^{-1} + nc_n^4 + c_n^{-1}). \tag{7.21}$$

Now the remaining problem is to derive an appropriate estimate for the quantity $\max_{1\leq j\leq n}|\bar{\Lambda}_j|$. For this purpose note that $\bar{\Lambda}_j = \bar{\lambda}_{j,1} + \bar{\lambda}_{j,2}$, where

$$\bar{\lambda}_{j,1} = \sum_{i=1}^{j} \frac{(\hat{e}_i\hat{e}_{i+k} - e_ie_{i+k})(\sigma^2(t_i) - \hat{\sigma}^2(t_i))}{\sigma^4(t_i)},$$

$$\bar{\lambda}_{j,2} = \sum_{i=1}^{j} \frac{e_ie_{i+k}(\sigma^2(t_i) - \hat{\sigma}^2(t_i))}{\sigma^4(t_i)}.$$

By Lemma 8.1, Corollary 8.1 of Section 8.1 and the estimate (7.19) it is easy to see that

$$\mathbb{E}\Big[\max_{1\leq j\leq n}|\bar{\lambda}_{j,1}|\Big] \leq \sum_{i=1}^{n} \frac{\|\hat{e}_i\hat{e}_{i+k} - e_ie_{i+k}\|_2}{\sigma^4(t_i)}\|\sigma^2(t_i) - \hat{\sigma}^2(t_i)\|_2 = O(\underline{\pi}_n), \tag{7.22}$$

$$\mathbb{E}\Big[\max_{1\leq j\leq \lfloor nb_n+nc_n\rfloor}|\bar{\lambda}_{j,2}|\Big] \leq \sum_{i=1}^{\lfloor nb_n+nc_n\rfloor} \frac{\|e_ie_{i+k}\|_2}{\sigma^4(t_i)}\|\sigma^2(t_i) - \hat{\sigma}^2(t_i)\|_2 = O(\pi_n), \tag{7.23}$$

$$\max_{n-\lfloor nb_n+nc_n\rfloor\leq j\leq n}|\bar{\lambda}_{j,2}| \leq |\bar{\lambda}_{n-\lfloor nb_n+nc_n\rfloor-1,2}| + \sum_{i=n-\lfloor nb_n+nc_n\rfloor}^{n} \frac{|e_ie_{i+k}|}{\sigma^4(t_i)}|\sigma^2(t_i) - \hat{\sigma}^2(t_i)|$$

$$\leq \max_{\lfloor nb_n+nc_n\rfloor\leq j\leq n-\lfloor nb_n+nc_n\rfloor-1}|\bar{\lambda}_{j,2}| + \sum_{i=n-\lfloor nb_n+nc_n\rfloor}^{n} \frac{|e_ie_{i+k}|}{\sigma^4(t_i)}|\sigma^2(t_i) - \hat{\sigma}^2(t_i)|$$

$$= \max_{\lfloor nb_n+nc_n\rfloor\leq j\leq n-\lfloor nb_n+nc_n\rfloor-1}|\bar{\lambda}_{j,2}| + O_p(\pi_n). \tag{7.24}$$

where the constants $\underline{\pi}_n$ and π_n are given by $\underline{\pi}_n = nb_n^2c_n^2 + \sqrt{\frac{n}{c_n}}b_n^2 + \sqrt{\frac{n}{b_n}}c_n^2 + \frac{1}{\sqrt{b_nc_n}}$, $\pi_n = (nb_n + nc_n)(b_n^2 + c_n^2 + \frac{1}{\sqrt{nb_n}} + \frac{1}{\sqrt{nc_n}})$, respectively.

In order to prove a corresponding estimate for the remaining term $\max_{\lfloor nb_n+nc_n\rfloor\leq j\leq n-\lfloor nb_n+nc_n\rfloor}|\bar{\lambda}_{j,2}|$

in (7.24) we study the asymptotic behavior of the quantity $\hat{\sigma}^2(t) - \sigma^2(t)$. By Lemma 8.4 in Section 8.1 it easily follows that

$$\sup_{t \in \mathfrak{T}_n} \left| \hat{\sigma}^2(t) - \sigma^2(t) - \frac{\mu_2 \ddot{\sigma}^2(t) c_n^2}{2} - \frac{1}{nc_n} \sum_{i=1}^{n} K_{c_n}(t_i - t)(\hat{e}_i^2 - \mathbb{E}(e_i^2)) \right| = O\left(c_n^3 + \frac{1}{nc_n} \right). \quad (7.25)$$

We now consider the decomposition

$$\sum_{i=1}^{n} K_{c_n}(t_i - t) \left(\hat{e}_i^2 - \mathbb{E}(e_i^2) - (e_i^2 - \mathbb{E}(e_i^2)) \right) = \sum_{i=1}^{n} K_{c_n}(t_i - t) Q_i,$$

where $Q_i = Q_{1,i} + Q_{2,i}$, $Q_{1,i} = 2e_i[\mu(t_i) - \hat{\mu}(t_i)]$, $Q_{2,i} = [\mu(t_i) - \hat{\mu}(t_i)]^2$. By Lemma 8.1 in Section 8.1 we obtain

$$\sup_{\lfloor nb_n \rfloor \leq i \leq n - \lfloor nb_n \rfloor} \left| \hat{\mu}_{b_n}(t_i) - \mu(t_i) - \frac{\mu_2 \ddot{\mu}(t_i)}{2} b_n^2 - \frac{1}{nb_n} \sum_{j=1}^{n} e_j K_{b_n}(t_j - t_i) \right| = O(b_n^3 + \frac{b_n}{n}).$$

The triangle inequality and Proposition 8.1 in Section 8.2 imply

$$\left\| \sup_{t \in \mathfrak{T}_n''} \left| \sum_{i=1}^{n} K_{c_n}(t_i - t) \left[Q_{1,i} - \frac{2e_i}{nb_n} \sum_{j=1}^{n} e_j K_{b_n}(t_i - t_j) - \mu_2 \ddot{\mu}(t_i) b_n^2 e_i \right] \right| \right\|_4 = O(nb_n^3 c_n^{3/4}), \quad (7.26)$$

where we use the notation $\mathfrak{T}_n'' = [b_n + c_n, 1 - b_n - c_n]$. Similar arguments as given in the calculation of $\max_{\lfloor nb_n \rfloor \leq i \leq n - \lfloor nb_n \rfloor} |A_{n,i}|$ in the proof of Lemma 7.1 and the summation by parts formula show

$$\left\| \sup_{t \in \mathfrak{T}_n''} \frac{2}{nb_n} \left| \sum_{i=1}^{n} K_{c_n}(t_i - t) e_i \sum_{j=1}^{n} e_j K_{b_n}(t_i - t_j) \right| \right\|_2 = O(b_n^{-1}),$$

$$\left\| \sup_{t \in \mathfrak{T}_n''} \left| \sum_{i=1}^{n} K_{c_n}(t_i - t) \mu_2 \ddot{\mu}(t_i) b_n^2 e_i \right| \right\|_2 = O(n^{1/2} b_n^2),$$

and (7.26) gives $\left\| \sup_{t \in \mathfrak{T}_n''} \left| \sum_{i=1}^{n} K_{b_n}(t_i - t) Q_{1,i} \right| \right\|_2 = O(nb_n^3 c_n^{3/4} + b_n^{-1} + n^{1/2} b_n^2)$. On the other hand, note that

$$\left\| \sup_{t \in \mathfrak{T}_n''} \left| \sum_{i=1}^{n} K_{c_n}(t_i - t) Q_{2,i} \right| \right\|_2 \leq R_{n,1} + R_{n,2}$$

45

where

$$R_{n,1} = \left\| \sup_{t \in \mathfrak{T}_n''} \left| \sum_{i=1}^n K_{c_n}(t_i - t) \left(\frac{1}{nb_n} \sum_{j=1}^n e_j K_{b_n}(t_i - t_j) + \frac{\mu_2 \ddot{\mu}(t_i)}{2} b_n^2 \right)^2 \right| \right\|_2$$

$$R_{n,2} = \left\| \sup_{t \in \mathfrak{T}_n''} \left| \sum_{i=1}^n K_{c_n}(t_i - t) \left(\mu(t_i) - \hat{\mu}(t_i) + \frac{1}{nb_n} \sum_{j=1}^n e_j K_{b_n}(t_i - t_j) + \frac{\mu_2 \ddot{\mu}(t_i)}{2} b_n^2 \right) \right. \right.$$
$$\left. \left. \times \left(\mu(t_i) - \hat{\mu}(t_i) - \frac{1}{nb_n} \sum_{j=1}^n e_j K_{b_n}(t_i - t_j) - \frac{\mu_2 \ddot{\mu}(t_i)}{2} b_n^2 \right) \right| \right\|_2.$$

Proposition 8.1 in Section 8.2 and similar calculations as given in the proof of Lemma 7.1 show that

$$R_{n,1} = O\left(nc_n c_n^{-1/2} \left(\frac{1}{nb_n} + b_n^4 \right) \right) = O(c_n^{1/2} b_n^{-1} + nc_n^{1/2} b_n^4),$$

while a further application of Lemma 8.1 in Section 8.1 yields

$$R_{n,2} = O\left(\frac{nb_n^3 c_n}{\sqrt{nb_n}} c_n^{-1/2} + nb_n^5 c_n^{1/2} \right) = O(\sqrt{n} b_n^{5/2} c_n^{1/2} + nb_n^5 c_n^{1/2}). \tag{7.27}$$

