Change Point Analysis of Correlation in Non-stationary Time Series

HOLGER DETTE, WEICHI WU and ZHOU ZHOU

RUHR-UNIVERSITÄT Bochum, UNIVERSITY COLLEGE London and UNIVERSITY OF Toronto

Abstract

A restrictive assumption in change point analysis is “stationarity under the null hypothesis of no change-point”, which is crucial for asymptotic theory but not very realistic from a practical point of view. For example, if change point analysis for correlations is performed, it is not necessarily clear that the mean, marginal variance or higher order moments are constant, even if there is no change in the correlation. This paper develops change point analysis for the correlation structures under less restrictive assumptions. In contrast to previous work, our approach does not require that the mean, variance and fourth order joint cumulants are constant under the null hypothesis. Moreover, we also address the problem of detecting relevant change points.

Key words: piecewise locally stationary process, change point analysis, relevant change points, second order structure, local linear estimation

1 Introduction

Change point analysis is a well studied subject in the statistical and econometric 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

1 Corresponding author. Department of Statistics, University College, Gower Street London, WC1E 6BT, UK
E-mail: w.wu@ucl.ac.uk
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. 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. Galeano and Peña (2007) and Aue et al. (2009) suggested nonparametric tests for structural breaks in the variance matrix of a multivariate time series. There exist also several papers discussing change point analysis in the second order structure of a time series. For example, Berkes et al. (2009) and Killick et al. (2013) considered the more classical problem of a change point in a correlation at fixed lag. Davis et al. (2006) and Preuss et al. (2015) proposed methods for detecting multiple breaks in piecewise stationary processes.

This list of references is by no means complete but an important and common feature of the cited references and most of the literature on testing for structural breaks in the covariance or correlation structure (at different lags) 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 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 for change point in the variance 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 time 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 correlation even if the mean and the variances change gradually in time. In this case the classical approach is not applicable.

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 lag-$k$ correlation. In Section 2 we introduce piecewise locally stationary processes as considered by Zhou (2013) who investigated the properties of the classical CUSUM test for the mean under non-stationarity. Section 3 is devoted to the “classical” change point problem for a (vector) of correlations at different lags in a piecewise locally stationary process. In the simplest case of one lag-1 autocorrelation, say $\rho_i = \text{Corr}(X_i, X_{i+1})$ the hypothesis can be formulated as

$$H_0 : \rho_i = \rho_j \text{ for all } i, j = 1, \ldots, n \text{ versus } H_1 : \rho_i \neq \rho_j \text{ for some } i \neq j.$$ (1.1)

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 is proved. The methodology is very general and applicable in many situations where the assumptions of classical tests are not satisfied. In particular, we neither assume that the mean, variance or higher order joint cumulants of the non-stationary sequence are constant nor that the change in the variance and the lag-$k$ correlation occur at the same location.
Furthermore, we show that the stochastic errors produced in the nonparametric estimation of the mean and variance function are asymptotically negligible in the second-order CUSUM statistic. This result is of particular interest, and non-trivial because the order of the stochastic errors of the nonparametric estimates is larger than the $1/\sqrt{n}$ convergence rate of the CUSUM test.

The situation is more complicated if one is interested in such sophisticated hypotheses 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

$$v_1 = \rho_1 = \ldots = \rho_k \neq v_2 = \rho_{k+1} = \ldots = \rho_n,$$

and is interested in hypotheses of the form

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

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 (2016) 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) are expensive and should therefore only be performed if changes have serious consequences. Testing a hypothesis of the form (1.3) to detect a structural break also avoids the consistency problem mentioned in Berkson (1938): any test will detect negligible changes in the parameter if the sample size is sufficiently large. Dette and Wied
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 testing problem in the form (1.3) is appealing, the construction of corresponding tests faces several mathematical challenges. In particular, even under the null hypothesis of a non-relevant change point one has to deal with the problem of non-stationarity. For example, Dette and Wied (2016) 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. In the context of change point analysis for correlations this means that the mean and the variances of the process have to be constant before and after the change point. From a practical point of view this assumption seems to be very strong and not very realistic.

Section 4 is devoted to the problem of testing the hypothesis of a non-relevant change in the several correlations at different lags. We use the CUSUM approach proposed in Dette and Wied (2016) to obtain a test for the hypothesis (1.3) and its analogue in the case of lag-k correlations. Asymptotic normality of a corresponding $L_2$-type statistic is established and a wild bootstrap method is developed that 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. 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 USD/CAD exchange rate series and illustrate the usefulness of the proposed methodology in identifying second order change points in modeling volatilities. All proofs and technical details are deferred to an online supplement (see also Dette et al. (2015)).
2 Piecewise locally stationary processes

We start writing some notations of frequent use. For an \( l \)-dimensional (random) vector \( \mathbf{v} = (v_1, \ldots, v_l) \), \( l \geq 1 \), let \( |\mathbf{v}| = (\sum_{i=1}^{l} v_i^2)^{1/2} \). A random vector \( \mathbf{V} \) is said to be in \( L^q \), \( q > 0 \), if \( \mathbb{E}(|\mathbf{V}|^q) < \infty \). In this case write \( \|\mathbf{V}\|_q = (\mathbb{E}|\mathbf{V}|^q)^{1/q} \), and \( \|\mathbf{V}\| = \|\mathbf{V}\|_2 \). The symbol \( \xrightarrow{\mathcal{D}} \) means weak convergence of real-valued random variables (convergence in distribution). For any interval \( \mathcal{I} \subset \mathbb{R} \) and nonnegative integer \( q \) let \( C^q(\mathcal{I}) \) be the set of \( q \) times continuously differentiable functions \( f : \mathcal{I} \to \mathbb{R} \) and \( C(\mathcal{I}) = C^0(\mathcal{I}) \). Let \( \{\varepsilon_i\}_{i \in \mathbb{Z}} \) denote a sequence of independent identically distributed (i.i.d.) random variables and denote by \( \mathcal{F}_i = \sigma(\ldots, \varepsilon_0, \ldots, \varepsilon_{i-1}, \varepsilon_i) \) the sigma field generated by \( \{\varepsilon_j|j \leq i\} \). We define the sigma field \( \mathcal{F}_i^{(j)} = \sigma(\ldots, \varepsilon_{j-1}, \varepsilon'_j, \varepsilon_{j+1}, \ldots, \varepsilon_i) \), 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. For any real number \( a \), write \( [a] \) be the largest integer which \( \leq a \). Let \( \mathbf{1}(\cdot) \) be the indicator function, \( \text{sign}(\cdot) \) be the usual sign function, such that \( \text{sign}(x) = \mathbf{1}(x \geq 0) - \mathbf{1}(x < 0) \). Define \( 0/0 = 1 \). Let \( a \wedge b \) denote \( \min(a, b) \) for \( a, b \in \mathbb{R} \). Throughout the paper we consider the case that type I error \( \alpha \leq 0.05 \). We discuss autocorrelation in the rest of the paper, and use the term “correlation” for “autocorrelation” for short. Our method can be applied to cross correlation without further difficulty.

We consider the model

\[
Y_i = \mu(t_i) + e_i, \quad i = 1, \ldots, n, \tag{2.1}
\]

where (for the sake of simplicity) \( t_i = i/n \) \( (i = 1, \ldots, n) \) and \( \mu(\cdot) \) is a smooth function. 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 changes of the mean in the situation of centered 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). 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 correlations. More precisely we consider an error process \( \{e_i\}_{i=1}^n \) in (2.1) that is piecewise locally stationary (PLS) with \( r \) breaks for some \( r \in \mathbb{N} \). Formally, we use a definition for a PLS process and the concept of “physical dependence measure for PLS” that 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 \( 0 = b_0 < b_1 < ... < b_r < b_{r+1} = 1 \) and nonlinear filters \( G_0, G_1, ..., G_r \), such that

\[
e_i = e_i(t_i), \quad \text{where } e_i(t) = G_j(t, F_i), \quad \text{if } b_j < t_i \leq b_{j+1}
\]

where \( F_i = \sigma(\ldots, \varepsilon_0, \ldots, \varepsilon_{i-1} \varepsilon_i) \), 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 \( k \)th physical dependence measure in \( L_p \)-norm as

\[
\delta_p(k) = \max_{0 \leq i \leq r} \sup_{b_i < t \leq b_{i+1}} \| G_i(t, F_k) - G_i(t, F_k^*) \|_p,
\]

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

The PLS process is a natural non-stationary extension of many well known statistical processes, with the dependence measure easy to calculate.

**Example 2.1. (PLS linear process)** For \( \{\varepsilon_i\}_{i \in \mathbb{Z}} \) take \( F_i = \sigma(\{\varepsilon_j|j \leq i\}) \), and consider the process

\[
G_j(t, F_i) = \sum_{s=0}^{\infty} a_{j,s}(t) \varepsilon_{i-s} b_j < t \leq b_{j+1} \quad 0 \leq j \leq r,
\]
where \(0 = b_0 < b_1 < ... < b_{r+1} = 1\) are unknown break points, \(a_{j,s}(t)\) for \(b_j < t \leq b_{j+1}\) \(0 \leq j \leq r\), \(s \in \mathbb{Z}\) are Lipschitz continuous functions. Straightforward calculations show that 
\[
\delta_p(k) = O(\max_{0 \leq j \leq r} \sup_{b_j < t \leq b_{j+1}} |a_{j,k}(t)|) \quad \text{provided } ||\varepsilon_0||_p < \infty.
\]
Model (2.2) is a time-varying MA process with possible abrupt changes. For smooth time-varying MA process, it could be shown, for example in Zhang and Wu (2012) that it well-approximates the locally stationary autoregressive processes that have been studied extensively in the literature (see for example Dahlhaus (1997) among others).

**Example 2.2. (PLS nonlinear process)** For \(\{\varepsilon_i\}_{i \in \mathbb{Z}}\) take \(F_i = \sigma(\{\varepsilon_j | j \leq i\})\) and consider the process

\[
G_j(t, F_i) = R_j(t, G_j(t, F_{i-1}), \varepsilon_i), \quad b_j < t \leq b_{j+1}, \quad 0 \leq j \leq r,
\]

(2.3)

where \(0 = b_0 < b_1 < ... < b_{r+1} = 1\) are unknown break points. Many important nonlinear time series have the form \(X_i = R(X_{i-1}, \varepsilon_i)\). Typical examples include (G)ARCH models (see Engle (1982) Bollerslev (1986)), threshold models (see Tong (1990)) and bilinear models. It can be shown similarly to Zhou and Wu (2009) that, under some mild conditions, \(\delta_p(k) = O(\chi^k)\) for some \(\chi \in (0,1)\), and that \(\chi\) can be evaluated as

\[
\chi := \max_{0 \leq j \leq r} \sup_{t \in (b_j, b_{j+1}]} \sup_{x \neq y} \frac{||R_j(t,x,\varepsilon_0) - R_j(t,y,\varepsilon_0)||_p}{|x - y|}.
\]

(2.4)

For our asymptotic analysis we list some conditions.

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

\[
||G_i(t, F_0) - G_i(s, F_0)||_\varepsilon \leq C|t - s|
\]

holds for \(i \geq 8\). In addition, \(\mathbb{E}[e_i] = 0\) for all \(1 \leq i \leq n\), and there is a variance function
\[ \sigma^2(\cdot) : [0,1] \to \mathbb{R}_+, \text{ such that } \sigma^2_i := \sigma^2(t_i) = \text{Var}(e_i), \text{ for } i = 1, \ldots, n. \]

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

(A3) \( \max_{0 \leq i \leq r} \sup_{t \in (b_i, b_{i+1})} \|G_i(t, F_0)\|_i < \infty \) for some \( \iota \geq 8 \).

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

Remark 2.1.

a) The bound of \( \max_{1 \leq i \leq n} \|e_i\|_p \) in Definition 2.1 does not depend on \( n \). This simplifies the assumptions and the proofs in the subsequent discussion. It is also possible to develop corresponding results for an \( n \)-dependent bound with added complications in the technical arguments.

b) For the sake of brevity we use the condition \( \iota \geq 8 \) in (A3) and (A4). Using additional technical arguments it can be shown that our methodology is still valid for innovations with a heavier tail (see also Section 5 for some simulation results with heavy-tailed distributions).

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

3 Tests for changes in correlations

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 \( \mu(\cdot) \) is an unknown deterministic trend. We are interested in testing nonparametrically the “classical” hypothesis of a change point in the correlations. 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

9
statistic are not directly available. For this we develop a bootstrap CUSUM-type test for the “classical” hypotheses of a change point in correlations based on residuals from a local linear fit. For the definition of the local linear estimator we assume 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 $K_b(\cdot) = K(\frac{\cdot}{b})$. We assume that $K \in C^2([-1,1])$. and, for convenience, we set $e_i = 0, \hat{e}_i = 0$ if $i > n$, where $n$ is the sample size.

Consider the problem of testing whether there are changes in correlations $\rho_{i,k} := Corr(Y_i, Y_{i+k})$ for some pre-specified lag-$k$'s, with

$$H_0 : \rho_{i,k} = \rho_{j,k} = \rho_k \text{ for all } i, j = 1, \ldots, n, k = r_1, \ldots, r_l$$

$$H_1 : \text{There exists } 1 \leq s \leq l \text{ and } i \neq j \text{ such that } \rho_{i,r_s} \neq \rho_{j,r_s}$$

where the integers $r_1 < r_2 < \cdots < r_l$ define the lags of interest. A test for the classical hypothesis for 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 locally stationary framework (including non constant mean or variance) requires estimates of the mean and variances. For this purpose consider the CUSUM statistic

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

where $\hat{S}_i = (\hat{S}_i^{(r_1)}, \ldots, \hat{S}_i^{(r_l)})$, $\hat{S}_i^{(j)} = \sum_{s=1}^{i} \frac{\hat{e}_s \hat{e}_{s+j}}{\partial^2(t_s)}$, $\hat{e}_s = Y_s - \hat{\mu}_{b_n}(t_s)$, $\hat{e}_{s+j} = Y_{s+j} - \hat{\mu}_{b_n}(t_{s+j})$, and $\hat{\mu}_{b_n}(\cdot)$ is the local linear estimator of the function $\mu(\cdot)$ with bandwidth $b_n$,

$$(\hat{\mu}_{b_n}(t), \hat{\mu}_{b_n}(t)) = \arg\min_{\beta_0, \beta_1} \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1(t_i - t))^2 K_{b_n}(t_i - t)$$

(see [Fan and Gijbels (1996)]).

We allow the variance to possibly have a structural break at a point, say $\tilde{t}_v$ that need not
coincide with the location of the change point in any of the lag- \( k \) correlations. We assume that \( \hat{\sigma}^2(\cdot) \) is Lipschitz continuous on the intervals \((0, \hat{t}_v)\) and \((\hat{t}_v, 1)\) and that there exists a constant \( \zeta > 0 \), such that \( \hat{t}_v \in [\zeta, 1 - \zeta] \). We define an estimator, say \( t^*_n \), of the change point \( \hat{t}_v \) in the variance by

\[
t^*_n = \arg\max_{\lfloor n\zeta \rfloor \leq i \leq n - \lfloor n\zeta \rfloor + 1} |M(i)|/n, \tag{3.5}
\]

where

\[
M(i) = \frac{1}{L} \left( \sum_{j=i-L+1}^{i} \hat{e}_j^2 - \sum_{j=i}^{i+L-1} \hat{e}_j^2 \right) \tag{3.6}
\]

and \( L \in \mathbb{N} \) is a regularization parameter that increases with \( n \). The maximum in (3.5) is not taken over the full range \( 1 \leq i \leq n \), as recommended in \cite{Andrews1993} (see also \cite{Qu2008}). We estimate \( \sigma^2(t_i) \) by \( \hat{\sigma}^2(t_i) = \hat{\sigma}_{c_n,b_n}^2(t, nt^*_n) \), where for \( k = 1, \ldots, n \)

\[
\hat{\sigma}_{c_n,b_n}^2(t, k) = \hat{\sigma}_{c_n,b_n}^2(t, k-1)1(t \leq k/n) + \hat{\sigma}_{c_n,b_n}^2(t, k+)1(t > k/n)
\]

and

\[
(\hat{\sigma}_{c_n,b_n}^2(t, k-), \hat{\sigma}_{c_n,b_n}^2(t, k-)) = \arg\min_{\beta_0, \beta_1} \sum_{i=1}^{k} (\hat{e}_i^2 - \beta_0 - \beta_1(t_i - t))^2 K_{c_n}(t_i - t),
\]

\[
(\hat{\sigma}_{c_n,b_n}^2(t, k+), \hat{\sigma}_{c_n,b_n}^2(t, k+)) = \arg\min_{\beta_0, \beta_1} \sum_{i=k+1}^{n} (\hat{e}_i^2 - \beta_0 - \beta_1(t_i - t))^2 K_{c_n}(t_i - t). \tag{3.7}
\]

We take the (non-observable) analogue of \( \hat{S}_i^{(j)} \) to be

\[
S_i^{(j)} = \sum_{s=1}^{i} W_s^{(j)}, \tag{3.8}
\]
where $W_s^{(j)} = \frac{e_s e_{s+j}}{\sigma(t_s) \sigma(t_{s+j})}$, and consider the random variable
\[
T_n = \max_{1 \leq i \leq n-r_l} \left| S_i - \frac{i}{n} S_n \right|,
\]
where $S_i = (S_i^{(r_1)}, ..., S_i^{(r_l)})$. It is easy to see that $W_i^{(j)}$ is $\mathcal{F}_{i+j}$ measurable and that the process $(W_i^{(j)})_{i=1}^{n-j}$ is PLS. Moreover, $\{W_i^{(k)} k = r_1, ..., r_l\}_{1 \leq i \leq n-r_l}$ can be modeled by an $l$-dimensional PLS process. Take $q$ as the number of break points, $0 = v_0 < v_1 < ... < v_{q+1} = 1$ as the corresponding locations of the breaks, and $\mathbf{H}$ as the corresponding nonlinear filters, $(W_i^{(r_1)}...W_i^{(r_l)})^T = \mathbf{H}_j(t_i, \mathcal{F}_{i+r_l})$ if $v_j < t_i \leq v_{j+1}$, $0 \leq j \leq q$.

The following result shows that $t_n^*$ is a consistent estimate of $\tilde{t}_v$; its proof can be found in Section A.1 of the online supplement.

**Lemma 3.1.** Assume that $nb_n^b \to 0$, $nb_n^3 \to \infty$ and that (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 $\sigma^2(\cdot)$ is Lipschitz continuous (here $\tilde{t}_v$ is the location of the change point of the variance such that $\zeta \leq \tilde{t}_v \leq 1 - \zeta$). Then the estimator $t_n^*$ defined in (3.5) satisfies $t_n^* - \tilde{t}_v = o_p(n^{-1-4/\iota} \log n)$.

**Remark 3.1.** The rate of convergence of the estimator $t_n^*$ is arbitrarily close to the optimal rate $n^{-1}$ subject to a logarithmic factor if (A1) and (A4) hold for any $\iota > 0$.

The rates of convergence of the estimators (3.4) and (3.7) are of the order $n^{-2/5}$ under suitable bandwidth conditions. Thus, a naive plug-in argument of $\hat{\mu}(t_i)$ does not lead to the crucial result that
\[
|\hat{T}_n - T_n| = o_p(\sqrt{n}),
\]
that is required for constructing the hypothesis testing procedure. In the Appendix we demonstrate that the estimate (3.10) is in fact valid using delicate arguments to overcome the slow rate of convergence of the non-parametric fit. Then the weak convergence of the statistic $\hat{T}_n/\sqrt{n}$ follows from the weak convergence of $T_n/\sqrt{n}$, which can be established under
an additional assumption.

(A5) The long run variance function

\[ \kappa^2(t) = \sum_{k=-\infty}^{\infty} \text{cov}(H_i(t, F_k), H_i(t, F_0)) \in \mathbb{R}^{l \times l} \quad \text{if } t \in (v_i, v_{i+1}], 0 \leq i \leq q, \]  

(3.11)

and \( \kappa^2(0) := \lim_{t \downarrow 0} \lambda_{\min}(\kappa^2(t)) \) exists with \( \inf_{t \in [0,1]} \lambda_{\min}(\kappa^2(t)) > 0 \), where for any positive semi-definite matrix \( A \), \( \lambda_{\min}(A) \) denotes the minimal eigenvalue of matrix \( A \).

The proof of the following result is deferred to the online appendix.

**Theorem 3.1.** Assume that \( b_n \to 0 \), \( c_n/b_n \to 0 \), \( c_n b_n^{-2} \to \infty \), \( n c_n^4 \to 0 \), \( nb_n^6 c_n^{-1/2} \to 0 \), \( nb_n^6 c_n^{1/2} \to \infty \) and suppose that (A1) - (A5) are satisfied with \( \iota \geq 8 \). Assume that the variance \( \sigma^2(t) > 0 \). Suppose that one of the following conditions is satisfied.

(i) \( \sigma^2(\cdot) \) is twice differentiable on \([0,1]\) and the second derivative \( \ddot{\sigma}^2(\cdot) \) is Lipschitz continuous.

(ii) \( \sigma^2(\cdot) \) has one abrupt change point \( \tilde{t}_v \in [\zeta, 1-\zeta] \), and on the intervals \([0, \tilde{t}_v) \) and \((\tilde{t}_v, 1]\), \( \sigma^2(t) \) is twice differentiable and the second derivative \( \ddot{\sigma}^2(\cdot) \) is Lipschitz continuous.

Then under the null hypothesis (3.1) we have

\[ \frac{1}{\sqrt{n}} \hat{T}_n \overset{d}{\to} K_1 := \sup_{t \in (0,1)} |U(t) - tU(1)|, \]  

(3.12)

where \( \{U(t)\}_{t \in [0,1]} \) is a zero mean \( l \)-dimensional Gaussian process with covariance function

\[ \gamma(t, s) = \int_{0}^{\min(t, s)} \kappa^2(r) dr. \]  

(3.13)

As a consequence of Theorem 3.1 we obtain - in principle - an asymptotic level \( \alpha \) 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 \( K_1 \) in (3.12). However, under non-stationarity
(more precisely under the PLS assumption), the function \( \kappa^2(t) \) defined in (3.11) and, as a consequence, the covariance structure of the Gaussian process \( \{U(t) - tU(1)\}_{t \in [0,1]} \) involves the complicated dependence structure of the data generating process.

Due to the PLS structure, the covariance structure of the Gaussian process \( U(\cdot) \) and the quantiles of the limiting distribution in Theorem 3.1 are hard to estimate. As an alternative, a data-driven critical value will be derived using a wild bootstrap method to mimic the distributional properties of the Gaussian process \( U(\cdot) \). Following Zhou (2013) we define for a fixed window size, say \( m \), the quantities

\[
\hat{\Phi}_{i,m} = \frac{1}{\sqrt{m(n - m + 1)}} \sum_{j=1}^{i} \left( \hat{S}_{j,m} - \frac{m}{n} \hat{S}_n \right) R_j, \quad i = 1, \ldots, n - m + 1, \tag{3.14}
\]

where \( \hat{S}_{j,m} = (S_{j,m}^{(r_1)}, \ldots, S_{j,m}^{(r_l)})^T \), \( \hat{S}_n = \hat{S}_{1,n} \), \( \hat{S}_{j,m}^{(k)} = \sum_{r=j}^{j+m-1} \frac{\hat{e}_r \hat{e}_r}{\hat{\sigma}_2(r)} \), and \( \{R_i\}_{i \in \mathbb{Z}} \) is a sequence of i.i.d standard normal distributed random variables independent of \( \{\varepsilon_i\}_{i \in \mathbb{Z}} \).

**Theorem 3.2.** If the conditions of Theorem 3.1 are satisfied and, for \( m \to \infty \), assume \( m/\sqrt{n} \to 0, \sqrt{m}(c_n^2 + (1/\sqrt{nc_n} + b_n^2 + 1/\sqrt{nb_n})c_n^{-1/4}) \log n \to 0 \) (conditional on \( \mathcal{F}_n \) in probability)

\[
M_n = \max_{m+1 \leq i \leq n-m+1} \left| \hat{\Phi}_{i,m} - \frac{i}{n - m + 1} \hat{\Phi}_{n-m+1,m} \right| \overset{P}{\to} \mathcal{K}_1,
\]

where the random variable \( \mathcal{K}_1 \) is defined in (3.12).

Theorem 3.2 provides an asymptotic level \( \alpha \) test for the hypothesis of constant correlations in model (2.1) with critical values obtained by resampling. The proof is deferred to the online supplement. The details of generating the critical values and performing the test are summarized in an algorithm.

**Algorithm 3.1.**

[1] Calculate the statistic \( \hat{T}_n \) at (3.3).

[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} \text{ defined in (3.14)} \text{ and calculate}

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

[3] If \( M(1) \leq M(2) \leq \ldots \leq M(B) \) denote the order statistics of \( M_1, \ldots, M_B \), null hypothesis of constant correlations is rejected at level \( \alpha \) when

\[ \hat{T}_n / \sqrt{n} > M_{\lfloor B(1-\alpha) \rfloor}. \] (3.15)

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 3.2.**

(1) If the sequence \( b_n \) is of order \( n^{-1/5} \) and \( m \) is of order \( n^{1/3} \), then the bandwidth conditions of Theorem 3.2 hold if the sequence \( c_n \) is of order \( n^{-\beta} \), where \( \beta \in \left( \frac{12}{14}, \frac{14}{15} \right) \).

(2) It follows by similar arguments as in the proof of Theorem 2, Proposition 3 of Zhou (2013) and Lemma B.1 and Lemma B.2 in the online supplement, that the bootstrap test (3.15) is consistent. For \( 1 \leq s \leq l \), write \( \rho_{rs}(t_i) = \rho_{i,r_s}, \rho(\cdot) = (\rho_{r_1}(\cdot), \ldots, \rho_{r_l}(\cdot))^T \). It can be shown that the bootstrap is able to detect local alternatives of the form \( \rho(\cdot) = \rho_0 + n^{-1/2} f(\cdot) \), where \( f(\cdot) \) is a nonconstant piecewise Lipschitz continuous \( l \)-dimensional vector function.

### 4 Relevant changes of correlations

After a change point has been detected and localized a modification of the statistical analysis is necessary, one that addresses the different features of the data generating process before and after the change point. Dette and Wied (2016) 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. Inference might be robust with respect to small changes of the correlation structure, but 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. Here we investigate the hypothesis (1.3) of a non-relevant change point for correlations in a general non-stationary context under the assumption of PLS.