Consequently, combining the arguments in (7.25)-(7.27), it follows that

$$\left\| \sup_{t \in \mathfrak{T}_n''} \left| \hat{\sigma}^2(t) - \sigma^2(t) - \frac{\mu_2 \ddot{\sigma}^2(t) c_n^2}{2} - \frac{1}{nc_n} \sum_{i=1}^n K_{b_n}(t_i - t) \left(e_i^2 - \mathbb{E}(e_i^2) \right) \right| \right\|_2 = O(\bar{\pi}_n), \tag{7.28}$$

where $\bar{\pi}_n = c_n^3 + \frac{1}{nc_n} + b_n^3 c_n^{-1/4} + \frac{1}{nb_n c_n} + \frac{b_n^2}{\sqrt{nc_n}} + c_n^{-1/2} b_n^{-1} n^{-1} + c_n^{-1/2} b_n^4 + b_n^{5/2}(nc_n)^{-1/2} + b_n^5 c_n^{-1/2}$.
Recall the definition of $\tilde{W}_i$, define $Z_i' = e_i^2 - \mathbb{E}e_i^2$, then it follows from (7.28) that

$$\mathbb{E} \left(\max_{\lfloor nb_n + nc_n \rfloor \leq j \leq n - \lfloor nb_n + nc_n \rfloor} \left| \bar{\lambda}_{j,2} + \sum_{i=1}^j \frac{\tilde{W}_i \left(\sum_{j=1}^n K_{c_n}(t_j - t_i) Z_i' + \mu_2 \ddot{\sigma}^2(t_i) nc_n^3/2 \right)}{\sigma^2(t_i) nc_n} \right| \right) \leq$$

$$\sum_{i=\lfloor nb_n + nc_n \rfloor}^{n - \lfloor nb_n + nc_n \rfloor} \|e_i e_{i+k}\|_2 \left\| \frac{\hat{\sigma}^2(t_i) - \sigma^2(t_i) - \frac{1}{nc_n} \sum_{j=1}^n K_{c_n}(t_j - t_i) Z_i' - \mu_2 \ddot{\sigma}^2(t_i) c_n^2/2}{\sigma^4(t_i)} \right\|_2 \tag{7.29}$$

$$= O(n\bar{\pi}_n).$$

By the Cauchy-Schwarz inequality we obtain $\|\tilde{W}_i - \tilde{W}_i^{(m)}\|_4 = O(\chi^{|i-m|}), \quad \|Z_i' - Z_i'^{(m)}\|_4 = $

46

$O(\chi^{|i-m|})$, where $Z_i'^{(m)} = (e_i^{(m)})^2 - \mathbb{E}(e_i^{(m)})^2$, $\tilde{W}_i^{(m)} = \frac{e_i^{(m)} e_{i+k}^{(m)}}{\sigma(t_i)^2}$, and

$$e_i^{(m)} = G_j(t_i, \mathcal{F}_i^{(m)}), \text{ if } b_j < t_i \leq b_{j+1}.$$

Hence, with similar arguments as given in the proof of Lemma 5 of Zhou and Wu (2010) we get

$$\max_{\lfloor nb_n+nc_n \rfloor \leq j \leq n - \lfloor nb_n+nc_n \rfloor} \mathbb{E}\left[\sum_{i=\lfloor nb_n+nc_n \rfloor}^{j} \frac{\tilde{W}_i \sum_{j=1}^{n} K_{c_n}(t_j - t_i) Z_j'}{\sigma^2(t_i) n c_n} \right] = O(c_n^{-1}).$$

Then by a similar m-dependent approximating technique as given in the proof of Lemma 7.1 we get

$$\max_{\lfloor nb_n+nc_n \rfloor \leq j \leq n - \lfloor nb_n+nc_n \rfloor} \left| \sum_{i=\lfloor nb_n+nc_n \rfloor}^{j} \frac{\sum_{j=1}^{n} \tilde{W}_i K_{c_n}(t_j - t_i) Z_j' - \mathbb{E}[\tilde{W}_i K_{c_n}(t_j - t_i) Z_j']}{\sigma^2(t_i) n c_n} \right| = O_p(c_n^{-1}).$$

Similarly, and more easily one obtains

$$\max_{\lfloor nb_n+nc_n \rfloor \leq j \leq n - \lfloor nb_n+nc_n \rfloor} \left| \sum_{i=1}^{j} e_i e_{i+k} \mu_2 \ddot{\sigma}^2(t_i) b_n^2 / (2\sigma^4(t_i)) \right| = O_p(nc_n^2). \tag{7.30}$$

Hence, it follows from (7.29) and (7.30) that $\max_{\lfloor nb_n+nc_n \rfloor \leq j \leq n - \lfloor nb_n+nc_n \rfloor} |\bar{\lambda}_{j,2}| = O_p(n\bar{\pi}_n + nc_n^2)$, which implies, observing (7.22) - (7.24),

$$\max_{\lfloor nb_n+nc_n \rfloor \leq j \leq n - \lfloor nb_n+nc_n \rfloor} |\bar{\Lambda}_j| = O_p(\pi_n + \underline{\pi}_n + n\bar{\pi}_n + nc_n^2).$$

Combining this result with the estimates (7.15), (7.16) and (7.21), and by our choice of the bandwidths, we have that

$$\max_{1 \leq j \leq n} \left| \sum_{i=1}^{j} (\hat{W}_i - W_i) \right| = O_p(nc_n^2 + nb_n^3 c_n^{-1/4} + b_n^{-1} c_n^{-1}),$$

which establishes the estimate (7.14) and completes the proof of Theorem 3.3.

7.1.5 Proof of Theorem 3.4

We restrict ourselves to the case of a variance function with no change point and the corresponding estimator (3.9). The statement for the estimator (3.12) follows by similar arguments as given in Theorem 4.4, where we deal with the problem of testing for relevant changes in the

correlation.

Recall the definition of $\hat{W}_i$ and W_i in (3.13) and (3.15) and introduce the notation $\tilde{W}_i = \frac{e_i e_{i+k}}{\sigma^2(t_i)}$ (again the superscript is omitted in our notation). We consider the corresponding partial sums $S_{j,m} = \sum_{r=j}^{j+m-1} W_r$, $\tilde{S}_{j,m} = \sum_{r=j}^{j+m-1} \tilde{W}_r$ and $\hat{S}_{j,m}^W = \sum_{r=j}^{j+m-1} \hat{W}_r$ and define $S_n = \sum_{r=1}^{n} W_r$, $\tilde{S}_n = \sum_{r=1}^{n} \tilde{W}_r$, $\hat{S}_n^W = \sum_{r=1}^{n} \hat{W}_r$. Similarly, define $(\Phi_{i,m}, \tilde{\Phi}_{m,n}(t))$, $(\Phi_{i,m}^o, \tilde{\Phi}_{m,n}^o(t))$ and $(\hat{\Phi}_{i,m}^W, \hat{\tilde{\Phi}}_{m,n}^W(t))$ by replacing $(\hat{S}_n, \hat{S}_{j,m})$ in the definitions (3.6), (7.11) of $\hat{\Phi}_{i,m}$ and $\hat{\tilde{\Phi}}_{m,n}(t)$ by $(S_n, S_{j,m})$, $(\tilde{S}_n, \tilde{S}_{j,m})$ and $(\hat{S}_n^W, \hat{S}_{j,m}^W)$, respectively. It then follows from the results Zhou (2013) that $\{\tilde{\Phi}_{m,n}(t)\}_{t\in[0,1]} \Rightarrow \{U_2(t)\}_{t\in[0,1]}$, where $\{U_2(t)\}_{t\in[0,1]}$ is a centered Gaussian process with mean 0 and covariance kernel (3.16). The assertion of Theorem 3.4 is now a consequence of the estimate

$$\sup_{t\in[0,1]} |\tilde{\Phi}_{m,n}(t) - \hat{\tilde{\Phi}}_{m,n}(t)| = O_p(\sqrt{m}\delta_n), \tag{7.31}$$

where $\delta_n = \left(c_n^2 + (\frac{1}{\sqrt{nc_n}} + b_n^2 + \frac{1}{\sqrt{nb_n}})c_n^{-1/4}\right)\log n$. For a proof of this estimate we consider the events

$$A_n = \left\{ \sup_{t\in[0,1]} |\hat{\mu}_{b_n}(t) - \mu(t)| \leq C\frac{\log n}{\sqrt{nb_n}b_n^{1/4}} + Cb_n^2 \log n \right\},$$

$$B_n = \left\{ \sup_{t\in[0,1]} |\hat{\sigma}^2(t) - \sigma^2(t)| \leq C\left(c_n^2 + (\frac{1}{\sqrt{nc_n}} + b_n^2 + \frac{1}{\sqrt{nb_n}})c_n^{-1/4}\right)\log n. \right\},$$

then by (8.5) and Corollary 8.2 in Section 8.1 it follows that $\lim_{n\to\infty} \mathbb{P}(A_n \cap B_n) = 1$. Using similar arguments as in the proof of the estimate (7.12) it can be shown that

$$\|(\Phi_{n,m}^o - \Phi_{n,m})I(A_n \cap B_n)\|^2 \leq C\left(\frac{m}{n^2}\right),$$

$$\|(\Phi_{n,m}^o - \hat{\tilde{\Phi}}_{n,m})I(A_n \cap B_n)\|^2 \leq Cm\delta_n^2.$$

The assertion of Theorem 3.4 now follows by a similar argument as given in the proof of Theorem 3.2. $\qquad \square$

7.2 Proof of Lemma 4.1 - 4.3

In order to simplify the notation define $G_n(m) = S_m - \frac{m}{n}S_n$, $\hat{G}_n(m) = \hat{S}_m - \frac{m}{n}\hat{S}_n$, where $S_m = \sum_{i=1}^{m} e_i^2$, $\hat{S}_m = \sum_{i=1}^{m} \hat{e}_i^2$. Then it is easy to see that the estimator $\tilde{t}_n$ of the change point in the variance function defined in (4.1) can be represented as $\tilde{t}_n = \frac{1}{n}\text{argmax}_{1\leq m\leq n}(\hat{G}_n(m))^2$.

Similarly, we introduce the notation

$$\check{G}_n(m) = \sum_{j=1}^{m} \frac{\hat{e}_j \hat{e}_{j+k}}{\hat{\sigma}^{2*}(t_j)} - \frac{m}{n} \sum_{j=1}^{n} \frac{\hat{e}_j \hat{e}_{j+k}}{\hat{\sigma}^{2*}(t_j)}$$

and obtain the representation $\hat{t}_n = \frac{1}{n} \text{argmax}_{1 \leq m \leq n} (\check{G}_n(m))^2$ for the estimator of the change point in the correlation function defined in (4.12).

7.2.1   Proof of Lemma 4.1

Recall that Section 4.1 considers the problem of testing for a non relevant change in the variance and that under null hypothesis, we have $\sigma^2(s) = \sigma^2$ for $s \leq \tilde{t}_v$ and $\sigma^2(s) = \sigma^2 + \gamma$ for $s > \tilde{t}_v$, where γ is an unknown (without loss of generality) positive constant. A simple calculation shows that

$$f_n(m) := \mathbb{E} G_n(m) = n(m(n)t(n) - m(n) \wedge t(n))\gamma, \tag{7.32}$$

where we used the notation $m(n) = m/n$ and $t(n) = \lfloor n\tilde{t}_v \rfloor / n$. By Proposition 5 of Zhou (2013), on a possibly richer probability space, there exist $i.i.d$ standard normal variables, say $\{V_i\}_{i \in \mathbb{Z}}$, such that

$$\max_{1 \leq i \leq n} \left| S_i - \mathbb{E}(S_i) - \sum_{j=1}^{i} \kappa_1(t_j) V_j \right| = o_p(n^{1/4} \log n), \tag{7.33}$$

where κ_1 is defined in assumption (A5). By the arguments given in Section 7.1.1 we have

$$\max_{1 \leq m \leq n} |G_n(m) - \hat{G}_n(m)| = O_p(\varrho_n), \tag{7.34}$$

where $\varrho_n = b_n^{-1} + nb_n^3 + \sqrt{nb_n}$. Now a similar reasoning as given in the proof of Lemma 5 of Zhou and Wu (2010), Assumption (A3) (A4) and (A5) yield that there exists a constant C such that $\kappa_1^2(s) \leq C$ for all $s \in [0,1]$. Then it is easy to see that $\|\Xi_n\|_2^2 = O(n)$. By Doob's inequality, we have that

$$\max_{1 \leq j \leq n} |\Xi_j| = O_p(\sqrt{n}), \tag{7.35}$$

49

and observing (7.33) we obtain

$$\max_{1 \le m \le n} \left| G_n^2(m) - \hat{G}_n^2(m) \right| = \max_{1 \le m \le n} |G_n(m) + \hat{G}_n(m)||G_n(m) - \hat{G}_n(m)| = O_p(n \varrho_n).$$

Define $\hat{V}_n(m) = \hat{G}_n^2(m) - \hat{G}_n^2(\lfloor nt \rfloor)$, note that $\hat{V}_n(\lfloor nt \rfloor) = 0$ and consider a constant $\alpha \in (\frac{1}{2}, \frac{2}{3})$, such that $n^{1-\alpha}/\varrho_n \to \infty$. Observing the definition (7.32) and the estimate (7.33), it follows that

$$\max_{1 \le m \le n} \left| G_n^2(m) - \left(f_n(m) + \Xi_m - \frac{m}{n} \Xi_n \right)^2 \right| = O_p(n^{5/4} \log n). \tag{7.36}$$

By (7.35), we have $\max_{1 \le m \le n}(\Xi_m - \frac{m}{n}\Xi_n)^2 = O_p(n)$, and together with (7.34) and (7.36) this yields

$$\max_{m \in \mathcal{M}_n} \hat{V}_n(m) = \max_{m \in \mathcal{M}_n} \left[G_n^2(m) - G_n^2(\lfloor nt \rfloor) \right] + O_p(n\rho_n) = \max_{m \in \mathcal{M}_n} \left\{ f_n^2(m) - f_n^2(\lfloor nt \rfloor) \right.$$
$$+ 2(f_n(m) - f_n(\lfloor nt \rfloor))\Xi_m + 2f_n(\lfloor nt \rfloor)(\Xi_m - \Xi_{\lfloor nt \rfloor})$$
$$\left. - 2\frac{m}{n} f_n(m)\Xi_n + 2\frac{\lfloor nt \rfloor}{n} f_n(\lfloor nt \rfloor)\Xi_n \right\} + O_p(n\varrho_n + n^{5/4}\log n), \tag{7.37}$$

where the maxima are taken over the set $\mathcal{M}_n = \{m \mid \lfloor n\tilde{t}_v \rfloor - \lfloor n^{1-\alpha/2} \rfloor \le m \le \lfloor n\tilde{t}_v \rfloor - \lfloor n^{1-\alpha} \rfloor \rfloor\}$. Observing the definition of $f_n(m)$ in (7.32) we have for some positive constant C,

$$\max_{m \in \mathcal{M}_n} [f_n^2(m) - f_n^2(\lfloor nt \rfloor)] \le -Cn^{2-\alpha}, \tag{7.38}$$

and (7.35) implies

$$\max_{m \in \mathcal{M}_n} [f_n(m) - f_n(\lfloor nt \rfloor)]\Xi_m = O_p(n^{3/2-\alpha/2} \log n), \tag{7.39}$$

$$\max_{m \in \mathcal{M}_n} \left[\frac{m}{n} f_n(m) - \frac{\lfloor nt \rfloor}{n} f_n(\lfloor nt \rfloor) \right]\Xi_n = O_p(n^{3/2-\alpha/2} \log n). \tag{7.40}$$

Using the representation $\Xi_m - \Xi_{\lfloor nt \rfloor} = \sum_{i=m+1}^{\lfloor nt \rfloor} \sigma(t_i)V_i$ and similar arguments as in the derivation of (7.35) yields

$$\max_{m \in \mathcal{M}_n} [\Xi_m - \Xi_{\lfloor nt \rfloor}] = O_p(n^{1/2(1-\alpha/2)} \log n).$$

Consequently,

$$\max_{m \in \mathcal{M}_n} f_n(\lfloor nt \rfloor)[\Xi_m - \Xi_{\lfloor nt \rfloor}] = O_p(n^{3/2-\alpha/4} \log n). \tag{7.41}$$

50

By our choice of α, (7.37) - (7.41) it now follows that

$$\mathbb{P}\left(\limsup_{n\to\infty} \max_{\lfloor n\tilde{t}_v\rfloor - \lfloor n^{1-\alpha/2}\rfloor \le m \le \lfloor n\tilde{t}_v\rfloor - \lfloor n^{1-\alpha}\rfloor} \hat{V}_n(m) = -\infty\right) = 1. \qquad (7.42)$$

On the other hand, similar arguments give the estimates

$$\max_{1\le m\le \lfloor n\tilde{t}_v\rfloor - \lfloor n^{1-\alpha/2}\rfloor} [f_n^2(m) - f_n^2(\lfloor nt\rfloor)] \le -Cn^{2-\alpha/2},$$

$$\max_{1\le m\le \lfloor n\tilde{t}_v\rfloor - \lfloor n^{1-\alpha/2}\rfloor} f_n(\lfloor nt\rfloor)[\Xi_m - \Xi_{\lfloor nt\rfloor}] = O_p(n^{3/2}\log n),$$

$$\max_{1\le m\le \lfloor n\tilde{t}_v\rfloor - \lfloor n^{1-\alpha/2}\rfloor} [\frac{m}{n}f_n(m) - \frac{\lfloor nt\rfloor}{n}f_n(\lfloor nt\rfloor)]\Xi_n = O_p(n^{3/2}\log n),$$

$$\max_{1\le m\le \lfloor n\tilde{t}_v\rfloor - \lfloor n^{1-\alpha/2}\rfloor} [f_n(m) - f_n(\lfloor nt\rfloor)]\Xi_m = O_p(n^{3/2}\log n),$$

and by our choice of α we obtain $\mathbb{P}(\limsup_{n\to\infty} \max_{1\le m\le \lfloor n\tilde{t}_v\rfloor - \lfloor n^{1-\alpha/2}\rfloor} \hat{V}_n(m) = -\infty) = 1$. Combined with (7.42) this gives $\mathbb{P}(\limsup_{n\to\infty} \max_{1\le m\le \lfloor n\tilde{t}_v\rfloor - \lfloor n^{1-\alpha}\rfloor} \hat{V}_n(m) = -\infty) = 1$, and it can be shown by similar arguments that $\mathbb{P}(\limsup_{n\to\infty} \max_{\lfloor n\tilde{t}_v\rfloor + \lfloor n^{1-\alpha}\rfloor \le m \le n} \hat{V}_n(m) = -\infty) = 1$. Consequently, it follows that

$$\lim_{n\to\infty} \mathbb{P}(|n\tilde{t}_n - \lfloor n\tilde{t}_v\rfloor| \le n^{1-\alpha}) = 1,$$

which proves part (ii) of Lemma 4.1. For the case that the variance has no jump at time t, the desired result follows from the fact that $\hat{G}_n(m)/\sqrt{n}$ converges weakly to some Gaussian process $\{U_1(s) - sU_1(1)\}_{s\in[0,1]}$, which implies $\tilde{t}_n \xrightarrow{\mathcal{D}} \tilde{T} = \operatorname{argmax}_{s\in(0,1)} |U_1(s) - sU_1(1)|$, where the Gaussian process $\{U_1(s)\}_{s\in[0,1]}$ is defined in Theorem 3.1.

7.2.2 Proof of Lemma 4.2

By a careful examination of the proof of Theorem 4.3 (see the next section) it follows that

$$\max_{1\le m\le n} |\check{G}_n(m) - \bar{G}_n(m)| = O_p(nc_n^2 + nb_n^3 c_n^{-1/4} + n^{1-\alpha}\log n + b_n^{-1}c_n^{-1}),$$

where

$$\bar{G}_n(m) = \sum_{j=1}^m \frac{e_j e_{j+k}}{\sigma^2(t_j)} - \frac{m}{n}\sum_{j=1}^n \frac{e_j e_{j+k}}{\sigma^2(t_j)}.$$

51

Write $\varrho'_n = nc_n^2 + nb_n^3 c_n^{-1/4} + b_n^{-1} c_n^{-1} + n^{1-\alpha} \log n,$ where α is defined in the proof of Lemma 4.1. Let $1/2 < \alpha' < 2/3$, such that $n^{1-\alpha'}/\varrho'_n \to \infty$. Then using similar arguments as given in the proof of Lemma 4.1 we can show that

$$\hat{t}_n - t_n = O_p(n^{-\alpha'}),$$

if there is a change in the correlation at time t. On the other hand, if there is no change in correlation, we have that $\hat{t}_n \xrightarrow{\mathcal{D}} T_{\max}$, where $T_{\max} = \operatorname{argmax}_{t \in (0,1)} |U_2(t) - t U_2(1)|$, and the stochastic process $\{U_2(s)\}_{s \in [0,1]}$ is defined in Theorem 3.3.