Consider model (2.1) and suppose that there exist time points $t_k \in (0, 1)$, $k = r_1, ..., r_l$, such that

$$
\rho_1^{(k)} = \rho_{1,k} = \ldots = \rho_{[nt_k],k}, \quad \rho_2^{(k)} = \rho_{[nt_k]+1,k} = \ldots = \rho_{n,k}.
$$

We are interested in testing the hypotheses

$$
H_0 : \mid \rho_1^{(k)} - \rho_2^{(k)} \mid \leq \delta_k \text{ for all } k = r_1, \ldots, r_l \quad (4.1)
$$

$$
H_1 : \text{There exists a lag } k \in \{r_1, \ldots, r_l\} \text{ such that } \mid \rho_1^{(k)} - \rho_2^{(k)} \mid > \delta_k, \quad (4.2)
$$

where $\delta_{r_1}, \ldots, \delta_{r_l}$ are given thresholds. Problems of this type have recently been discussed in [Dette and Wied (2016)] under assumptions that are not practically tenable. In the PLS framework, these assumptions will be relaxed. However, under these more general assumptions, the construction of a test and the investigation of its asymptotic properties is substantially more difficult, as described in the following paragraphs.

We denote by, for $1 \leq s \leq l$, $\Delta_{rs} = \rho_2^{(rs)} - \rho_1^{(rs)}$ the (unknown) difference before and after the change point and assume here that, under the null hypothesis of a non-relevant change in the correlations, the variance function $\sigma^2(\cdot)$ has either no jumps or has a jump at a point, say $\hat{t}_v$, that need not coincide with any of the change point $t_k$ in the correlation structure. We define the CUSUM process, for $k = r_1, ..., r_l$, by

$$
\hat{V}_n^{(k)}(s) = \frac{1}{n} \sum_{j=1}^{[ns]} \hat{e}_j \hat{e}_{j+k} - \frac{[ns]}{n} \sum_{j=1}^{n} \hat{e}_j \hat{e}_{j+k}
$$

where $\hat{e}_i = Y_i - \hat{\mu}_{bn}(t_i)$ denotes the nonparametric residuals from the local linear fit while using the convention that $\hat{e}_i = 0$ for $i > n$. The estimator for the change point of the
correlation structure at lag-\(k\) is taken to be

\[ \hat{t}_n^{(k)} = \arg\max_{1 \leq m \leq n} \left( \hat{V}_n^{(k)}(m/n) \right)^2/n. \]  
\(4.4\)

The statistic \(\hat{t}_n^{(k)}\) depends on the estimator \(t_n^*\) for the change point in the variance as defined in (3.5). The estimator is consistent (a proof can be found in the online supplement.)

**Lemma 4.1.** Suppose that one of the following conditions holds.

(i) Conditions of Lemma 3.1 are satisfied.

(ii) \(\sigma^2(\cdot)\) is twice differentiable on \([0,1]\) and the second derivative \(\bar{\sigma}^2(\cdot)\) is Lipschitz continuous.

In addition, suppose the conditions for the bandwidths \(b_n\) and \(c_n\) of Theorem 3.1 hold. Then for any \(k = r_s, 1 \leq s \leq l\), the estimate \(\hat{t}_n^{(k)}\) of the change point in the correlation structure at lag-\(k\) defined by (4.4) satisfies

\[ \hat{t}_n^{(k)} \overset{D}{\to} T_{\text{max}}^{(k)} \text{ if } |\Delta_k| = 0 \]  
\(4.5\)

\[ |\hat{t}_n^{(k)} - t_k| = O_p(n^{-\nu}), \text{ if } |\Delta_k| > 0 \]  
\(4.6\)

for some \(\nu \in (1/2, 2/3)\), where \(T_{\text{max}}^{(k)}\) is a \([0,1]\)-valued random variable.

The test for the hypothesis of a non-relevant change is based on the statistic

\[ \hat{T}_n^{(k),r} = \frac{3}{(\hat{t}_n^{(k)})^2(1 - \hat{t}_n^{(k)})^2} \int_0^1 (\hat{V}_n^{(k)}(s))^2 ds, \]  
\(4.7\)

where the process \(\{\hat{V}_n^{(k)}(s), 0 \leq s \leq 1\}\) is defined in (4.3). We show that \(\hat{T}_n^{(r_s),r}\) is a consistent estimator of \(\Delta_{r_s}^2 = (\rho_1^{(r_s)} - \rho_2^{(r_s)})^2\) for \(s = 1, \ldots, l\), and provide its asymptotic distribution.

**Theorem 4.1.** Assume that the conditions for the bandwidths \(b_n\) and \(c_n\) of Theorem 3.1 hold and that (A1) - (A4) are satisfied with \(\iota \geq 16\).
(i) If $\Delta_k \neq 0$ for $k = r_1, \ldots, r_l$, then

$$
\left\{ \sqrt{n} \frac{T_n^{(r_s),r} - \Delta^2_{r_s}}{|\Delta_{r_s}|} \right\}_{s=1}^l \xrightarrow{D} Z := \left\{ \frac{Z^{(r_s)} \Delta_{r_s}}{|\Delta_{r_s}|} \right\}_{s=1}^l ,
$$

where

$$Z^{(r_s)} := \frac{6}{t_{r_s}^2 (1 - t_{r_s})^2} \int_0^1 \left( U^{(s)}(u) - u U^{(s)}(1) \right) (u t_{r_s} - u \wedge t_{r_s}) \, du ,
$$

with the process $\{U(u)\}_{u \in [0,1]} = \{(U^{(1)}(u), \ldots, U^{(l)}(u))^T\}_{u \in [0,1]}$ as defined in Theorem 3.1.

(ii) If $\Delta_{r_s} = 0$ for some $1 \leq s \leq l$, then $\hat{T}_n^{(r_s),r} = O_P(1/n)$, the $s$th coordinate of the process on the left side of (4.8) degenerates.

A careful inspection of the proof of Theorem 4.1 shows that (4.8) remains that for any estimator of the change point in the correlation structure that satisfies (4.5) and (4.6) (for $\upsilon > 1/2$) for any given fixed lag-k’s. Theorem 4.1 yields an asymptotic level $\alpha$ test for the hypothesis (4.1) of a non-relevant change in the correlation structure by rejecting $H_0$, whenever

$$
\hat{T}_{n,\text{max}} := \max_{1 \leq s \leq l} \frac{\hat{T}_n^{(r_s),r} - \delta^2_{r_s}}{\delta_{r_s}} > \tilde{v}_{1-\alpha} \sqrt{n} \quad (4.10)
$$

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

$$\max_{1 \leq s \leq l} \left\{ \frac{Z^{(r_s)} \Delta_{r_s}}{|\Delta_{r_s}|} \right\},$$

$Z^{(r_s)}$ as defined in (4.9). This distribution is a maximum of $l$-variate centered normal distributions with a covariance depending on the data generating process in a complicated way, in particular on the long run variance as (3.11). We construct a bootstrap procedure for generating the critical values with the asymptotically correct nominal level.

Recall the definition of the estimator $T_n^{(k)}$ of the change point in the correlation structure.
in (4.4). Consider the statistics

\[ \hat{\Delta}_{n,1}^{(k)} = \frac{1}{[nt_n^{(k)}]} \sum_{j=1}^{[nt_n^{(k)}]} \frac{\hat{e}_j \hat{e}_{j+k}}{\hat{\sigma}_n^2(t_j)}, \quad \hat{\Delta}_{n,2}^{(k)} = \frac{1}{n-[nt_n^{(k)}]} \sum_{j=[nt_n^{(k)}]+1}^{n} \frac{\hat{e}_j \hat{e}_{j+k}}{\hat{\sigma}_n^2(t_j)} \]

and take

\[ \hat{\Delta}_n^{(k)} = \hat{\Delta}_{n,2}^{(k)} - \hat{\Delta}_{n,1}^{(k)} \]  \hspace{1cm} (4.11)

as an estimator of the difference \( \Delta_k = \rho_2^{(k)} - \rho_1^{(k)} \). We have consistency of \( \hat{\Delta}_n^{(k)} \). (The proof is deferred to the online supplement.)

**Lemma 4.2.** If the conditions of Theorem 4.1 hold, then

\[ \hat{\Delta}_n^{(k)} - \Delta_k = O_P\left(\frac{\log n}{\sqrt{n}}\right) \]

for \( k = r_1, ..., r_l \).

Let

\[ \hat{A}_j^{(k)} = \frac{\hat{e}_j \hat{e}_{j+k}}{\hat{\sigma}_n^2(t_j)} - \hat{\Delta}_n^{(k)} 1(j \geq [nt_n^{(k)}]), \]  \hspace{1cm} (4.12)

and let \( \{R_j\}_{j \in \mathbb{Z}} \) be a sequence of i.i.d. standard normal distributed random variables independent of \( \{F_i\}_{i \in \mathbb{Z}} \). We introduce the partial sums \( \hat{S}_{j,m}^{A,(k)} = \sum_{r=j}^{j+m-1} \hat{A}_r^{(k)}, \hat{S}_n^{A,(k)} = \sum_{r=1}^{n} \hat{A}_r^{(k)} \)

and define \( \hat{S}_{j,m}^A = (\hat{S}_{j,m}^{A,(r_1)}, ..., \hat{S}_{j,m}^{A,(r_l)}), \hat{S}_n^A = (\hat{S}_{n}^{A,(r_1)}, ..., \hat{S}_{n}^{A,(r_l)}) \),

\[ \hat{\Phi}_{i,m}^A = \frac{1}{\sqrt{m(n-m+1)}} \sum_{j=1}^{i} \left( \hat{S}_{j,m}^A - \frac{m}{n} \hat{S}_n^A \right) R_j. \]  \hspace{1cm} (4.13)

Let \( \hat{\Phi}_{i,m}^{A,(s)} \) be the \( s \)th component of \( \hat{\Phi}_{i,m}^A \). Then the following result is proved in Section A.3 of the online supplement.

**Theorem 4.2.** Suppose the conditions of Theorem 4.1 hold and that \( m \to \infty, m \log n/\sqrt{n} \to \)
0, $\sqrt{m}(c_n^2 + \frac{1}{\sqrt{nc_n}} + b_n^2 + \frac{1}{\sqrt{mb_n}})c_n^{-1/4}) \log n \to 0$. If for $1 \leq s \leq l$,

$$M_n^{r,(rs)} = \frac{1}{n} \frac{6}{(\hat{t}_n^{(rs)})^2 (1 - \hat{t}_n^{(rs)})^2} \sum_{m+1 \leq i \leq n-m+1} \left( \hat{\Phi}_{i,m,r}^{A,(s)} - \frac{i}{n-m+1} \hat{\Phi}_{n-m+1,m,r}^{A,(s)} \right) \left( \frac{i \hat{t}_n^{(rs)}}{n} - \frac{i}{n} \wedge \hat{t}_n^{(rs)} \right)$$

then (conditional on $\mathcal{F}_n$ in probability)

$$(M_n^{r,(r_1)}, ..., M_n^{r,(r_l)}) \overset{D}{\to} \hat{Z} := \{ Z^{(rs)} \}_{s=1}^l$$

(4.14)

where the random variables $\{ Z^{(rs)} \}_{s=1}^l$ are defined in Theorem 4.1.

The bootstrap test for the hypothesis (1.3) of a non-relevant change in the correlation structure results as follows.

**Algorithm 4.1.**

1. Calculate the statistics $\hat{T}_{n}^{(r_u),r}$ defined in (4.7) for $u = 1, \ldots, l$. For given $\delta = (\delta_{r_1}, ..., \delta_{r_l})^T$, calculate $\hat{T}_{n,max}$ by (4.10).

2. Generate $B$ conditionally i.i.d copies $\{ \hat{\Phi}_{i,m,r}^{A} \}_{i=1}^{n-m+1}$ $(r = 1, 2, ..., B)$ of the sequence $\{ \hat{\Phi}_{i,m}^{A} \}_{i=1}^{n-m+1}$ defined in (4.13). Calculate $M_n^{A,r} := \max_{1 \leq u \leq l}(M_n^{A,(r_u)})$ where, for $1 \leq u \leq l$,

$$M_n^{A,(r_u)} = \frac{1}{n} \frac{6 \text{sign}(\hat{\Delta}_n^{(r_u)})}{(\hat{t}_n^{(r_u)})^2 (1 - \hat{t}_n^{(r_u)})^2} \sum_{i=m+1}^{n-m+1} \left( \hat{\Phi}_{i,m,r}^{A,(u)} - \frac{i}{n-m+1} \hat{\Phi}_{n-m+1,m,r}^{A,(u)} \right) \left( \frac{i \hat{t}_n^{(r_u)}}{n} - \frac{i}{n} \wedge \hat{t}_n^{(r_u)} \right).$$

3. If $M_{(1)}^A \leq M_{(2)}^A \leq \ldots \leq M_{(B)}^A$ denote the order statistics of $M_1^A, \ldots, M_B^A$, reject the null hypothesis (1.3) of a non-relevant change in the correlations at level $\alpha$ if

$$\hat{T}_{n,max} > \frac{M_{(B(1-\alpha))}}{\sqrt{n}}.$$

(4.15)

The $p$-value of this test is given by $1 - B^*$, where $B^* = \max\{ r : \frac{M_r^A}{\sqrt{n}} \leq \hat{T}_{n,max} \}$.