7.2.3   Proof of Lemma 4.3

Recall the definition of (4.17) and define $\Delta_{n,1} = \frac{1}{\lfloor nt \rfloor} \sum_{j=1}^{\lfloor nt \rfloor} W_j$, $\Delta_{n,2} = \frac{1}{n - \lfloor nt \rfloor} \sum_{j=\lfloor nt \rfloor+1}^{n} W_j$ (the superscript k is again omitted). From the proof of Theorem 3.3 we have that

$$\Delta_{n,1} - \mathbb{E}[\Delta_1] = O_p\left(\frac{1}{\sqrt{n}}\right), \quad \Delta_{n,2} - \mathbb{E}[\Delta_2] = O_p\left(\frac{1}{\sqrt{n}}\right).$$

Since $\Delta = \mathbb{E}[\Delta_2] - \mathbb{E}[\Delta_1]$ we have $\Delta_n := \Delta_{n,2} - \Delta_{n,1} = \Delta + O_p(1/\sqrt{n})$. In order to prove this estimate we introduce the notation $\mathcal{A}_n = \{|\hat{t}_n - t| \le \frac{C \log n}{\sqrt{n}}\}$. Then by Lemma 4.2, we have that $\lim_{n \to \infty} \mathbb{P}(\mathcal{A}_n) = 1$. This yields

$$(\Delta_{n,1} - \hat{\Delta}_{n,1}) I(\mathcal{A}_n) = I(\mathcal{A}_n)(A_n + B_n + C_n),$$

where

$$A_n = \sum_{j=1}^{\lfloor nt \rfloor} \frac{W_j}{\lfloor nt \rfloor} - \sum_{j=1}^{\lfloor n\hat{t}_n \rfloor} \frac{W_j}{\lfloor nt \rfloor}, \quad B_n = \sum_{j=1}^{\lfloor n\hat{t}_n \rfloor} \left(\frac{W_j}{\lfloor nt \rfloor} - \frac{\hat{W}_j}{\lfloor nt \rfloor}\right), \quad C_n = \sum_{j=1}^{\lfloor n\hat{t}_n \rfloor} \left(\frac{\hat{W}_j}{\lfloor nt \rfloor} - \frac{\hat{W}_j}{\lfloor n\hat{t}_n \rfloor}\right). \quad (7.43)$$

It is easy to see that $I(\mathcal{A}_n)A_n = O_p(\frac{\log n}{\sqrt{n}})$. Using the same arguments as in the proof of Theorem 4.3, we obtain $I(\mathcal{A}_n)B_n = o(\sqrt{n}/n) = o(1/\sqrt{n})$ and

$$
\begin{aligned}
I(\mathcal{A}_n) \cdot C_n &= I(\mathcal{A}_n) \sum_{j=1}^{\lfloor n\hat{t} \rfloor} \hat{W}_j \frac{\lfloor n\hat{t} \rfloor - \lfloor nt \rfloor}{\lfloor nt \rfloor \lfloor n\hat{t} \rfloor} \leq C \cdot I(\mathcal{A}_n) \sum_{j=1}^{\lfloor n\hat{t} \rfloor} \hat{W}_j \frac{\log n}{n\sqrt{n}} \\
&= C \cdot I(\mathcal{A}_n) \Big(\sum_{j=1}^{\lfloor n\hat{t} \rfloor} W_j + o_p(\sqrt{n}) \Big) \frac{\log n}{n\sqrt{n}} \\
&= I(\mathcal{A}_n) \cdot \Big(\sum_{j=1}^{\lfloor nt \rfloor} W_j + \sum_{j=\lfloor nt \rfloor}^{\lfloor n\hat{t} \rfloor} W_j I(t \leq \hat{t}) - \sum_{j=\lfloor n\hat{t} \rfloor}^{\lfloor nt \rfloor} W_j I(t > \hat{t}) + o_p(\sqrt{n}) \Big) \frac{\log n}{n\sqrt{n}} \\
&= O_p\Big(\frac{\log n}{\sqrt{n}} \Big). \tag{7.44}
\end{aligned}
$$

Combining (7.43) - (7.44) and using Proposition 8.2 in Section 8.2 shows $\Delta_{n,1} - \hat{\Delta}_{n,1} = O_p\big(\frac{\log n}{\sqrt{n}}\big)$. Similarly, we have $\Delta_{n,2} - \hat{\Delta}_{n,2} = O_p\big(\frac{\log n}{\sqrt{n}}\big)$, and the assertion of the lemma follows.

7.3   Proof of Theorem 4.3 and 4.4

7.3.1   Proof of Theorem 4.3

We consider the non-observable analogue

$$
T_n^r = \frac{3}{t^2(1-t)^2} \int_0^1 U_n^2(s)\,ds.
$$

of the statistic $\hat{T}_n^r$ defined in (4.22), where the process U_n is given by

$$
U_n(s) = \frac{1}{n} \sum_{j=1}^{\lfloor ns \rfloor} \frac{e_j e_{j+k}}{\sigma(t_j)\sigma(t_{j+k})} - \frac{s}{n} \sum_{j=1}^{n} \frac{e_j e_{j+k}}{\sigma(t_j)\sigma(t_{j+k})}.
$$

It follows from the proof of Theorem 3.3 that

$$
\{\sqrt{n}(U_n(s) + (s \wedge t - st)\Delta)\}_{s\in[0,1]} \Rightarrow \{U_2(s) - sU_2(1)\}_{s\in[0,1]}, \tag{7.45}
$$

whenever $\Delta \neq 0$. The continuous mapping theorem, elementary calculations, and the identity $3\int_0^1 [st - s \wedge t]^2 ds = t^2(1-t)^2$ imply $\sqrt{n}(T_n^r - \Delta^2) \xrightarrow{\mathcal{D}} \mathcal{Z}_2(\Delta)$, where the random variable $\mathcal{Z}_2(\Delta)$

is defined in Theorem 4.3. Finally, we show the estimate

$$\sqrt{n}(T_n^r - \hat{T}_n^r) = O_p(n^{1/2}c_n^2 + n^{1/2}b_n^3 c_n^{-1/4} + n^{-1/2}\zeta_n + n^{-1/2}b_n^{-1}c_n^{-1}), \tag{7.46}$$

where $\zeta_n = n^{1-\alpha} + nc_n\left(n^{-\alpha}c_n^{-5/4} + c_n^2 + (\frac{1}{\sqrt{nc_n}} + b_n^2 + \frac{1}{\sqrt{nb_n}})c_n^{-1/4}\right)$, $\alpha = 1 - 4/\iota - \nu$ for some $\nu > 0$, which completes the proof.

For a proof of this remaining estimate we consider the case $\Delta > 0$, where it follows from Lemma 3.1 that $|t_n^* - \tilde{t}_v| = O_p(n^{-\alpha})$. The proof of the statement in the case $\Delta = 0$ (which corresponds to $\hat{t}_n \xrightarrow{\mathcal{D}} T_{\max}$) is easier and omitted for the sake of brevity. For the event $A_n = \{|t_n^* - \tilde{t}_v| \leq \frac{1}{n^\alpha}\}$ it follows from Lemma 3.1 that $\lim_{n\to} \mathbb{P}(A_n) = 1$. Observe that we have from Lemma 8.4

$$\sup_{t \in T_n} \left| \hat{\sigma}_n^{2*}(t) - \sigma^2(t) - \frac{\mu_2 \ddot{\sigma}^2(t)c_n^2}{2} - \frac{1}{nc_n}\sum_{i=1}^n (\hat{e}_i^2 - \mathbb{E}e_i^2)K_{b_n}(t_i - t) \right| = O\left(c_n^3 + \frac{c_n}{n}\right),$$

where $T_n = [c_n, \tilde{t}_v - n^{-\alpha} - c_n] \cup [\tilde{t}_v + n^{-\alpha} + c_n, 1 - c_n]$. Also (8.9) still holds. Now similar calculations as given in the proof of (7.14) yield

$$\max_{1 \leq l \leq \lfloor n(\tilde{t}_v - n^{-\alpha} - c_n)\rfloor} \left| \sum_{j=1}^l \left(\frac{e_j e_{j+k}}{\sigma(t_j)\sigma(t_{j+k})} - \frac{\hat{e}_j \hat{e}_{j+k}}{\hat{\sigma}_n^{2*}(t_j)} \right) \right| I(A_n) = O_p(nc_n^2 + nb_n^3 c_n^{-1/4} + b_n^{-1}c_n^{-1}),$$

$$\max_{\lfloor n(\tilde{t}_v - n^{-\alpha} - c_n)\rfloor \leq l \leq \lfloor n(\tilde{t}_v + n^{-\alpha} + c_n)\rfloor} \left| \sum_{j=\lfloor n(\tilde{t}_v - n^{-\alpha} - c_n)\rfloor}^l \left(\frac{e_j e_{j+k}}{\sigma(t_j)\sigma(t_{j+k})} - \frac{\hat{e}_j \hat{e}_{j+k}}{\hat{\sigma}_n^{2*}(t_j)} \right) \right| I(A_n) = O_p(\zeta_n),$$

$$\max_{\lfloor n(\tilde{t}_v + n^{-\alpha} + c_n)\rfloor \leq l \leq \lfloor ns\rfloor} \left| \sum_{j=\lfloor n(\tilde{t}_v + n^{-\alpha} + c_n)\rfloor}^l \left(\frac{e_j e_{j+k}}{\sigma(t_j)\sigma(t_{j+k})} - \frac{\hat{e}_j \hat{e}_{j+k}}{\hat{\sigma}_n^{2*}(t_j)} \right) \right| I(A_n) = O_p(nc_n^2 + nb_n^3 c_n^{-1/4} + b_n^{-1}c_n^{-1}),$$

where we used Corollary 8.3 for the second statement. So we have

$$\sup_{0 \leq s \leq 1} n|U_n(s) - \hat{U}_n(s)|I(A_n) = O_p(nc_n^2 + nb_n^3 c_n^{-1/4} + b_n^{-1}c_n^{-1} + \zeta_n).$$

Using the same arguments and Proposition 8.2 in Section 8.2 we obtain

$$\sup_{0 \leq s \leq 1} n|U_n(s) - \hat{U}_n(s)| = O_p(nc_n^2 + nb_n^3 c_n^{-1/4} + b_n^{-1}c_n^{-1} + \zeta_n).$$

54

From (7.45) it follows that $\int_0^1 |U_n(s)| ds = O_p(1)$. Consequently, we have

$$n^{1/2} \int_0^1 [U_n^2(s) - \hat{U}_n^2(s)] ds \leq \sup_{0 \leq s \leq 1} n^{1/2} |U_n(s) - \hat{U}_n(s)| \int_0^1 |U_n(s) + \hat{U}_n(s)| ds$$

$$\leq 2n^{1/2} \sup_{0 \leq s \leq 1} |U_n(s) - \hat{U}_n(s)| \int_0^1 |U_n(s)| ds + n^{1/2} \sup_{0 \leq s \leq 1} |U_n(s) - \hat{U}_n(s)|^2$$

$$= O_p(n^{1/2} c_n^2 + n^{1/2} b_n^3 c_n^{-1/4} + n^{-1/2} \zeta_n + n^{-1/2} b_n^{-1} c_n^{-1}),$$

and the remaining estimate (7.46) follows.