If only one lag is considered, then the term $\text{sign}(\hat{\Delta}_n^{(r_u)})$ in the definition of $M_n^{A,(r_u)}$ can be dropped by the symmetry of a centered Gaussian process.
Remark 4.1. We investigate the power of the test \((4.15)\). Let \(\bar{v}_{r_s,1-\alpha}\) be the \((1-\alpha)\)-quantile of the distribution of the random variable \(\max\{Z^{(r_s)}\text{sign}(\Delta_{r_s})\}_{s=1}^{l}\). If \(\Delta_{r_1}^2 > \delta_{r_1}^2\), then we obtain from Theorem 4.1 an approximation for the power of the test \((4.10)\) as

\[
\beta_n(\delta, \Delta) := \mathbb{P}\left(\hat{T}_{n, \max} > \bar{v}_{1-\alpha} \sqrt{\frac{\Delta_{r_1}}{n}}\right) \geq \mathbb{P}\left(\hat{T}_{n, (r_1)} > \frac{\Delta_{r_1}^2 - \delta_{r_1}^2}{\sqrt{\frac{\Delta_{r_1}}{n}}} + \frac{\bar{v}_{1-\alpha} \delta_{r_1}}{\sqrt{\frac{\Delta_{r_1}}{n}}}\right)
\]

where \(\Psi_{r_1}\) is the distribution function of the random variable \(Z^{(r_1)}\) (in fact a centered normal distribution). Therefore, under the alternative of a relevant change for some lag \(r_1\), \(\Delta_{r_1}^2 > \delta_{r_1}^2\), we have \(\beta_n(\delta, \Delta) \to 1\) as \(n \to \infty\), which provides the consistency of the test \((4.15)\). Under the null hypothesis \(0 < \Delta_{r_s}^2 \leq \delta_{r_s}^2\) for \(1 \leq s \leq l\), we have

\[
1 - \beta_n(\delta, \Delta) = \mathbb{P}\left(\hat{T}_{n, \max} \leq \frac{\bar{v}_{1-\alpha}}{\sqrt{\frac{\Delta_{r_1}}{n}}}\right) \geq \mathbb{P}\left(\max_{1 \leq s \leq l} \left\{\frac{\Delta_{r_s} Z^{(r_s)}}{\delta_{r_s}} + \sqrt{\frac{\Delta_{r_s}^2 - \delta_{r_s}^2}{\delta_{r_s}}}\right\} \leq \frac{\bar{v}_{1-\alpha}}{\sqrt{\frac{\Delta_{r_1}}{n}}}\right) + o(1)
\]

Consequently, if \(0 < \Delta_{r_s}^2 \leq \delta_{r_s}^2\) \((1 \leq s \leq l)\) and

\[
i f l^* := \#\{s \in \{1, \ldots, l\} \mid |\Delta_{r_s}| = \delta_{r_s}\}
\]

denotes the number of coordinates where the “true” difference between the lag-\(r_s\) correlations is at the boundary of the null hypothesis, we have

\[
\lim_{n \to \infty} \beta_n(\delta, \Delta) = \begin{cases} 
0 & \text{if } l^* = 0 \\
\alpha & \text{if } l^* = l \\
< \alpha & \text{if } 1 \leq l^* \leq l - 1.
\end{cases}
\]

If there exist some lags, without loss of generality \(r_1, \ldots, r_k\), with \(\Delta_{r_i} = 0\) \((1 \leq i \leq k)\) and
\( k < l \), then it follows that \( \hat{T}_n^{(r_1),r} = O_P(1/n) \) for all \( 1 \leq i \leq k \), and it is easy to see that a result similar to (4.19) holds. Moreover, if \( \Delta_{r_s} = 0 \) for all \( s = 1, \ldots, l \), then \( \hat{T}_n^{(r_s),r} = O_P(1/n) \) for \( s = 1, \ldots, l \) and \( \lim_{n \to \infty} \beta_n(\delta, \Delta) = 0 \) (since \( \alpha \leq 0.5 \) and \( \bar{v}_{1-\alpha} > 0 \)). Summarizing these calculations shows that the test (4.15) has, in fact, asymptotic level \( \alpha \).

We can also use (4.16) to investigate the power as a function of the parameter \( \delta \) in the hypothesis (1.3): for sufficiently large \( n \) the power \( \beta_n(\delta, \Delta) \) is approximately 1 if \( \delta \to 0 \) and \( \sqrt{n} \delta \to \infty \), and 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 (4.15).

**Remark 4.2.** In applications of the test (4.15) for a non-relevant change in the correlation, the thresholds \( \Delta_{r_s} \) are usually very small, and this can lead to a less accurate approximation of the nominal level. Consider, for example, the univariate test for a relevant change in the lag-1 correlation. We obtain from the proof of in Theorem 4.1 for the estimating object of statistic defined in (4.7) the stochastic expansion (omitting the subscript)

\[
\sqrt{n}(T_n^2 - \Delta^2) = \frac{6\Delta}{t^2(1-t)^2} \int (U(s) - sU(1))(st - s \wedge t)ds + \frac{3}{\sqrt{n}t^2(1-t)^2} \int (U(s) - sU(1))^2 ds + o_p(n^{-1/2}),
\]

(4.20)

where \( t \) is the jump time in lag-1 correlation and the process \( \{U(t)\}_{t \in [0,1]} \) is defined in Theorem 3.1. The second term vanishes asymptotically. However, when \( \Delta \) is small and the sample size is not too large, the first and second term on the right hand side of (4.20) could be comparable in size. The bootstrap methodology proposed in this paper provides us with a convenient way to solve this problem. We propose to replace \( \hat{T}_{n,\text{max}} \) in (4.10) by \( \max_{1 \leq u \leq l}(\hat{T}_n^{(r_u),r} - \delta^2_{r_u}) \), and to replace the statistic \( M_{n,r}^{A,(r_u)} \) in step [2] of Algorithm 4.1 by the statistic

\[
M_{n,r}^{A,(r_u)} = \frac{1}{n} \frac{6\text{sign}(\hat{\Delta}_{n}^{(r_u)})\delta_{r_u}}{(\hat{t}_n^{(r_u)})^2(1-\hat{t}_n^{(r_u)})^2} \sum_{i=m+1}^{n-m+1} \left( \Phi_{i,m,r}^{A,(u)} - \frac{i}{n-m+1} \Phi_{n-m+1,m,r}^{A,(u)} \right) \left( \frac{\hat{t}_n^{(r_u)}}{n} - \frac{i}{n} \wedge \hat{t}_n^{(r_u)} \right) + N_{n,r_u}^A
\]

22
where
\[
N^{A,(r_u)}_{n,r} = \frac{1}{n^{3/2}} \frac{3}{(i_n^{(r_u)})^2(1-i_n^{(r_u)})^2} \sum_{i=m+1}^{n-m+1} \left( \hat{\Phi}^{A,(u)}_{i,m,r} - \frac{i}{n-m+1} \hat{\Phi}^{A,(u)}_{n-m+1,m,r} \right)^2.
\]

**Remark 4.3.** Straightforward calculation shows that the computational time complexity of Algorithms 3.1 and 4.1 is \( O(Bn + \alpha(n)) \), where \( n \) is the length of time series, \( \alpha(n) \) is the time cost of obtaining \( \{\tilde{e}_i\}_{1 \leq i \leq n} \) and \( \{\tilde{\sigma}^2(t_i)\}_{1 \leq i \leq n} \) which depends on the particular optimization method that users choose, and \( B \) is the number of bootstrap replications that is mainly determined by the nominal level. As a rule of thumb, for a nominal level of 5\%, our experience shows that \( B = 2000 \) is sufficient, though we use \( B = 4000 \) and 8000 in our simulations and data analysis, respectively.

**Remark 4.4.** For the change point test defined in Algorithm 3.1 the alternative hypothesis allows for multiple change points and one could use a similar approach as in Section 5 of Qu (2008). For the test of relevant change points defined in Algorithm 4.1 we propose to proceed in two steps: we use Algorithm 3.1 and the binary segmentation technique to deal with multiple change points (see Vostrikova (1981)); if this procedure identifies the potential relevant change points \( 0 = t_0 < t_1 < \ldots < t_s < t_{s+1} = 1 \), we perform a test for a relevant change point in every two consecutive intervals \( (t_l, t_{l+2}] \) for \( 0 \leq l \leq s - 1 \).

**Remark 4.5.** The behaviour of the test statistics may not be close to the limiting distribution when the sequence is short, especially under piecewise local stationarity. As a result, the finite performance of those tests only based on the limiting distribution may not be satisfactory under non-stationarity. Thanks to the bootstrap procedure, our proposed method works reasonably well and is not very sensitive to the length of the sequence. This is also justified by the simulation results for sample sizes 300, 500, 800 in Section C.3 of the supplementary material. As a rule of thumb, we recommend our method when the length of sequence is larger than 300.
Figure 1: *Typical sample paths of the processes corresponding to model (I) - (IV).*

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 quadratic mean function $\mu(t) = 8(-t - 0.5)^2 + 0.25$ and a sequence of independent identically random variables $\{\varepsilon_j\}_{j \in \mathbb{Z}}$ in the definition of the errors $e_i = G_j(t_i, F_i)$ in model (2.1) where $F_i = \sigma(\ldots, \varepsilon_0, \ldots, \varepsilon_i)$, if not mentioned otherwise. The dependence structures differ by choice of the nonlinear filters $G_j$. The sample size was $n = 500$ and all results were based on 4000 simulation runs. In each run, the critical values were generated by $B = 4000$ bootstrap replications. We use the Epanechnikov kernel; We analyzed the impact of different kernel functions on the performance of the tests and saw no substantial differences. Some of these investigations are summarized in Section C.2 of the supplementary material.

5.1 Change point tests for correlations

We investigate properties of the tests for changes in the lag-1 and lag-2 correlations. For this purpose we consider these models.
(I) $G(t, F_i) = H(t, F_i) \sqrt{1 - (t - 0.5)^2}/2$, where $H(t, F_i) = 0.2H(t, F_{i-1}) + \varepsilon_i$.

(II,IIIt) $G(t, F_i) = H(t, F_i) \sqrt{c(t)}/2$ for $t \leq 0.5$, and $G(t, F_i) = H(t, F_i) \sqrt{d(t)}/2$ for $t > 0.5$, where $c(t) = 1 - (t - 0.5)^2$, $d(t) = 1 - \frac{1}{2} \sin t$, and $H(t, F_i) = 0.2H(t, F_{i-1}) + \varepsilon_i$.

(III,IIIt) $G(t, F_i) = H(t, F_i) \sqrt{1 - (t - 0.5)^2}/2$, where $H(t, F_i) = 0.2H(t, F_{i-1}) + \varepsilon_i$ for $t \leq 0.5$, and $H(t, F_i) = 0.4H(t, F_{i-1}) + \varepsilon_i$ for $t > 0.5$.

(IV) $G(t, F_i) = H(t, F_i) \sqrt{1 - (t - 0.5)^2}/2$, where $H(t, F_i) = 0.5H(t, F_{i-1}) + 0.1H(t, F_{i-2}) + \varepsilon_i$ for $t \leq 0.5$, and $H(t, F_i) = 0.3H(t, F_{i-1}) + 0.2H(t, F_{i-2}) + \varepsilon_i$ for $t > 0.5$.

For models (I) (II) (III) (IV) the innovations were $\varepsilon_i \sim_{i.i.d} N(0,1)$, and for model (IIIt) and (IIIIt) $\varepsilon_i \sim_{i.i.d} t(5)/\sqrt{5/3}$. Model (I) was for locally stationary processes. The variance of the process was time-varying, but the correlation remained constant. Model (II,IIIt), (III,IIIIt) and model (IV) were piecewise locally stationary processes, where the variances had an abrupt change. Before and after the jump, the variance varied smoothly. The correlations of model (I) and (II,IIIt) were constant, while the correlations of model (III,IIIIt,IV) had a break at $t = 0.5$ and were used to illustrate the approximation of the optimal level of the test for the hypothesis of a non-relevant change point, as discussed at the end of this section. Model (IV) is a tvAR(2) model with a change in the lag-1 and lag-2 scaled AR coefficients. Typical trajectories corresponding to these processes are depicted in Figure 1.

Change point analysis on the basis of the tests proposed in Section 3 and 4 requires the choices of two bandwidths in the local linear estimates of the mean and variance. We used a generalized cross validation method (GCV) introduced by Zhou and Wu (2010) to select the bandwidth for estimating the mean function. Then we applied this cross validation procedure again to select the bandwidth for estimating the variance function. The parameters $L$ and $\zeta$ in the estimator (3.5) were chosen as $L = \lfloor 3n^{1/3} \rfloor$ and $\zeta = 0.2$, respectively. For the choice of window size $m$ in Section 3 and 4 we used the minimal volatility method (MV) in Zhou (2013).
For the nominal level we display in Table 1 the rejection probabilities of the test for the hypothesis (3.1) of a “classical” change point, where various bandwidths $b_n$ from the interval $[0.075, 0.225]$ were considered. At each fixed $b_n$, the bandwidth $c_n$ for estimating the variance was calculated by cross validation. The last row of the table shows the simulated rejection probabilities for the case that both bandwidths $b_n$ and $c_n$ were calculated by cross validation. In the 1st-3rd column we display results of the test (3.15) for models I, II and IIIt, where we used lag-1 correlation. The 4th column (denoted by II$^\star$) corresponds to model II, where lag-1 and lag-2 correlations were used simultaneously in the test (3.15). We observed a reasonable approximation of the nominal level, only slightly affected by the choice of the bandwidth $b_n$. Moreover, generalized cross validation yielded a good approximation of the nominal level in all cases under consideration.