7.3.2 Proof of Theorem 4.4

Recall the definition of $\hat{\Delta}_n$, $\hat{A}_j$, $\hat{\Phi}_{i,m}^A$ in (4.17), (4.18), (4.19) and define

$$A_j = \frac{e_j e_{j+k}}{\sigma(t_j)\sigma(t_{j+k})} - \Delta I(j \geq \lfloor nt \rfloor),$$

$$\Phi_{i,m}^A = \frac{1}{\sqrt{m(n-m+1)}} \sum_{j=1}^{n-m+1} (S_{j,m}^A - \frac{m}{n} S_n^A) R_j,$$

where $S_{j,m}^A = \sum_{r=j}^{j+m-1} A_r$, $S_n^A = \sum_{r=1}^n A_r$. We introduce the processes

$$\Phi_{m,n}^A(s) = \Phi_{\lfloor ns \rfloor,m}^A + (ns - \lfloor ns \rfloor)(\Phi_{\lfloor ns \rfloor+1,m}^A - \Phi_{\lfloor ns \rfloor,m}^A),$$

$$\hat{\Phi}_{m,n}^A(s) = \hat{\Phi}_{\lfloor ns \rfloor,m}^A + (ns - \lfloor ns \rfloor)(\hat{\Phi}_{\lfloor ns \rfloor+1,m}^A - \hat{\Phi}_{\lfloor ns \rfloor,m}^A). \tag{7.47}$$

and note that it follows from Zhou (2013) that $\{\Phi_{m,n}^A(s)\}_{s \in [0,1]} \Rightarrow \{U_2(s)\}_{s \in [0,1]}$ conditional on $\mathcal{F}_n$. The assertion of Theorem 4.4 is therefore a consequence of the estimate

$$\sup_{s \in (0,1)} |\Phi_{m,n}^A(s) - \hat{\Phi}_{m,n}^A(s)| = O_p(\sqrt{m} \rho_n), \tag{7.48}$$

where $\rho_n = \left(n^{-\alpha} c_n^{-5/4} + c_n^2 + (\frac{1}{\sqrt{nc_n}} + b_n^2 + \frac{1}{\sqrt{nb_n}}) c_n^{-1/4} \right) \log n$. To show (7.48), define $\mathcal{A}_n = \{|\hat{t}_n - t| \leq \frac{C \log n}{\sqrt{n}}\}$, $\mathcal{B}_n = \{|\hat{\Delta}_n - \Delta| \leq \frac{C \log^2 n}{\sqrt{n}}\}$, $\mathcal{C}_n = \{\sup_{t \in (0,1)} |\hat{\mu}_{b_n}(t) - \mu(t)| \leq C \frac{\log n}{\sqrt{nb_n} b_n^{1/4}} + C b_n^2 \log n\}$,

$$\mathcal{D}_{1,n} = \left\{ \sup_{t \in [0, \tilde{t}_v - n^{-\alpha})} |\hat{\sigma^{2*}}(t) - \sigma^2(t)| \leq C \rho_n \right\},$$

$$\mathcal{D}_{2,n} = \left\{ \sup_{t \in (\tilde{t}_v + n^{-\alpha}, 1]} |\hat{\sigma^{2*}}(t) - \sigma^2(t)| \leq C \rho_n \right\},$$

where C is some sufficiently large constant, α is defined in Corollary 8.3. By construction it follows $\lim_{n\to\infty} \mathbb{P}(\mathcal{W}_n) = 1$, where $\mathcal{W}_n = \mathcal{A}_n \cap \mathcal{B}_n \cap \mathcal{C}_n \cap \mathcal{D}_{1,n} \cap \mathcal{D}_{2,n}$. Let $\tilde{A}_j = \frac{\hat{e}_j \hat{e}_{j+k}}{\hat{\sigma}^{2*}(t_j)} - \Delta I(j \geq \lfloor nt \rfloor)$. Similarly to the definition (4.19) and (7.47), we define $\tilde{S}^A_{i,m}$, $\tilde{\Phi}^A_{i,m}$, $\tilde{\Phi}^A_{m,n}(t)$, where the random variables $\hat{A}_i$ are replaced by $\tilde{A}_i$. From the proof of the estimate (7.31) it follows

$$\sup_{t\in(0,1)} |\tilde{\Phi}^A_{m,n}(t) - \Phi^A_{m,n}(t)| = O_p\Big(\sqrt{m}\rho_n + \sqrt{m}n^{-\alpha/2}\Big). \tag{7.49}$$

On the other hand,

$$\mathbb{E}[(\tilde{S}^A_{j,m} - \hat{S}^A_{j,m})^2 I(\mathcal{W}_n)]$$

$$=\mathbb{E}\Big[\sum_{r=j}^{j+m-1} I(\mathcal{W}_n)\big(\Delta\big(I(r \geq \lfloor nt \rfloor) - I(r \geq \lfloor n\hat{t}_n \rfloor)\big)\big) + \sum_{r=j}^{j+m-1} I(\mathcal{W}_n)(\Delta - \hat{\Delta})I(r \geq \lfloor n\hat{t}_n \rfloor)\Big]^2$$

$$\leq C\mathbb{E}\Big[\sum_{r=j}^{j+m-1} I(\mathcal{W}_n)\big(\Delta\big(I(r \geq \lfloor nt \rfloor) - I(r \geq \lfloor n\hat{t}_n \rfloor)\big)\big)\Big]^2 + \Big(\frac{m\log^2 n}{\sqrt{n}}\Big)^2.$$

Note that $\mathbb{E}[\sum_{r=j}^{j+m-1} I(\mathcal{W}_n)(\Delta(I(r \geq \lfloor nt \rfloor) - I(r \geq \lfloor n\hat{t}_n \rfloor)))]^2 = 0$ if $j < \lfloor nt - C\log n\sqrt{n} \rfloor - m + 1$ or $j \geq \lfloor nt + C\log n\sqrt{n} \rfloor$, and is bounded by m^2 if $j \in [\max(1, \lfloor nt - C\log n\sqrt{n} \rfloor + m - 1), \min(n, \lfloor nt + C\log n\sqrt{n} \rfloor)]$. Thus we have

$$\frac{1}{(n-m+1)m} \sum_{j=1}^{n-m+1} \mathbb{E}[(\tilde{S}^A_{j,m} - \hat{S}^A_{j,m})^2 I(\mathcal{W}_n)] \leq \frac{Cm\log n}{\sqrt{n}} + \frac{Cm\log^4 n}{n}, \tag{7.50}$$

and similar arguments lead to

$$\frac{1}{(n-m+1)m} \sum_{j=1}^{n-m+1} \mathbb{E}\Big[(\tilde{S}^A_n - \hat{S}^A_n)^2 I(\mathcal{W}_n)\frac{m^2}{n^2}\Big] \leq \frac{Cm\log^4 n}{n}. \tag{7.51}$$

Now similar arguments as given in the proof of estimate (7.31) together with (7.50) and (7.51), yield

$$\sup_{t\in(0,1)} |\tilde{\Phi}^A_{m,n}(t) - \hat{\Phi}^A_{m,n}(t)| = O_p\Big(\Big(\frac{m\log n}{\sqrt{n}}\Big)^{1/2}\Big),$$

and the assertion (7.48) follows from (7.49).

8 More technical details

8.1 Uniform bounds for nonparametric estimates

The following two lemmas provide uniform bounds for the estimate $\hat{\mu}_{b_n}$ in the interior $\mathfrak{T}_n = [b_n, 1 - b_n]$ and at the boundary $\mathfrak{T}'_n = (0, b_n] \cup [1 - b_n, 1)$ of the interval $[0, 1]$.

Lemma 8.1. *If assumptions (A1)-(A3) are satisfied and $b_n \to 0$, $nb_n \to \infty$, we have*

$$\sup_{t \in \mathfrak{T}_n} \left| \hat{\mu}_{b_n}(t) - \mu(t) - \frac{\mu_2 \ddot{\mu}(t)}{2} b_n^2 - \frac{1}{nb_n} \sum_{i=1}^{n} e_i K_{b_n}(t_1 - t) \right| = O(b_n^3 + \frac{b_n}{n}),$$

where $\mathfrak{T}_n = [b_n, 1 - b_n]$.

Proof. With the notations

$$S_{n,l}(t) = \frac{1}{nb_n} \sum_{i=1}^{n} \left(\frac{t_i - t}{b_n} \right)^l K_{b_n}(t_i - t), \qquad R_{n,l}(t) = \frac{1}{nb_n} \sum_{i=1}^{n} Y_i \left(\frac{t_i - t}{b_n} \right)^l K_{b_n}(t_i - t),$$

$(l = 0, 1, ...)$ we obtain the representation

$$\begin{bmatrix} \hat{\mu}_{b_n}(t) \\ b_n \hat{\dot{\mu}}_{b_n}(t) \end{bmatrix} = \begin{bmatrix} S_{n,0}(t) & S_{n,1}(t) \\ S_{n,1}(t) & S_{n,2}(t) \end{bmatrix}^{-1} \begin{bmatrix} R_{n,0}(t) \\ R_{n,1}(t) \end{bmatrix} =: S_n^{-1}(t) R_n(t), \tag{8.1}$$

for the local linear estimate $\tilde{\mu}_{b_n}$, where the last identity defines the 2×2 matrix $S_n(t)$ and the vector $R_n(t)$ in an obvious manner. By elementary calculation and a Taylor expansion we have

$$S_n(t) \begin{bmatrix} \hat{\mu}_{b_n}(t) - \mu(t) \\ b_n(\hat{\dot{\mu}}_{b_n}(t) - \dot{\mu}(t)) \end{bmatrix} = \begin{bmatrix} \frac{1}{nb_n} \sum_{i=1}^{n} e_i K_{b_n}(t_i - t) + \frac{1}{2} \ddot{\mu}(t) \mu_2 b_n^2 \\ \frac{1}{nb_n} \sum_{i=1}^{n} e_i K_{b_n}(t_i - t)(\frac{t_i - t}{b_n}) \end{bmatrix} + O(b_n^3 + b_n/n)$$

uniformly with respect to $t \in \mathfrak{T}_n$. Note that $S_{n,0}(t) = 1 + O(\frac{1}{nb_n})$ and $S_{n,1}(t) = O(\frac{1}{nb_n})$, uniformly with respect to $t \in \mathfrak{T}_n$, which yields

$$\sup_{t \in \mathfrak{T}_n} \left| \hat{\mu}_{b_n}(t) - \mu(t) - \frac{\mu_2 \ddot{\mu}(t)}{2} b_n^2 - \frac{1}{nb_n} \sum_{i=1}^{n} e_i K_{b_n}(t_i - t) \right| = O(b_n^3 + \frac{b_n}{n}).$$

Therefore the lemma follows from the definition of the estimate $\hat{\mu}_{b_n}$ in (3.2). $\qquad \square$

Lemma 8.2. *Assume that the conditions of Lemma 8.1 hold, then*

$$\sup_{t\in\mathfrak{T}'_n}\left|c(t)(\hat{\mu}_{b_n}(t)-\mu(t))-\frac{1}{nb_n}\sum_{i=1}^{n}\left[\nu_{2,b_n}(t)-\nu_{1,b_n}(t)\left(\frac{t_i-t}{b_n}\right)\right]e_iK_{b_n}(t_i-t)+\right.$$

$$\left.\frac{b_n^2}{2}\ddot{\mu}(t)(\nu_{2,b_n}^2(t)-\nu_{1,b_n}(t)\nu_{3,b_n}(t))\right|=O(b_n^3+\frac{b_n}{n}),$$

where $\mathfrak{T}'_n=[0,b_n]\cup[1-b_n,1]$, $\nu_{j,b_n}(t)=\int_{-t/b_n}^{(1-t)/b_n}x^jK(x)dx$ *and* $c(t)=\nu_{0,b_n}(t)\nu_{2,b_n}(t)-\nu_{1,b_n}^2(t)$.

Proof. For any $t\in[0,b_n]\cup[1-b_n,1]$, using (8.1), we obtain

$$S_n(t)\begin{bmatrix}\hat{\mu}_{b_n}(t)-\mu(t)\\b_n(\hat{\dot{\mu}}_{b_n}(t)-\dot{\mu}(t))\end{bmatrix}=\begin{bmatrix}\frac{1}{nb_n}\sum_{i=1}^{n}[Y_i-\mu(t)-\dot{\mu}(t)(t_i-t)]K_{b_n}(t_i-t)\\\frac{1}{nb_n}\sum_{i=1}^{n}[Y_i-\mu(t)-\dot{\mu}(t)(t_i-t)]K_{b_n}(t_i-t)(\frac{t_i-t}{b_n})\end{bmatrix},$$

and a Taylor expansion yields

$$S_n(t)\begin{bmatrix}\hat{\mu}_{b_n}(t)-\mu(t)\\b_n(\hat{\dot{\mu}}_{b_n}(t)-\dot{\mu}(t))\end{bmatrix}=\begin{bmatrix}\frac{1}{nb_n}\sum_{i=1}^{n}e_iK_{b_n}(t_i-t)+\frac{b_n^2}{2}\nu_{2,b_n}(t)\ddot{\mu}(t)\\\frac{1}{nb_n}\sum_{i=1}^{n}e_iK_{b_n}(t_i-t)(\frac{t_i-t}{b_n})+\frac{b_n^2}{2}\nu_{3,b_n}(t)\ddot{\mu}(t)\end{bmatrix}+O(b_n^3+b_n/n)$$

(8.2)

uniformly with respect to $t\in[0,b_n]\cup[1-b_n,1]$. On the other hand, uniformly with respect to $t\in[0,b_n]\cup[1-b_n,1]$, we have that

$$S_n(t)=\begin{bmatrix}\nu_{0,b_n}(t)&\nu_{1,b_n}(t)\\\nu_{1,b_n}(t)&\nu_{2,b_n}(t)\end{bmatrix}+O(\frac{1}{nb_n}).$$

(8.3)

Therefore, combining (8.2) and (8.3), it follows that

$$c(t)(\hat{\mu}_{b_n}(t)-\mu(t))=\frac{1}{nb_n}\sum_{i=1}^{n}\left[\nu_{2,b_n}(t)-\nu_{1,b_n}(t)\left(\frac{t_i-t}{b_n}\right)\right]e_iK_{b_n}(t_i-t)+$$

$$\frac{b_n^2}{2}\ddot{\mu}(t)\left(\nu_{2,b_n}^2(t)-\nu_{1,b_n}(t)\nu_{3,b_n}(t)\right)+O\left(b_n^3+\frac{b_n}{n}\right)$$

uniformly with respect to $t\in[0,b_n]\cup[1-b_n,1]$. □

The next lemma concerns the order of deviations of $\hat{\mu}_{b_n}$ from μ in the $\|\cdot\|_4$-norm.

Lemma 8.3. *Assume that assumptions (A1)-(A4) are satisfied and that $nb_n^3\to\infty$, $nb_n^6\to0$,*

then

$$\sup_{t\in[0,1]} \|\hat{\mu}_{b_n}(t) - \mu(t)\|_4 = O(b_n^2 + (nb_n)^{-1/2}), \tag{8.4}$$

$$\left\|\sup_{t\in[0,1]} |\hat{\mu}_{b_n}(t) - \mu(t)|\right\|_4 = O(b_n^2 + (nb_n)^{-1/2}b_n^{-1/4}). \tag{8.5}$$

Proof. Observing the stochastic expansion in Lemma 8.1 we first evaluate $\|\sum_{i=1}^n e_i K_{b_n}(t_i - t)\|_4$ and $\|\frac{\partial}{\partial t}\sum_{i=1}^n e_i K_{b_n}(t_i - t)\|_4$. Recalling the definition of $\mathcal{P}_i$ we note that

$$\left\|\sum_{i=1}^n e_i K_{b_n}(t_i - t)\right\|_4 = \left\|\sum_{i=1}^n \sum_{k=0}^\infty \mathcal{P}_{i-k} e_i K_{b_n}(t_i - t)\right\|_4 \le \sum_{k=0}^\infty \left\|\sum_{i=1}^n \mathcal{P}_{i-k} e_i K_{b_n}(t_i - t)\right\|_4.$$

Since for each k, $\mathcal{P}_{i-k} e_i K_{b_n}(t_i - t)$, $1 \le i \le n$ is a martingale difference sequence, it follows from Burkholder's inequality

$$\left\|\sum_{i=1}^n \mathcal{P}_{i-k} e_i K_{b_n}(t_i - t)\right\|_4^2 \le C\left\|\left(\sum_{i=1}^n (\mathcal{P}_{i-k} e_i K_{b_n}(t_i - t))^2\right)^{1/2}\right\|_4^2$$

$$\le C\sum_{i=1}^n \|(\mathcal{P}_{i-k} e_i K_{b_n}(t_i - t))^2\|_2 = C\sum_{i=1}^n \|(\mathcal{P}_{i-k} e_i K_{b_n}(t_i - t))\|_4^2,$$

and condition (A4) implies $\|\sum_{i=1}^n \mathcal{P}_{i-k} e_i K_{b_n}(t_i - t)\|_4 = O(\sqrt{nb_n}\chi^k)$, uniformly with respect to $t \in [0,1]$. This yields

$$\sup_{t\in[0,1]} \left\|\sum_{i=1}^n e_i K_{b_n}(t_i - t)\right\|_4 = O(\sqrt{nb_n}). \tag{8.6}$$

Similar arguments show $\sup_{t\in[0,1]} \|\frac{\partial}{\partial t}\sum_{i=1}^n e_i K_{b_n}(t_i - t)\|_4 = O(\sqrt{nb_n}b_n^{-1})$, and by Proposition 8.1 in Section 8.2 it follows that

$$\|\sup_{t\in[0,1]} |\sum_{i=1}^n e_i K_{b_n}(t_i - t)/(nb_n)|\|_4 = O((nb_n)^{-1/2}b_n^{-1/4}), \tag{8.7}$$

and by Lemma 8.1, we obtain

$$\left\|\sup_{t\in\mathfrak{T}_n} \left|(\hat{\mu}_{b_n}(t) - \mu(t))^2 - \left(\frac{1}{nb_n}\sum_{i=1}^n e_i K_{b_n}(t_i - t) + \frac{\mu_2\ddot{\mu}(t)}{2}b_n^2\right)^2\right|\right\|_2 = O\left(\frac{\chi_n}{\sqrt{nb_n}b_n^{1/4}} + \chi_n^2\right),$$

where $\chi_n = b_n^3 + b_n/n$. Hence $\|\sup_{t\in\mathfrak{T}_n}(\hat{\mu}_{b_n}(t) - \mu(t))^2\|_2 = O(\frac{1}{nb_n^{3/2}} + b_n^4)$. By similar argument

59

and Lemma 8.2 it follows that $\|\sup_{t\in\mathfrak{T}'_n}(\hat{\mu}_{b_n}(t)-\mu(t))^2\|_2 = O(\frac{1}{nb_n^{3/2}}+b_n^4)$, and a combination of the last two estimates gives (8.5). On the other hand, Lemma 8.1 and (8.6) also imply

$$\sup_{t\in\mathfrak{T}_n}\left\|(\hat{\mu}_{b_n}(t)-\mu(t))^2 - \left(\frac{1}{nb_n}\sum_{i=1}^n e_i K_{b_n}(t_i-t) + \frac{\mu_2\ddot{\mu}(t)}{2}b_n^2\right)^2\right\|_2 = O\left(\frac{\chi_n}{\sqrt{nb_n}}+\chi_n^2\right),$$

which further yields $\sup_{t\in\mathfrak{T}_n}\|(\hat{\mu}_{b_n}(t)-\mu(t))^2\|_2 = O((\frac{1}{\sqrt{nb_n}}+b_n^2)^2)$. Similar arguments show the estimate $\sup_{t\in\mathfrak{T}'_n}\|(\hat{\mu}_{b_n}(t)-\mu(t))^2\|_2 = O((\frac{1}{\sqrt{nb_n}}+b_n^2)^2)$, which proves the remaining estimate (8.4). $\qquad\square$

The following results give a uniform bound for the p-mean of $\hat{\sigma}^2(t)-\sigma^2(t)$, where $\hat{\sigma}^2 = \hat{\sigma}^2_{c_n,b_n}$ is the variance estimator defined in (3.9). We begin with a uniform asymptotic stochastic expansion for the difference $\hat{\sigma}^2(t)-\sigma^2(t)$.