In Table 2 we show corresponding results for the test (4.15) of a non-relevant change point, where in all cases the simulated type I error was calculated for a boundary point of the null hypothesis. Thus $l^* = l$ in (4.19) and, by the discussion in Remark 4.1 the nominal level of the test should be close to $\alpha$ at this point. In the 1st and 2nd columns we show the simulated type I error of the test (4.15) for a relevant change in lag-1 correlation with $\delta = \Delta = 0.3$ for Models III and IIIt, respectively. In the 3rd column of Table 2 we display the simulated level of the test for the hypotheses (4.1) for a relevant change in lag-1 and lag-2 correlations for model III, where $\delta_1 = \Delta_1 = 0.3$ and $\delta_2 = \Delta_2 = 0.15$, respectively. Finally, the 4th column shows corresponding results for the locally stationary AR(2) model (IV) where again lag-1 and lag-2 correlations were considered (here $(\delta_1, \delta_2) = (\Delta_1, \Delta_2) = (0.18, 0.065)$). Once again, all displayed results correspond to the boundary, and at interior points of the null hypothesis the type I error of the test (4.15) is usually smaller (see the discussion in Remark 4.1).

Figure 2 shows the simulated rejection probabilities of the tests for the hypothesis (4.1) of a non-relevant change in the lag-1 correlation for model III as a function of the parameter
Table 1: Simulated Type I error of the test for the classical hypothesis \((3.1)\) of a change in the correlation for various bandwidths and the bandwidth calculated by generalized cross validation (last line). Columns 1 – 3: test \((3.15)\) based on the lag-1 correlation for Models I, II and II*. Column 4: test \((3.15)\) based on lag-1 and lag-2 correlations for Model II.

| model | I (5%) | I (10%) | II (5%) | II (10%) | IIt (5%) | IIt (10%) | II* (5%) | II* (10%) |
|-------|--------|---------|---------|----------|---------|---------|---------|---------|
| \(b_n/\alpha\) | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% |
| 0.075 | 5.625 | 11.6 | 4.375 | 9.6 | 4.825 | 10.025 | 4.725 | 10.35 |
| 0.1 | 5.2 | 10.8 | 4.3 | 9.775 | 4.9 | 10.925 | 5.325 | 10.675 |
| 0.125 | 4.025 | 9.35 | 4.05 | 9.275 | 4.075 | 8.875 | 4.425 | 8.95 |
| 0.15 | 4.575 | 10.075 | 3.75 | 8.4 | 4.35 | 9.75 | 3.95 | 9.45 |
| 0.175 | 4.1 | 8.675 | 3.85 | 8.75 | 3.65 | 8.4 | 4.175 | 9.05 |
| 0.2 | 3.725 | 8.6 | 3.575 | 8.15 | 3.525 | 8.225 | 3.825 | 8.35 |
| 0.225 | 3.925 | 8.675 | 3.2 | 8.025 | 4.025 | 8.65 | 3.95 | 8.625 |
| GCV | 4.275 | 9.625 | 4.575 | 9.425 | 4.25 | 9.475 | 4.75 | 9.8 |

Table 2: Simulated Type I error of the test for the hypothesis \((4.1)\) of a relevant change in the correlation for various bandwidths and the bandwidth calculated by generalized cross validation (last line). Columns 1 and 2: tests based on the lag-1 correlation for Models III and III*. Column 3 and 4: test based on lag-1 and lag-2 correlations for Models III and IV.

| model | III (5%) | III (10%) | III* (5%) | III* (10%) | IIIIt (5%) | IIIIt (10%) | III (5%) | III (10%) | IV (5%) | IV (10%) |
|-------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| \(b_n/\alpha\) | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% | 5% | 10% |
| 0.075 | 5.275 | 9.575 | 6.65 | 10.775 | 4.925 | 9.9 | 5.6 | 10.85 |
| 0.1 | 6.2 | 10.825 | 6.325 | 10.575 | 5.45 | 10.475 | 5 | 9.825 |
| 0.125 | 6.05 | 11.3 | 6.425 | 11.125 | 5 | 10.025 | 4.875 | 9.55 |
| 0.15 | 5.8 | 10.25 | 6.575 | 11.075 | 5.25 | 10.35 | 4.025 | 8.325 |
| 0.175 | 5.775 | 10.1 | 6.075 | 10.8 | 4.85 | 10.275 | 4.175 | 8.5 |
| 0.2 | 5.775 | 9.425 | 5.575 | 9.9 | 4.775 | 10.1 | 4 | 8.925 |
| 0.225 | 5.3 | 10.15 | 5.45 | 9.95 | 4.525 | 9.425 | 3.75 | 8.2 |
| GCV | 5.45 | 9.9 | 5.875 | 10.55 | 5.45 | 10.675 | 4.875 | 9.6 |
$\delta \in [0, 2\Delta]$. The significance level was chosen as 0.1. As expected the probability of rejection decreases with $\delta$ (see also the discussion in Remark 4.1). More simulation results for different sample sizes can be found in Section C.3 of the supplementary material.

![Figure 2: Simulated rejection probabilities of the test for lag-1 correlation as a function of the threshold $\delta \in [0, 2\Delta]$ in the hypothesis (1.3) for Model III.](image)

Importantly, the symmetry of the innovations do not affect the asymptotic properties of the tests, since the rates of Gaussian approximations of partial sums from skewed random variables are of the same order as in the symmetric case. To investigate if there exist differences in the finite sample properties we took model II and III, with the $i.i.d.$ Gaussian innovations replaced by $i.i.d.$ $(\chi^2(5) - 5)/\sqrt{10}$ random variables. In Table 3 we display the simulated type I error of the test (3.15) for a change point in the lag-1 correlation in model II and of the test (4.15) for a relevant change point in the lag-1 correlation in model III. The corresponding results for a symmetric error can be found in Tables 1 and 2; we only observe minor differences in the approximation of the nominal level between the symmetric and non-symmetric case.

**Remark 5.1.** Throughout, the bandwidth $b_n$ is assumed to be the same over the whole sequence. As pointed out by one referee, it might be of interest to investigate a time-dependent bandwidth $b_n$ with respect to its potential to deal with local stationarity. Using similar ar-
Table 3: Simulated type I error of the test (3.15) for a change point in the lag-1 correlation in model II and of the test (4.15) for a relevant change point test in lag-1 correlation (at the boundary point of the null) in model III with $(\chi^2(5) - 5)/\sqrt{10}$ innovations. The last column represents the simulated Type I error if the bandwidth is $b_n$ selected by GCV.

| model | $b_n$  | 0.075 | 0.1  | 0.125 | 0.15  | 0.175 | 0.2  | 0.225 | GCV |
|-------|-------|-------|------|-------|-------|-------|------|-------|-----|
| II    | 5%    | 4.6   | 4.55 | 3.6   | 2.6   | 2.95  | 3.25 | 2.9   | 4.2 |
|       | 10%   | 10.35 | 8.95 | 8.7   | 7.15  | 7.65  | 7.35 | 6.8   | 9.3 |
| III   | 5%    | 4.85  | 5.1  | 4.95  | 6.9   | 5.25  | 6    | 5.35  | 5.4 |
|       | 10%   | 8.9   | 8.85 | 9.7   | 11.7  | 9.9   | 10.35| 9.85  | 9.95|

Arguments as in Zhou and Wu (2009) we can obtain the optimal time varying bandwidth as

$$b_n(t) = \left| \frac{\hat{\kappa}(t)}{\hat{\sigma}(t)} \right| b_n,$$

(5.1)

where $b_n$ is the time invariant bandwidth obtained by the GCV method, $\hat{\kappa}^2$ and $\hat{\sigma}^2$ are estimates of the long run variance and the variance of the random variables $e_t$, respectively. It is hard to accurately estimate $\kappa^2$ in a PLS model due to the unknown break points. In the case of local stationarity, an estimate of $\kappa^2$ was proposed by Zhou and Wu (2010) and we used this method to investigate the differences between a local and global bandwidth in the locally stationary model I. The simulated levels of the corresponding bootstrap tests are shown in Table 4 and we observe that the performance of the procedure with a time dependent bandwidth is quite similar to the one using a constant bandwidth.

Table 4: Simulated type I error of the test (3.15) for a change point in the lag-1 correlation in model I using the time varying bandwidth (5.1). The last column represents the simulated Type I error if the bandwidth is $b_n$ selected by GCV.

| $b_n$ | 0.075 | 0.1  | 0.125 | 0.15  | 0.175 | 0.2  | 0.225 | GCV |
|-------|-------|------|-------|-------|-------|------|-------|-----|
| Model I | 5%    | 5.05 | 4.9  | 4.3   | 3.95  | 4.2  | 3.15  | 3.55 | 4.15|
|        | 10%   | 10.45| 9.75 | 9.5   | 9.85  | 8.9  | 7.55  | 8    | 10.7|
5.2 Some robustness considerations

As was pointed out by a referee it might be of interest to investigate the approximation of the nominal level if the assumption of PLS is violated. For this purpose we considered modifications of the models II and III introduced in the previous section. Let \( \{\eta_i, i \in \mathbb{Z}\} \) denote i.i.d. standard normal distributed and \( \{\varepsilon_i, i \in \mathbb{Z}\} \) denote i.i.d. \( t \)-distributed random variables with 5 degrees of freedom, normalized such that they have variance 1. We consider the processes

\[
(II_0) \ G_i = H_i \sqrt{c(i/n)} / 2 \text{ for } i/n \leq 0.5, \quad \text{and} \quad G_i = H_i \sqrt{d(i/n)} / 2 \text{ for } i/n > 0.5,
\]

where \( c(t) = 1 - (t - 0.5)^2, \quad d(t) = 1 - \frac{1}{2} \sin t \) and \( H_i = 0.2H_{i-1} + \varepsilon_i \) for \( i/n \leq 0.5 \), and \( H_i = 0.2H_{i-1} + \eta_i \) for \( i/n > 0.5 \).

\[
(III_0) \ G_i = H_i \sqrt{1 - (i/n - 0.5)^2} / 2, \text{ where} \quad H_i = 0.1H_{i-1} + \varepsilon_i \text{ for } i/n \leq 0.5, \quad \text{and} \quad H_i = 0.4H_{i-1} + \eta_i \text{ for } i/n > 0.5.
\]

These models are not PLS in the sense of Definition 2.1. In Table 5 we show the simulated type I error of the test (3.15) for a change point in the lag-1 correlation in model II_0 and of the test (4.15) for a relevant change point test in the lag-1 correlation in model III_0. We observe reasonable approximations of the nominal level in all cases under consideration. In Table 5: Simulated type I error of the test (3.15) for a change point in the lag-1 correlation in model II_0 and of the test (4.15) for a relevant change point test in the lag-1 correlation (at the boundary point of the null) in model III_0. The last column represents the simulated Type I error if the bandwidth is \( b_n \) selected by GCV.

|     | \( b_n \) | 0.075 | 0.1  | 0.125 | 0.15 | 0.175 | 0.2  | 0.225 | GCV  |
|-----|-----------|-------|------|-------|------|-------|------|-------|------|
| \( II_0 \) | 5% | 4.45  | 3.9  | 3.8   | 4.15 | 2.9   | 3    | 2.85  | 3.6  |
|     | 10% | 9.8   | 9.2  | 9.2   | 8.4  | 7     | 7.55 | 6.7   | 8.55 |
| \( III_0 \) | 5% | 6.25  | 5.75 | 5.9   | 5.2  | 6.2   | 6.05 | 4.7   | 4.6  |
|     | 10% | 10.15 | 10   | 10.1  | 9.7  | 11.4  | 10.8 | 8.8   | 8.35 |

fact, it follows from Zhou (2012) that model \( II_0 \) and \( III_0 \) can be approximated by the two PLS models.
(II0).  \( G(t, F_i) = H(t, F_i)\sqrt{c(t)/2} \) for \( t \leq 0.5 \), and \( G(t, F_i) = H(t, F_i)\sqrt{d(t)/2} \) for \( t > 0.5 \),
where \( c(t) = 1 - (t - 0.5)^2 \), \( d(t) = 1 - \frac{1}{2}\sin t \) and \( H(t, F_i) = 0.2H(t, F_{i-1}) + \varepsilon_i \) for \( t \leq 0.5 \), and \( H(t, F_i) = 0.2H(t, F_{i-1}) + \eta_i \) for \( t > 0.5 \).

(III0).  \( G(t, F_i) = H(t, F_i)\sqrt{1 - (t - 0.5)^2/2} \), where \( H(t, F_i) = 0.1H(t, F_{i-1}) + \varepsilon_i \) for \( t \leq 0.5 \), and \( H(t, F_i) = 0.4H(t, F_{i-1}) + \eta_i \) for \( t > 0.5 \),
where \( F_i = (\eta_{-\infty}, \varepsilon_{-\infty}, \ldots, \eta_0, \varepsilon_0, \ldots, \eta_i, \varepsilon_i) \). In summary, the proposed test procedures work reasonably well as long as the underlying processes are not too different from PLS processes. An important class of non-stationary processes that are not PLS and cannot be handled by our methodology are the unit root non-stationary processes.

Figure 3: Simulated power. Upper left panel: test for a constant lag-1 correlation defined in (3.15) (model (I')). Upper right panel: test for constant lags-1 and lag-2 correlation defined in (3.15) (model (I')). Lower left panel: test for the hypothesis of a non-relevant change in the lag-1 correlation defined in (4.15) (model (II')). Lower right panel: test for the hypothesis of a non-relevant change in the lag-1 and lag-2 correlation defined in (4.15) (model (II')).
5.3 Power properties

In this section we investigate the power of the proposed tests in two scenarios. Let \( \{\varepsilon_i\}_{i \in \mathbb{Z}} \) be i.i.d \( N(0,1) \).