Lemma 8.4. *Suppose that Assumptions (A1)-(A4) are satisfied, $c_n \to 0$, $nc_n \to \infty$, and the variance function σ^2 is strictly positive, twice differentiable with a Lipschitz continuous second derivative $\ddot{\sigma}^2$. Then the estimate $\hat{\sigma}^2 = \hat{\sigma}^2_{c_n,b_n}$ defined in (3.9) satisfies*

$$\sup_{t\in\mathfrak{T}_n}\left|\hat{\sigma}^2(t)-\sigma^2(t)-\frac{\mu_2\ddot{\sigma}^2(t)c_n^2}{2}-\frac{1}{nc_n}\sum_{i=1}^n(\hat{e}_i^2-\mathbb{E}e_i^2)K_{c_n}(t_i-t)\right| = O\left(c_n^3+\frac{c_n}{n}\right), \qquad (8.8)$$

$$\sup_{t\in\mathfrak{T}'_n}\left|c(t)(\hat{\sigma}^2(t)-\sigma^2(t))-\frac{1}{nc_n}\sum_{i=1}^n\left[\nu_{2,c_n}(t)-\nu_{1,c_n}(t)\left(\frac{t_i-t}{c_n}\right)\right][\hat{e}_i^2-\mathbb{E}(e_i^2)]K_{c_n}(t_i-t)+\right.$$
$$\left.\frac{c_n^2}{2}\ddot{\sigma}^2(t)(\nu^2_{2,c_n}(t)-\nu_{1,c_n}(t)\nu_{3,c_n}(t))\right| = O\left(c_n^3+\frac{c_n}{n}\right), \qquad (8.9)$$

where $c(t)$, $\nu_{j,c_n}(t)$ are defined in Lemma 8.2, $\mathfrak{T}_n = [c_n, 1-c_n]$ and $\mathfrak{T}'_n = [0,c_n]\cup[1-c_n,1]$.

Proof. Following the argument given in the proof of Lemma 8.1, we have that

$$S_n(t)\begin{bmatrix}(\hat{\sigma}^2(t)-\sigma^2(t))\\ c_n(\hat{\sigma}^2(t)-\dot{\sigma}^2(t))\end{bmatrix} = \begin{bmatrix}\frac{1}{nc_n}\sum_{i=1}^n(\hat{e}_i^2-\sigma^2(t)-\dot{\sigma}^2(t)(t_i-t))K_{c_n}(t_i-t)\\ \frac{1}{nc_n}\sum_{i=1}^n(\hat{e}_i^2-\sigma^2(t)-\dot{\sigma}^2(t)(t_i-t))(\frac{t_i-t}{c_n})K_{c_n}(t_i-t)\end{bmatrix},$$

where $S_n(t)$ is defined in the proof of Lemma 8.1. The lemma now follows by the same arguments as given in the proof of Lemma 8.1 and Lemma 8.2, which are omitted for the sake of brevity.$\square$

Corollary 8.1. *Suppose that the conditions of Lemma 8.4 hold with $\iota \geq 8$, then*

$$\sup_{t\in[0,1]}\|\hat{\sigma}^2(t)-\sigma^2(t)\|_4 = O\left(c_n^2+\frac{1}{\sqrt{nc_n}}+b_n^2+\frac{1}{\sqrt{nb_n}}\right).$$

60

Proof. By (8.8), we have for some large constant C,

$$\sup_{t \in \mathcal{T}_n} \|\hat{\sigma}^2(t) - \sigma^2(t)\|_4 \le Cc_n^2 + \sup_{t \in \mathcal{T}_n} \left\|\frac{1}{nc_n} \sum_{i=1}^n (\hat{e}_i^2 - \mathbb{E}e_i^2) K_{c_n}(t_i - t)\right\|_4 \le$$

$$Cc_n^2 + \sup_{t \in \mathcal{T}_n} \left\|\frac{1}{nc_n} \sum_{i=1}^n (e_i^2 - \mathbb{E}e_i^2) K_{c_n}(t_i - t)\right\|_4 + \sup_{t \in \mathcal{T}_n} \left\|\frac{1}{nc_n} \sum_{i=1}^n (e_i^2 - \hat{e}_i^2) K_{c_n}(t_i - t)\right\|_4.$$

It is easy to verify that the first term satisfies

$$\sup_{t \in \mathcal{T}_n} \left\|\frac{1}{nc_n} \sum_{i=1}^n (e_i^2 - \mathbb{E}e_i^2) K_{c_n}(t_i - t)\right\|_4 = O\left(\frac{1}{\sqrt{nc_n}}\right).$$

By the proof of Lemma 8.3, we obtain (note that $\iota \ge 8$)

$$\sup_{t \in [0,1]} \|\hat{\mu}_{b_n}(t) - \mu(t)\|_8 = O\left(b_n^2 + \frac{1}{\sqrt{nb_n}}\right),$$

which yields (note that $\hat{e}_i = e_i + \mu(t_i) - \hat{\mu}_{b_n}(t_i)$)

$$\sup_{t \in \mathcal{T}_n} \left\|\frac{1}{nc_n} \sum_{i=1}^n (e_i^2 - \hat{e}_i^2) K_{c_n}(t_i - t)\right\|_4 = O\left(b_n^2 + \frac{1}{\sqrt{nb_n}}\right).$$

Hence $\sup_{t \in \mathcal{T}_n} \|\hat{\sigma}^2(t) - \sigma^2(t)\|_4 = O(c_n^2 + \frac{1}{\sqrt{nc_n}} + b_n^2 + \frac{1}{\sqrt{nb_n}})$. Similarly, using the estimate (8.9) we obtain $\sup_{t \in \mathcal{T}_n'} \|\hat{\sigma}^2(t) - \sigma^2(t)\|_4 = O(c_n^2 + \frac{1}{\sqrt{nc_n}} + b_n^2 + \frac{1}{\sqrt{nb_n}})$, which completes the proof. $\square$

Corollary 8.2. *Suppose the conditions of Lemma 8.4 hold, with $\iota \ge 8$. Then*

$$\left\|\sup_{t \in (0,1)} |\hat{\sigma}^2(t) - \sigma^2(t)|\right\|_4 = O\left(c_n^2 + \left(\frac{1}{\sqrt{nc_n}} + b_n^2 + \frac{1}{\sqrt{nb_n}}\right)c_n^{-1/4}\right).$$

Proof. The lemma follows from Proposition 8.1 in Section 8.2, the triangle inequality and simple calculations. Note that the first assumption of Proposition 8.1 is satisfied by the arguments in Corollary 8.1. The second assumption regarding the derivative can be shown by similar arguments as given in (8.6) and (8.7). $\square$

Corollary 8.3. *Suppose that the conditions of Corollary 8.2 hold. Let $\alpha = 1 - 4/\iota - \nu$ for some $\nu > 0$, and ι is defined in condition (A3) and condition (A4). Then we have*

$$\sup_{t \in [0, \tilde{t}_v - n^{-\alpha}] \cup [\tilde{t}_v + n^{-\alpha}, 1]} |\sigma_n^{2*}(t) - \sigma^2(t)| = O_p(\rho_n) ,$$

61

where $\rho_n = n^{-\alpha}c_n^{-5/4} + c_n^2 + \big(\frac{1}{\sqrt{nc_n}} + b_n^2 + \frac{1}{\sqrt{nb_n}}\big)c_n^{-1/4}$

Proof. Define

$$\bar{S}_{n,l}^k = \frac{1}{nc_n}\sum_{i=1}^{k}\Big(\frac{t_i - t}{c_n}\Big)^l K_{c_n}(t_i - t), \ \ \bar{R}_{n,l}^k = \frac{1}{nc_n}\sum_{i=1}^{k}\Big(\frac{t_i - t}{c_n}\Big)^l K_{c_n}(t_i - t)\hat{e}_i^2,$$

$$\underline{S}_{n,l}^k = \frac{1}{nc_n}\sum_{i=k}^{n}\Big(\frac{t_i - t}{c_n}\Big)^l K_{c_n}(t_i - t), \ \ \underline{R}_{n,l}^k = \frac{1}{nc_n}\sum_{i=k}^{n}\Big(\frac{t_i - t}{c_n}\Big)^l K_{c_n}(t_i - t)\hat{e}_i^2.$$

Recall the definition of $\sigma_n^{2*}(t)$ in (3.12). With the notation $k^* = \lfloor nt_n^* \rfloor$ elementary calculations yield

$$\tilde{\sigma}_1^2(t) := \tilde{\sigma}_{1,k^*}^2(t) = \big(\bar{R}_{n,0}^{k^*}(t)\bar{S}_{n,2}^{k^*}(t) - \bar{R}_{n,1}^{k^*}(t)\bar{S}_{n,1}^{k^*}(t)\big)/\big(\bar{S}_{n,0}^{k^*}(t)\bar{S}_{n,2}^{k^*}(t) - (\bar{S}_{n,1}^{k^*}(t))^2\big), \tag{8.10}$$

$$\tilde{\sigma}_2^2(t) := \tilde{\sigma}_{2,k^*+1}^2(t) = \big(\underline{R}_{n,0}^{k^*+1}(t)\underline{S}_{n,2}^{k^*+1}(t) - \underline{R}_{n,1}^{k^*+1}(t)\underline{S}_{n,1}^{k^*+1}(t)\big)/\big(\underline{S}_{n,0}^{k^*+1}(t)\underline{S}_{n,2}^{k^*+1}(t) - (\underline{S}_{n,1}^{k^*+1}(t))^2\big).$$