(I') \( G(t, \mathcal{F}_i) = H(t, \mathcal{F}_i)\sqrt{c(t)/2} \) for \( t \leq 0.5 \), and \( G(t, \mathcal{F}_i) = H_1(t, \mathcal{F}_i)\sqrt{c(t)/2} \) for \( t > 0.5 \), where \( c(t) = 1 - (t - 0.5)^2 \), \( H(t, \mathcal{F}_i) = 0.2H(t, \mathcal{F}_{i-1}) + \varepsilon_i \) for \( t \leq 0.5 \), and \( H(t, \mathcal{F}_i) = (0.2 - \lambda)H(t, \mathcal{F}_{i-1}) + \varepsilon_i \) for \( t > 0.5 \).

(II') \( G(t, \mathcal{F}_i) = H(t, \mathcal{F}_i)\sqrt{1-(t-0.5)^2}/2 \), where \( H(t, \mathcal{F}_i) = (0.1 - \lambda)H(t, \mathcal{F}_{i-1}) + \varepsilon_i \) for \( t \leq 0.5 \), and \( H(t, \mathcal{F}_i) = 0.4H(t, \mathcal{F}_{i-1}) + \varepsilon_i \) for \( t > 0.5 \).

Model (I') is used to study the power of the test (3.15) for the “classical” hypothesis of no change point in the correlation for various values of \( \lambda \), where \( \lambda = 0 \) corresponds to the null hypothesis of a constant correlation. In the upper panel of Figure 3 we show the simulated power of the test for a constant lag-1 correlation while in the upper right panel corresponding results of the test for a constant lag-1 and lag-2 correlations are displayed. We observe a decrease in power, which can be explained by the observation, that in model (I') the jump size of the lag 2-correlation is \( |0.2^2 - (0.2 - \lambda)^2| \), which is not monotone with respect to \( \lambda \). The power properties of the test (4.15) of a change is investigated in model (II'). In the lower left panel of Figure 3 we display the simulated rejection probabilities for the hypotheses of a non-relevant change in the lag-1 correlation, that is

\[ H_0 : \Delta_1 \leq 0.3 \text{ versus } H_1 : \Delta_1 > 0.3, \]

where \(-0.6 \leq \lambda \leq 0\) corresponds to the null hypothesis. In the lower right panel we investigate the hypotheses for lag-1 and lag-2 correlations, that is

\[ H_0 : \Delta_1 \leq 0.3 \text{ and } \Delta_2 \leq 0.15 \text{ versus } H_1 : \Delta_1 > 0.3 \text{ or } \Delta_2 > 0.15, \]

where \(0.1 - \sqrt{0.31} \leq \lambda \leq 0\) corresponds to the null hypothesis. We observe a decrease in
power (note again that, the jump size of the lag 2-correlation is $|0.4^2 - (0.1 - \lambda)^2|$, which is not monotone with respect to $\lambda$). We conclude that in all cases under consideration the proposed methodology can detect (relevant) changes in the correlation structures with reasonable size.

**Remark 5.2.** The power of the proposed tests depends sensitively on the choice of the bandwidth $b_n$. Ideally, if the errors are i.i.d or the series is strictly stationary, the optimal bandwidth can be calculated by an Edgeworth-expansion-based method (see Gao and Gijbels (2008)) such that the power is optimized. However, the extension of this approach to a PLS scenario is non-trivial, and is out of the present scope but an interesting problem for future work. In the case of a stationary null hypothesis, we have also compared the power of our test presented with algorithms specifically designed for stationary processes. We observed that our approach has decent power; these results are presented in Section C.1 of the supplementary material.

### 6 Data Analysis

Table 6: Tests for the existence of a change point in the lag-1 and lag-2 correlations of the USD/CAD series, respectively. $v_0^*$ 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 date.

|                | lag 1-Correlation | lag 2-Correlation |
|----------------|-------------------|-------------------|
|                | Whole  | Before | After | Whole  | Before | After |
| Test Stat.     | 1.28*  | 0.50   | 0.99   | 1.38** | 0.40   | 2.74  |
| $v_{90\%}$     | 1.23   | 0.70   | 1.01   | 1.15   | 0.72   | 3.95  |
| $v_{95\%}$     | 1.36   | 0.78   | 1.12   | 1.3    | 0.80   | 4.70  |
| $b_n$          | 0.34   | 0.18   | 0.54   | 0.34   | 0.29   | 0.15  |
| $m$            | 18     | 9      | 22     | 18     | 18     | 12    |
| $c_n$          | 0.13   | 0.24   | 0.19   | 0.13   | 0.14   | 0.34  |

We analyze the daily exchange rate of U.S. dollar/Canadian dollar from Nov 18th, 2011 to Jun 24th, 2016. The data can be obtained from https://www.federalreserve.gov/releases
Figure 4: Percentage change (left panel) and the squared percentage change (right panel) of exchange rate of USD/CAD. The line in the right panel is the fitted mean for the squared percentage change.
Figure 5: *p*-values of the bootstrap test for a relevant change in the lag-1, lag-2 and lag-(1, 2) correlations for the squared percentage change of USD/CAD for different values of the threshold $\delta$. The horizontal line marks the significance level 0.05.
The series contains 1154 data points. During the period, USD/CAD has changed drastically in the range (0.9710, 1.4592). The wide range of the exchange rate motivates us to further investigate the robustness of the volatility of the percentage change of the series during the period. Figure 4 shows the percentage change and squared percentage change of the exchange rate data. The pattern of the squared percentage change of exchange rate displays non-stationarity. For the test over the whole period, the GCV method selects \( b_n = 0.34 \) and \( c_n = 0.3 \) and the MV method select \( m = 18 \). For this section, the critical values were generated by 8000 bootstrap replications. We used the statistic (3.5) to estimate the abrupt change points in the variance with \( \zeta = 0.10 \) and \( L = 31 \), and identified a variance change point \( t_n^* = 795 \), which corresponds to Jan 15th, 2015.

Let \( X_t \) represent the squared percentage change at day \( t \), and consider the relationship between \( X_t \) and \( X_{t-i} \), \( i = 1, 2, 3 \). We performed our test on lag 1, 2 and 3 simultaneously to check two null hypothesis: (i) all three correlations are 0, (ii) all three correlations stay constant during the time considered. For (ii), we use the testing procedure in Section 3. For testing (i), we modified the test procedure in Section 3 by setting \( T_n \) in (3.9) as \( T_n = \max_{1 \leq i \leq n} |S_i| \). The test statistic is the corresponding quantity \( \hat{T}_n \) that replaces the error in \( T_n \) by the local linear residuals. The critical value was generated by the bootstrap sample of \( \max_{m+1 \leq i \leq n-m+1} |\hat{\Phi}_{i,m}| \) where \( \hat{\Phi}_{i,m} \) is defined in (3.14). For null hypothesis (i), the test statistics was 4.05, with simulated \( p \)-value 1.2%. For null hypothesis (ii), the test statistics was 2.09 with simulated \( p \)-value 4.5%. Hence there is moderately strong evidence that there are non-zero and non-constant correlations among the three lags.

We analyzed the correlation at lags 1, 2, 3 separately. We tested the constancy in the lag-1 correlation for squared percentage change. The \( p \)-value for the test of no change points in the lag-1 correlation was 7.6%. (see Table 3), and the \( p \)-value for null hypothesis (i) of zero lag-1 correlation was 4.1%. Next we used the statistic (4.4) to identify the location of the change point of the first order correlation and found \( \hat{t}_n = 397 \), which corresponds to Jun 18th, 2013. We investigated the existence of further changes in the lag-1 correlation before
and after the Jun 18th, 2013 and concluded that there are no further structural breaks in the lag-1 correlation during the two periods at 5% significance level, with the p-value 40% and 11%, respectively, for the first and second period. For the first period, the test statistic for zero lag-1 correlation was 1.98 with p-value < 1%. For the second period, the test statistic for zero lag-1 correlation was 2.25 with p-value 1.1%. The identified change point in the lag-1 correlation is close to the date that USD/CAD significantly exceeded the boundary 1. Before this date, the exchange rate was slightly fluctuating around 1, and after this point the exchange rate increased over 1.4 and never returned to 1.

For lag-2, the testing result are also presented in Table 6. There p-value for the test of hypothesis of zero lag-2 correlation is < 1%, while the p-value of constant lag-2 correlation is 2.7%. The location of the jump time for the lag-2 correlation is 695 which corresponds to Aug 25, 2014. We also investigated the lag-2 correlation before and after Aug 25, 2014. For the hypothesis of constant lag-2 correlation, The p-values were 71% and 26% for the first and second period, respectively. For the hypothesis of zero lag-2 correlation, the p-values were 1% and 18% before and after the jump, respectively. The identified change point in the lag-2 correlation is close to the date where the Crude oil price drastically decreased from 100 USD per barrel to 50 USD per barrel. The oil price has a great impact on the economy of Canada, which is one of the decisive factors of the exchange rate.

For lag-3, the test statistic for no changes in correlation was 0.96 with p-value 43.3%. The test statistic for zero lag-3 correlation was 1.11 with p-value 58.9%. We conclude that correlations of the squared percentage changes concentrate in lag-1 and lag-2, with change points existing in both lags. Interestingly, the time of change for lag-1 and 2 correlations are different.

We further performed tests from Section 4 for relevant changes in lag-1, lag-2 correlations separately and jointly (the trajectory we considered was \( \delta_1 = \delta_2 \)) for the USD/CAD data. The estimates of the lag-1 correlation before and after the break point were −0.056 and 0.079 while, for the lag-2 correlation, the estimates before and after the jump were 0.092
and $-0.034$. The $p$-values of the tests for a relevant change in the lag-1/lag-2 correlation for different values of the threshold $\delta$ are displayed in Figure 5. At 5% significance level, we conclude that there are relevant changes with size $\delta = 0.032$ in the lag-1 correlation, $\delta = 0.024$ in the lag-2 correlation, and size $\delta = 0.026$ in lag-1 or lag-2 correlation. The $p$-values of the tests for relevant changes in the lag-1 or 2 correlation for different values of the threshold $\delta$ for the USD/CAD are displayed in Figure 5.

The correlations of the squared series are closely related to the ARCH effect. For example, Baillie and Chung (2001) estimated the GARCH model via the autocorrelations of the square of the process. Our method shows that the USD/CAD from late 2011-mid 2016 may not be well fitted by a simple ARCH/GARCH model due to the changes in the correlation structure. Further, the negative first order correlation in the first period shows that USD/CAD from late 2011-mid 2013 may not be well fitted by usual ARCH/GARCH model, due to their restriction of positive coefficients. Other models, for example the EGARCH model should be considered. We have also identified very different pattern of squared percentage changes of USD/CAD in the three lags considered: there is weak evidence against the null hypothesis of constant lag-1 correlation, strong evidence against constant lag-2 correlation, no evidence against constant lag-3 correlation, and strong evidence against constant lag-1, lag-2 and lag-3 correlations. We have no evidence against zero lag-3 correlation, while we have strong evidence against the hypotheses of zero lag-1 or lag-2 correlations.

7 Supplementary Materials

The supplementary materials contains the proofs of theorems and additional simulation results.

Acknowledgements. The work of H. Dette has been supported in part by the Collaborative Research Center “Statistical modeling of nonlinear dynamic processes” (SFB 823,
Teilprojekt A1, C1) of the German Research Foundation (DFG). Z. Zhou’s research has been supported in part by NSERC of Canada. The authors would like to thank Martina Stein who typed parts of this manuscript with considerable technical expertise. The authors would also like to thank three referees and an associate editor for their constructive comments on an earlier version of this paper.

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.

Ahamada, I. and Jouini, J. and Boutahar, M. (2004). Detecting multiple breaks in time series covariance structure, A nonparametric approach based on the evolutionary spectral density. *Applied Economics* 36,1095–1101.

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.

Baillie, Richard T., and Huimin Chung. (2001). Estimation of GARCH models from the autocorrelations of the squares of a process *Journal of Time Series Analysis* 22(6),631–650.

Berger, J. O. and Delampady, M. (1987). Testing precise hypotheses. *Statistical Science* 2(3),317–335.
Berkes, I. and Gombay, E. and Horvath, L. (2009). Testing for changes in the co-variance structure of linear processes. *Journal of Statistical Planning and Inference* 139,2044–2063.

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.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics* 31(3), 307–327.

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.

Dahlhaus, R., (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics* 25(1), 1-37.

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 Preuß P. and Vetter, M. (2011). A measure of stationarity in locally stationary processes with applications to testing. *Journal of the American Statistical Association* 106,1113–1124.

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

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

Dwivedi, Y. and Subba Rao, S. (2011). A test for second order stationarity of a time series based on the discrete Fourier transform. *Journal of Time Series Analysis* 32,68–91.
Engle, Robert F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* (1982), 987–1007.

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.

Gao, J. and Gijbels, I. (2008). Bandwidth Selection in Nonparametric Kernel Testing *Journal of the American Statistical Association* 103 1584–1594.

Tong, H. (1990). Non-linear time series, a dynamical system approach *Oxford*, *Oxford University Press*.

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.

Jin, L. and Wang, S. and Wang, H. (2015). A new non-parametric stationarity test of time series in the time domain. *Journal of the Royal Statistical Society, Series B* 77,893–922.

Killick, R. and Eckley, I. A. and Jonathan, P. (2013). A wavelet-based approach for detecting changes in second order structure within nonstationary time series. *Electronic Journal of Statistics* 7,1167–1183.

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),625–44.

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.

Paparoditis, E. (2009). Testing temporal constancy of the spectral structure of a time series. *Bernoulli* 15,1190–1221.

Paparoditis, E. (2010). Validating stationarity assumptions in time series analysis by rolling local periodograms. *Journal of the American Statistical Association* 105,839–851.

Paparoditis, E. and Preuß P. (2015). On local power properties of frequency domain-based tests for stationarity. *Scandinavian Journal of Statistics* 43,664-682.