We will show below that

$$\sup_{|\tilde{t}_v - t_h| \leq n^{-\alpha}} \sup_{t \in [0, \tilde{t}_v - n^{-\alpha}]} |\tilde{\sigma}_{1,h}^2(t) - \sigma_1^2(t)| = O_p(\rho_n),$$

$$\sup_{|\tilde{t}_v - t_h| \leq n^{-\alpha}} \sup_{t \in [\tilde{t}_v + n^{-\alpha}, 1]} |\tilde{\sigma}_{2,h}^2(t) - \sigma_2^2(t)| = O_p(\rho_n).$$

The assertion of Corollary 8.3 then follows since $t_n^* - \tilde{t}_v = O_p(n^{-\alpha})$. For a proof of the two remaining estimates we assume without loss of generality $t_h > \tilde{t}_v$ and define

$$\bar{\mathcal{R}}_{n,l}^h(t) = \frac{1}{nc_n}\Big(\sum_{i=1}^{\lfloor n\tilde{t}_v \rfloor}\Big(\frac{t_i - t}{c_n}\Big)^l K_{c_n}(t_i - t)\hat{e}_i^2 + \sum_{i=\lfloor n\tilde{t}_v + 1 \rfloor}^{h}\Big(\frac{t_i - t}{c_n}\Big)^l K_{c_n}(t_i - t)\tilde{e}_i^2\Big),$$

where $\{\tilde{e}_i^2 | \lfloor n\tilde{t}_v \rfloor + 1 \leq i \leq t_h\}$ are independent of $\{\mathcal{F}_i, i \in \mathbb{Z}\}$ with bounded ι_{th} moment, $\mathbb{E}(\tilde{e}_i) = 0$, $\text{Var}(\tilde{e}_i) := \tilde{\sigma}^2(t_i)$ such that $\ddot{\tilde{\sigma}}^2(t_i)$ are Lipschitz continuous in $[\tilde{t}_v, t_h]$ and $\ddot{\tilde{\sigma}}^2(\tilde{t}_v^+) = \ddot{\tilde{\sigma}}^2(\tilde{t}_v) = \ddot{\sigma}^2(\tilde{t}_v^-) = \ddot{\sigma}_1^2(\tilde{t}_v-)$. Define

$$\tilde{\tilde{\sigma}}_{1,h}^2(t) = \big(\bar{\mathcal{R}}_{n,0}^h(t)\bar{S}_{n,2}^h(t) - \bar{\mathcal{R}}_{n,1}^h(t)\bar{S}_{n,1}^h(t)\big)/\big(\bar{S}_{n,0}^h(t)\bar{S}_{n,2}^h(t) - \big(\bar{S}_{n,1}^h(t)\big)^2\big). \tag{8.11}$$

Using similar arguments as given in the proof of Corollary 8.2 it follows that

$$\sup_{\lfloor n\tilde{t}_v \rfloor < h < \lfloor n\tilde{t}_v \rfloor + \lfloor n^{1-\alpha} \rfloor} \sup_{t \in [0,\tilde{t}_v - n^{-\alpha}]} |\tilde{\tilde{\sigma}}_{1,h}^2(t) - \sigma_1^2(t)| = O_p\left(c_n^2 + \left(\frac{1}{\sqrt{nc_n}} + b_n^2 + \frac{1}{\sqrt{nb_n}}\right)c_n^{-1/4}\right). \quad (8.12)$$

On the other hand, by Proposition 8.1 we can show that for $l = 0, 1$

$$\sup_{\lfloor n\tilde{t}_v \rfloor < h < \lfloor n\tilde{t}_v \rfloor + \lfloor n^{1-\alpha} \rfloor} \sup_{t \in [0,\tilde{t}_v - n^{-\alpha}]} |\bar{\mathcal{R}}_{n,l}^h(t) - \bar{R}_{n,l}^h(t)| = O_p(n^{-\alpha} c_n^{-5/4}).$$

Combining (8.10) and (8.11), the last estimate implies that

$$\sup_{\lfloor n\tilde{t}_v \rfloor < h < \lfloor n\tilde{t}_v \rfloor + \lfloor n^{1-\alpha} \rfloor} \sup_{t \in [0,\tilde{t}_v - n^{-\alpha}]} |\tilde{\tilde{\sigma}}_{1,h}^2(t) - \tilde{\sigma}_1^2(t)| = O_p(n^{-\alpha} c_n^{-5/4}).$$

By (8.12) we have

$$\sup_{\lfloor n\tilde{t}_v \rfloor < h < \lfloor n\tilde{t}_v \rfloor + \lfloor n^{1-\alpha} \rfloor} \sup_{t \in [0,\tilde{t}_v - n^{-\alpha}]} |\tilde{\sigma}_{1,h}^2(t) - \sigma_1^2(t)| = O_p\left(\rho_n\right), \quad (8.13)$$

and it is easy to see that

$$\sup_{\lfloor n\tilde{t}_v \rfloor < h < \lfloor n\tilde{t}_v \rfloor + \lfloor n^{1-\alpha} \rfloor} \sup_{t \in [\tilde{t}_v + n^{-\alpha}, 1]} |\tilde{\sigma}_{2,h}^2(t) - \sigma_2^2(t)| = O_p(\rho_n)$$

$$\sup_{\lfloor n\tilde{t}_v \rfloor - \lfloor n^{1-\alpha} \rfloor < h < \lfloor n\tilde{t}_v \rfloor \rfloor} \sup_{t \in [0,\tilde{t}_v - n^{-\alpha}]} |\tilde{\sigma}_{1,h}^2(t) - \sigma_1^2(t)| = O_p(\rho_n),$$

$$\sup_{\lfloor n\tilde{t}_v \rfloor - \lfloor n^{1-\alpha} \rfloor < h < \lfloor n\tilde{t}_v \rfloor} \sup_{t \in [\tilde{t}_v + n^{-\alpha}, 1]} |\tilde{\sigma}_{2,h}^2(t) - \sigma_2^2(t)| = O_p\left(\rho_n\right). \quad (8.14)$$

The assertion now follows from (8.13)–(8.14), the definition of $\sigma_n^{2*}(t)$, and the fact that $t_n^* - \tilde{t}_v = o_p(n^{-\alpha})$. $\qquad\square$

8.2   Two additional technical results

Proposition 8.1. *Let $\{\Upsilon_n(t)\}_{t \in [0,1]}$ be a sequence of stochastic processes with differentiable paths. Assume that for some $p \geq 1$ and any $t \in [0,1]$, $\|\Upsilon_n(t)\|_p = O(m_n), \|\dot{\Upsilon}_n(t)\|_p = O(l_n)$, where m_n, l_n are sequences of real numbers, $m_n = O(l_n)$, then*

$$\left\| \sup_{t \in [0,1]} |\Upsilon_n(t)| \right\|_p = O\left(m_n \left(\frac{m_n}{l_n}\right)^{-\frac{1}{p}}\right).$$

In particular, if $p = 2$, we have $\|\sup_{t \in [0,1]} |\Upsilon_n(t)|\|_2 = O(\sqrt{m_n l_n})$.

Proof. For a sequence b_n define $\tilde{b}_n = \lfloor b_n^{-1} \rfloor$ and let $\tau_i = ib_n$, $i = 1, 2, ..., \tilde{b}_n$ and $\tau_i = 1$ for $i = \tilde{b}_n + 1$. Then by the triangle inequality, we have

$$\sup_{t \in (0,1)} |\Upsilon_n(t)| \leq \max_{0 \leq i \leq \tilde{b}_n + 1} |\Upsilon_n(\tau_i)| + \max_{1 \leq i \leq \tilde{b}_n + 1} Z_{in},$$

where $Z_{in} = \sup_{\tau_i - b_n < t < \tau_i} |\Upsilon_n(t) - \Upsilon_n(\tau_i)|$. Observing the inequalities

$$\|Z_{in}\|_p \leq \left\| \int_{\tau_i - b_n}^{\tau_i} |\dot{\Upsilon}(t)| dt \right\|_p \leq \int_{\tau_i - b_n}^{\tau_i} \|\dot{\Upsilon}_n(t)\|_p dt = O(b_n l_n)$$

and $\max_{1 \leq i \leq \tilde{b}_n + 1} Z_{in}^p \leq \sum_{i=1}^{\tilde{b}_n + 1} Z_{in}^p$, we have

$$\left\| \max_{1 \leq i \leq \tilde{b}_n + 1} Z_{in} \right\|_p = O((l_n^p b_n^{(p-1)})^{1/p}) = O(l_n b_n^{(p-1)/p}).$$

Similarly, we obtain the estimate $\| \max_{0 \leq i \leq \tilde{b}_n + 1} |\Upsilon_n(t_i)| \|_p = O_p(b_n^{-1/p} m_n)$, and picking $b_n = m_n/l_n$ proves the assertion. $\qquad \square$

Proposition 8.2. *Suppose A_n are sets such that $\mathbb{P}(A_n) \to 0$ as $n \to \infty$, and $X_n I(\bar{A}_n) = O_p(1)$. Then $X_n = O_p(1)$.*

Proof. For any $\epsilon > 0$, let N be a large constant such that $\mathbb{P}(A_n) \leq \epsilon/2$ for $n \geq N$, and M be a large constant such that $\mathbb{P}(|X_n|I(\bar{A}_n) \geq M/2) \leq \epsilon/2$ for $n \geq N$. Then

$$\mathbb{P}(|X_n| \geq M) \leq \mathbb{P}(|X_n|I(A_n) \geq M/2) + \mathbb{P}(|X_n|I(\bar{A}_n) \geq M/2)$$
$$\leq \mathbb{P}(A_n) + \mathbb{P}(|X_n|I(\bar{A}_n) \geq M/2) \leq \epsilon$$

for all $n \geq N$. $\qquad \square$