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

Preuß P. and Vetter, M. and Dette, H. (2013). A test for stationarity based on empirical processes. *Bernoulli* 19,2715–2749.

Preuss, P., Puchstein, R., and Dette, H. (2015). Detection of multiple structural breaks in multivariate time series. *Journal of the American Statistical Association* 110,654-668.

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.
Vostrikova, L. J. (1981). Detecting ‘disorder’ in multidimensional random processes. *Soviet Mathematics Doklady* 24,55–59.

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.

Zhang, T., and Wu, W. B. (2012). Inference of time-varying regression models. *The Annals of Statistics* 40(3), 1376–1402.

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

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

Zhou, Z (2012). Measuring Nonlinear Dependence in Multivariate Time Series, A Distance Correlation Approach. *Journal of Time Series Analysis* 33, 438–457.

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.
A Proofs of main results

In this section we provide proofs of the main results, where some of the technical details are deferred to the Appendix [see Section B]. In the following discussion we will also make frequent use of the projection operator $\mathcal{P}_j(\cdot) = \mathbb{E}(\cdot | \mathcal{F}_j) - \mathbb{E}(\cdot | \mathcal{F}_{j-1})$. Throughout this section the symbol $\Rightarrow$ denotes weak convergence of a stochastic process in $C(0,1)$ with the uniform topology. 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 for $l \in \mathbb{Z}$. For series $a_n$ and $b_n$, denotes $a_n \approx b_n$ if $a_n = O(b_n)$ and $b_n = O(a_n)$.

A.1 Proof of Theorem 3.1, 3.2 and Lemma 3.1

We will start with an auxiliary Lemma, which will be used in the proof of Theorem 3.1.

Lemma A.1. Under conditions of Theorem 3.1, we have that for any fixed lag-$k$,

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

where $S_i^\circ = \sum_{s=1}^i e_i e_{i+k}$, $\hat{S}_i^\circ = \sum_{s=1}^i \hat{e}_i \hat{e}_{i+k}$.

Proof. First note that

$$S_i^\circ - \hat{S}_i^\circ = A_{n,i} + B_{n,i} + C_{n,i}, \quad (A.2)$$

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

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

Observing the estimate (B.4) in Section B.1, we have that

$$\max_{1 \leq i \leq n} B_{n,i} = O_p(b_n^{-1} + nb_n^4), \quad (A.3)$$
By Lemma B.1 (which is proved in Section B.1) it follows that

\[
\max_{[n \beta_n] \leq i \leq n - [n \beta_n]} \left| A_{n,i} - 2 \sum_{j=[n \beta_n]+1}^{i} a_{n,j} - 2 \sum_{j=1}^{[n \beta_n]} e_{j+k}(\hat{\mu}_{n}(t_j) - \mu(t_j)) \right| = O_p(n \chi_n),
\]

\[
\max_{n-[n \beta_n]+1 \leq i \leq n} \left| A_{n,i} - 2 \sum_{j=[n \beta_n]}^{n-[n \beta_n]} a_{n,j} - 2 \sum_{j=1}^{[n \beta_n]-1} e_{j+k}(\hat{\mu}_{n}(t_j) - \mu(t_j)) - 2 \sum_{j=n-[n \beta_n]+1}^{i} e_{j+k}(\hat{\mu}_{n}(t_j) - \mu(t_j)) \right| = O_p(n \chi_n),
\]

where \( \chi_n = b_n^3 + \frac{1}{n \beta_n} \), and

\[
a_{n,j} = \frac{e_{j+k}}{n \beta_n} \sum_{s=1}^{n} K_b \left( \frac{s-j}{n} \right) e_s \quad (j = 1, \ldots, n).
\]

(A.4)

A further application of the estimate (B.4) in Section B.1 and the Cauchy-Schwarz inequality gives

\[
\left\| \max_{1 \leq j \leq [n \beta_n]} \left| \sum_{i=1}^{j} e_{j+k}(\mu(t_j) - \hat{\mu}_{n}(t_j)) \right| \right\|_2 \leq \sum_{i=1}^{[n \beta_n]} \| e_{j+k} \|_4 \| \mu(t_j) - \hat{\mu}_{n}(t_j) \|_4
\]

\[
= O(\sqrt{n \beta_n} + nb_n^3 + \frac{1}{n \beta_n}),
\]

\[
\left\| \max_{n-[n \beta_n]+1 \leq j \leq n} \left| \sum_{i=n-[n \beta_n]+1}^{j} e_{j+k}(\mu(t_j) - \hat{\mu}_{n}(t_j)) \right| \right\|_2 = O(\sqrt{n \beta_n} + nb_n^3 + \frac{1}{n \beta_n}).
\]

This implies that

\[
\max_{1 \leq i \leq n} | A_{n,i} | \leq \max_{[n \beta_n] \leq i \leq n-[n \beta_n]} | \hat{A}_{n,i} | + O_p(\sqrt{n \beta_n} + nb_n^3 + \frac{1}{n \beta_n}),
\]

(A.5)

where \( \hat{A}_{n,i} = 2 \sum_{j=[n \beta_n]}^{i} a_{n,j} \) and \( a_{n,j} \) is defined in (A.4).

In the following we derive an estimate for the first term on the right-hand side of (A.5). For this purpose we consider the random variables \( \tilde{e}_{s,m} = \mathbb{E}(e_s | \varepsilon_s, \ldots, \varepsilon_{s-m}) \) and note that the sequence \( (\tilde{e}_{s,m})_{s=1}^{n} \) is \( m \)-dependent. Now define \( a_{n,j}^{(m)} = e_{j+k} \sum_{s=1}^{n} K_{b_n} \left( \frac{s-j}{n} \right) \tilde{e}_{s,m} / (n \beta_n) \) and

\[
\hat{A}_{n,i}^{(m)} = 2 \sum_{j=[n \beta_n]}^{i} a_{n,j}^{(m)},
\]

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

\[
\max_{1 \leq j \leq n} \left\| \sum_{s=1}^{n} K_{b_n} \left( \frac{s-j}{n} \right) (\tilde{e}_{s,m} - e_s) \right\|_4 \leq C \sqrt{n \beta_n m \chi^m}
\]
for some constant $\chi \in (0, 1)$. By the Cauchy-Schwartz inequality it follows that
\[
\left\| \max_{\lfloor nb_n \rfloor \leq i \leq n - \lfloor nb_n \rfloor} | \bar{A}_{n,i} - \bar{\bar{A}}_{n,i} | \right\|_2 \leq \frac{2}{nb_n} \sum_{j=\lfloor nb_n \rfloor}^{n} |e_{j+k}| \sum_{s=1}^{n} K_{bn}(\frac{s-j}{n})(\bar{e}_{s,m} - e_s) \right\|_2 \tag{A.6}
\]
\[
= O(\sqrt{mn} \chi^{-1/2}).
\]
Write $\tilde{a}_{n,j}^{(m)} = \bar{e}_{j+k,m} \sum_{s=1}^{n} K_{bn}(\frac{s-j}{n})\bar{e}_{s,m}/(nb_n)$ and $\bar{A}_{n,i}^{(m)} = 2 \sum_{j=\lfloor nb_n \rfloor}^{i} \tilde{a}_{n,j}^{(m)}$ it is easy to see that
\[
\left\| \max_{\lfloor nb_n \rfloor \leq i \leq n - \lfloor nb_n \rfloor} | \bar{A}_{n,i}^{(m)} - \bar{\bar{A}}_{n,i}^{(m)} | \right\|_2 \leq \frac{2}{nb_n} \sum_{j=1}^{n} |e_{j+k} - \bar{e}_{j+k,m}| \sum_{s=1}^{n} K_{bn}(\frac{s-j}{n})\bar{e}_{s,m} \right\|_4 \tag{A.7}
\]
Now an elementary calculation via Burkholder’s inequality shows
\[
\max_{1 \leq j \leq n} \left\| \frac{1}{nb_n} \sum_{s=1}^{n} K_{bn}(\frac{s-j}{n})\bar{e}_{s,m} \right\|_4 = O\left(\frac{1}{\sqrt{nb_n}}\right),
\]
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 (A.7)
\[
\left\| \max_{\lfloor nb_n \rfloor \leq i \leq n - \lfloor nb_n \rfloor} | \bar{A}_{n,i}^{(m)} - \bar{\bar{A}}_{n,i}^{(m)} | \right\|_2 = O(\sqrt{n/b_n} \chi^m),
\]
and an application of (A.6) yields
\[
\left\| \max_{\lfloor nb_n \rfloor \leq i \leq n - \lfloor nb_n \rfloor} | \bar{A}_{n,i}^{(m)} - \bar{\bar{A}}_{n,i}^{(m)} | \right\|_2 = O(\sqrt{n/b_n} m \chi^m). \tag{A.8}
\]
A tedious but straightforward calculation shows that $P_{j-l}(\tilde{e}_{j,m} \bar{e}_{i,m}) = 0$ for $l > 2m$. For example, if $i \geq j - m$, then by definition, $\tilde{e}_{j,m} \bar{e}_{i,m}$ is $\sigma(\tilde{e}_{j-2m}, \tilde{e}_{j-2m+1}, \ldots, \tilde{e}_{i+j})$ measurable. Consequently, $\mathbb{E}(\tilde{e}_{j,m} \bar{e}_{i,m} | F_{j-l}) = \mathbb{E}(\tilde{e}_{j,m} \bar{e}_{i,m} | F_{j-l-1}) = \mathbb{E}(\tilde{e}_{j,m} \bar{e}_{i,m})$ if $l > 2m$, which gives $P_{j-l}(\tilde{e}_{j,m} \bar{e}_{i,m}) = 0$. 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 $P_{j-l}(\tilde{e}_{j,m} \bar{e}_{i,m}) = 0$ for $l > 2m$ we obtain
\[
\left\| \max_{\lfloor nb_n \rfloor \leq i \leq n - \lfloor nb_n \rfloor} | \bar{A}_{n,i}^{(m)} - \mathbb{E} \bar{A}_{n,i}^{(m)} | \right\|_2 \leq 2 \sum_{l=0}^{2m} \left\| \max_{\lfloor nb_n \rfloor \leq i \leq n - \lfloor nb_n \rfloor} | \sum_{j=\lfloor nb_n \rfloor}^{i} P_{j-l}(\tilde{a}_{n,j}^{(m)}) | \right\|_2 \tag{A.9}
\]
Similar arguments as given in the proof of Theorem 1 in Wu (2005) show
\[
\| P_{j-l}(\tilde{a}_{n,j}^{(m)}) \|_2 \leq \frac{M}{n} \| \tilde{e}_{j+k,m} \sum_{s=1}^{n} \bar{e}_{s,m} K_{bn}(\frac{s-j}{n}) - \bar{e}_{(j-l)}(j-l) \sum_{s=1}^{n} \bar{e}_{s,m} K_{bn}(\frac{s-j}{n}) \|_2,
\]
and by the triangle inequality it follows that
\[ \|P_{j-l} \tilde{a}_{n,j}^{(m)} \|_2 \leq M(Z_{1,j} + Z_{2,j}), \]
where the terms \( Z_{1,j} \) and \( Z_{2,j} \) are defined by
\[
Z_{1,j} = \frac{1}{nb_n} \left\| \sum_{s=1}^{n} K_{bn} \left( \frac{s-j}{n} \right) \hat{\epsilon}_{s,m}^{(j-l)} \tilde{\epsilon}_{s,m} \right\|_2, \\
Z_{2,j} = \frac{1}{nb_n} \left[ \hat{\epsilon}_{j+k,m}^{(j-l)} \tilde{\epsilon}_{j+k,m} \right] \sum_{s=1}^{n} K_{bn} \left( \frac{s-j}{n} \right) \tilde{\epsilon}_{s,m}^{(j-l)} \right\|_2,
\]
\[ \hat{\epsilon}_{s,m}^{(j)} = \mathbb{E} \hat{\epsilon}_{s,m} | \hat{\epsilon}_{s-m}, \ldots, \hat{\epsilon}_{j}, \ldots, \hat{\epsilon}_{s} \] for \( s - m \leq j \leq s, \) \( \hat{\epsilon}_{s}^{(j)} = G(t_{s}, F_{s}) \) for \( b_{j} < t_{s} \leq b_{j+1} \) and we use the convention \( \hat{\epsilon}_{s}^{(j)} \) for \( j < s - m \) or \( j > s. \) Elementary calculations show that for \( l \geq 0 \)
\[ \left\| \sum_{s=1}^{n} K_{bn} \left( \frac{s-j}{n} \right) e_{s,m}^{(j-l)} \right\|_4 = O(\sqrt{nb_n}), \quad 1 \leq j \leq n, \]
while by definition \( \| \hat{\epsilon}_{j,m}^{(j-l)} - \hat{\epsilon}_{j,m} \|_4 = 0 \) for \( l > m. \) On the other hand, if \( 1 \leq j \leq n, 0 \leq l \leq m, \)
we have by Assumption (A4)
\[ \| \hat{\epsilon}_{j+k,m}^{(j-l)} - \hat{\epsilon}_{j+k,m} \|_4 \leq M \chi_{l+k}, \]
which gives \( Z_{2,j} = O\left( \frac{\chi_{l}}{\sqrt{nb_n}} \right). \) Observing that \( \hat{\epsilon}_{s,m}^{(j-l)} - \hat{\epsilon}_{s,m} = 0 \) if \( s \geq j - l + m + 1 \) or \( s \leq j - l - 1, \)
it is easy to see that \( Z_{1,j} = O\left( \frac{m}{nb_n} \right). \) It now follows from Doob’s inequality
\[ \left\| \max_{[nb_n] \leq i \leq n-[nb_n]} \left\{ \sum_{j=[nb_n]}^{i} P_{j-l} \tilde{a}_{n,j}^{(m)} \right\} \right\|_2 = O\left( \sqrt{n} \left( \frac{\chi_{l}}{\sqrt{nb_n}} + \frac{m}{nb_n} \right) \right), \]
and we obtain from (A.9) that
\[ \left\| \max_{[nb_n] \leq i \leq n-[nb_n]} \left| \tilde{A}_{n,i}^{(m)} - \mathbb{E} \tilde{A}_{n,i}^{(m)} \right| \right\|_2 = O\left( \frac{m^2}{n^{1/2}b_n} + (b_n)^{-1/2} \right). \quad (A.10) \]
Finally, similar arguments as given in the proof of Lemma 5 in Zhou and Wu (2010) show
\[ \max_{[nb_n] \leq i \leq n-[nb_n]} \mathbb{E} \left| \tilde{A}_{n,i}^{(m)} \right| = O\left( \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\chi_{i-j}}{(nb_n)} \right) = O(b_n^{-1}). \]
Observing (A.5), (A.8) and (A.10) and taking \( m = M \log n \) for a sufficiently large constant \( M > 0 \)
yields \( \max_{1 \leq i \leq n} |A_{n,i}| = O(p(\sqrt{nb_n} + nb_n^3 + b_n^{-1})). \) Similarly \( \max_{1 \leq i \leq n} |C_{n,i}| = O(p(\sqrt{nb_n} + nb_n^3 + b_n^{-1})). \)
Consequently, the assertion (A.1) follows from (A.2), (A.3) and these two estimates.
A.1.1 Proof of Lemma 3.1

Define \( N(i) = \frac{1}{L} \left( \sum_{j=i-L+1}^{i} e_{j}^{2} - \sum_{j=i+L-1}^{i+1} e_{j}^{2} \right) \) and recall the definition of \( M(i) \) in (3.6). By similar arguments as given in the proof of Lemma B.3 (note that \( \iota > 8 \)) we have \( \|M(i) - N(i)\|_{4} = b_{n}^{2} + \frac{1}{\sqrt{n}b_{n}}, \) and Proposition B.1 yields

\[
\max_{L \leq i \leq n-L+1} |M(i) - N(i)| = O_{p}\left(n^{1/4}b_{n}^{2} + \frac{1}{n^{1/4}b_{n}^{1/2}}\right). \tag{A.11}
\]

Consider the case that \( i \in B := \{i : |t_{i} - \hat{t}_{i}| > 2L\} \). Then by our assumption on the variance function, there exists a large constant \( C \), such that \( |\mathbb{E}N(i)| \leq CL/n \) for \( L \leq i \leq n-L+1, \) \( i \in B \). By Lemma B.3 and Lemma B.4 it now follows \( \|N(i) - \mathbb{E}N(i)\|_{1/2} \leq CL^{-1/2} (L \leq i \leq n-L+1, \) \( i \in B \)), which gives

\[
\max_{L \leq i \leq n-L+1, i \in B} |N(i)| = O_{p}(L^{-1/2}n^{2/4} + L/n).
\]

Combining this estimate with (A.11) yields

\[
\max_{L \leq i \leq n-L+1, i \in B} |M(i)| = O_{p}\left(L^{-1/2}n^{2/4} + L/n + n^{1/4}b_{n}^{2} + \frac{1}{n^{1/4}b_{n}^{1/2}}\right).
\]

Similarly, we can show that \( M([n\hat{t}_{i}]) = \sigma(t_{i}^{+}) - \sigma(t_{i}^{-}) + O_{p}\left(n^{1/4}b_{n}^{2} + \frac{1}{n^{1/4}b_{n}^{1/2}} + L^{-1/2} + L/n\right). \)

Let \( L = [n^{4/\iota} \log \log n] \). The choice of \( L \) implies that

\[
\lim_{n \to \infty} \mathbb{P}\left( |M([n\hat{t}_{i}])| > \max_{L \leq i \leq n-L+1, i \in B} |M(i)| \right) = 1,
\]

which completes the proof of Lemma 3.1. \( \square \)

A.1.2 Proof of Theorem 3.1

We restrict ourselves to the case of a variance function with 1 abrupt change point. The situation that the variance changes smoothly with time could be shown similarly and easier, with the fact that \( t^{*}_{n} \in [\zeta, 1 - \zeta] \). Recall that for any fixed lag-\( k \), \( S_{i}^{(k)} = \sum_{j=1}^{i} W_{j}^{(k)}, \) \( \hat{S}_{i}^{(k)} = \sum_{j=1}^{i} \hat{W}_{j}^{(k)} \) where

\[
W_{j}^{(k)} = \frac{e_{i}e_{i+k}}{\sigma(t_{i})\sigma(t_{i+k})}, \quad \hat{W}_{j}^{(k)} = \frac{\hat{e}_{i}e_{i+k}}{\hat{\sigma}^{2}(t_{i})}, \quad T_{n,i}^{(k)} = |S_{i}^{(k)} - \frac{i}{n} S_{n}^{(k)}|.
\]

We will show the estimate

\[
\max_{1 \leq i \leq n-k} |\hat{S}_{i}^{(k)} - S_{i}^{(k)}| = O_{p}(nc_{n}^{2} + nb_{n}^{3}c_{n}^{-1/4} + b_{n}^{-1}c_{n}^{-1} + n^{1-\nu'}), \tag{A.12}
\]

5
which implies \( \max_{1 \leq i \leq n-k} |\hat{T}(k)_{n,i} - T(k)_{n,i}| = O_p(nc^2_n + nb^3_n c_n^{-1/4} + b^{-1}_n c_n^{-1} + n^{-1-\nu'}) \), where \( \nu' \) is a constant which satisfies \( \nu' \in \left( \frac{1}{2}, 1 - \frac{4}{c} \right) \). Define

\[
T_n = \max_{1 \leq i \leq n-r} |(\hat{T}^{(r_1)}_{n,i}, \ldots, \hat{T}^{(r_j)}_{n,i})^T|,
\]

then it follows from Section 5 in Zhou (2013) that \( T_n/\sqrt{n} \) converges weakly to the distribution of the random variable \( K_1 \) defined in Theorem 3.1. By our choice of the bandwidth \( b_n \) we have \( nc^2_n + nb^3_n c_n^{-1/4} + b^{-1}_n c_n^{-1} + n^{-1-\nu'} = o(\sqrt{n}) \), and the assertion of Theorem 3.1 follows from (A.12). 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 \( \bar{S}_i^{(k)} \) in the proof of the estimate (A.12). With the notation \( \bar{S}_i = \frac{\hat{\sigma}_{i+k}}{\sigma^2(t_i)} \) we obtain

\[
\max_{1 \leq j \leq n} \left| S_j - \bar{S}_j \right| \leq \max_{1 \leq j \leq n} \sum_{i=1}^{j} \frac{|e_i| e_{i+k} \cdot |\sigma(t_i) - \sigma(t_{i+k})|}{\sigma^2(t_i) \sigma(t_{i+k})} = O_p(1), \tag{A.13}
\]

where we have used the fact that the variance function is Lipschitz continuous before and after \( t_v \). Let \( \bar{S}_j = \sum_{i=1}^{j} \frac{\hat{\sigma}_{i+k}}{\sigma^2(t_i)} \), where the estimate \( \hat{\sigma}(t_i) \) has been replaced by the “true” variance \( \sigma^2(t_i) \). By Lemma A.1, it can be seen that by similar argument,

\[
\max_{1 \leq j \leq n} \left| S_j - \bar{S}_j \right| = O_p(\sqrt{nb_n + nb^3_n + b^{-1}_n}). \tag{A.14}
\]

Define

\[
\Lambda_j := (\bar{S}_j - S_j) = \sum_{i=1}^{j} \frac{\hat{\sigma}_{i+k}}{\sigma^2(t_i)} (-\hat{\sigma}(t_i) + \sigma^2(t_i)),
\]

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

\[
\bar{\Lambda}_j := \sum_{i=1}^{j} \frac{\hat{\sigma}_{i+k}}{\sigma^4(t_i)} (-\hat{\sigma}^2(t_i) + \sigma^2(t_i)) = \sum_{i=1, t_i \notin [t_v - n^{-\nu'}, t_v + n^{-\nu'}]}^{j} \frac{\hat{\sigma}_{i+k}}{\sigma^4(t_i)} (+\hat{\sigma}^2(t_i) + \sigma^2(t_i)) + O_p(n^{-1-\nu'})
\]

(here the estimator in the denominator has been replaced by the true variance function, and the remaining order is due to Lemma B.3), and obtain

\[
\max_{1 \leq j \leq n} |\Lambda_j - \bar{\Lambda}_j| \leq \sum_{i=1}^{n} \frac{\hat{\sigma}_{i+k}}{\sigma^2(t_i)} \frac{2(\hat{\sigma}^2(t_i) - \sigma^2(t_i))}{\sigma^2(t_i) \sigma^4(t_i)} 1(t_i \notin [t_v - n^{-\nu'}, t_v + n^{-\nu'}]) + O_p(n^{-1-\nu'}). \tag{A.15}
\]

For the expectation of the right-hand side it follows

\[
E \left[ \sum_{i=1}^{n} \frac{\hat{\sigma}_{i+k}}{\sigma^2(t_i)} \frac{2(\hat{\sigma}^2(t_i) - \sigma^2(t_i))}{\sigma^2(t_i) \sigma^4(t_i)} 1(t_i \notin [t_v - n^{-\nu'}, t_v + n^{-\nu'}]) \right] \leq C \sum_{i=1}^{n} \frac{\|\hat{\sigma}_i\|_4 \|e_{i+k}\|_4 \|\sigma^2(t_i) - \sigma^2(t_i)\|_2}{\sigma^2(t_i) \sigma^4(t_i) \sigma^2(t_i)} 1(t_i \notin [t_v - n^{-\nu'}, t_v + n^{-\nu'}]) \tag{A.16}
\]
By Lemma B.3 of Section B.1, we have that
\[
\|\hat{\mu}_{bn}(t) - \mu(t)\|_4 = O\left(b_n^2 + \frac{1}{\sqrt{nb_n}}\right),
\] (A.17)
which implies \(\|\hat{e}_i\|_4 \leq C\). On the other hand, Corollary B.2 in Section B.1 shows
\[
\max_{1 \leq i \leq n} \|(\hat{\sigma}^2(t_i) - \sigma^2(t_i))^2 \mathbf{1}(t_i \not\in [\tilde{t}_v - n^{-\nu'}, \tilde{t}_v + n^{-\nu'}])\|_2 = O\left(b_n^4 + \frac{1}{nb_n} + c_n^4 + \frac{1}{nc_n}\right),
\]
and we obtain from (A.15), (A.16) and Proposition B.1 in Section B.2 the estimate
\[
\max_{1 \leq j \leq n} |\Lambda_j| \leq \max_{1 \leq j \leq n} |\hat{\Lambda}_j| + \max_{1 \leq j \leq n} |\Lambda_j - \hat{\Lambda}_j|
\]
\[
= \max_{1 \leq j \leq n} |\hat{\Lambda}_j| + O_p(nb_n^4 + b_n^{-1} + nc_n^4 + c_n^{-1}).
\]
Now the remaining problem is to derive an appropriate estimate for the quantity \(\max_{1 \leq j \leq n} |\hat{\Lambda}_j|\). For this purpose note that \(\hat{\Lambda}_j = \hat{\lambda}_{j,1} + \hat{\lambda}_{j,2} + O_p(n^{-1}),\) where
\[
\hat{\lambda}_{j,1} = \sum_{i=1}^{s} \frac{(\hat{e}_{i+k} - e_{i+k})(\sigma^2(t_i) - \hat{\sigma}^2(t_i))}{\sigma^4(t_i)} \mathbf{1}(t_i \not\in [\tilde{t}_v - n^{-\nu'}, \tilde{t}_v + n^{-\nu'}]),
\]
\[
\hat{\lambda}_{s,j,2} = \sum_{s=i}^{j} \frac{e_{i+k}(\sigma^2(t_i) - \hat{\sigma}^2(t_i))}{\sigma^4(t_i)} \mathbf{1}(t_i \not\in [\tilde{t}_v - n^{-\nu'}, \tilde{t}_v + n^{-\nu'}]).
\]
and \(\hat{\lambda}_{i,2} = \hat{\lambda}_{1,2} + \hat{A}_i = \{t_i \not\in [\tilde{t}_v - n^{-\nu'}, \tilde{t}_v + n^{-\nu'}]\}\) for short. By Lemma B.1, Corollary B.2 of Section B.1 and the estimate (A.17) it is easy to see that
\[
\mathbb{E}\left[\max_{1 \leq j \leq n} |\hat{\lambda}_{j,1}|\right] \leq \sum_{i=1}^{n} \frac{\|\hat{e}_{i+k} - e_{i+k}\|_2}{\sigma^4(t_i)} \|\sigma^2(t_i) - \hat{\sigma}^2(t_i)\|_2 \mathbf{1}(A_i) = O(\pi_n),
\] (A.19)
\[
\mathbb{E}\left[\max_{1 \leq j \leq \lfloor nb_n + nc_n \rfloor} |\hat{\lambda}_{j,2}|\right] \leq \sum_{i=1}^{\lfloor nb_n + nc_n \rfloor} \frac{|e_{i+k}|}{\sigma^4(t_i)} \|\sigma^2(t_i) - \hat{\sigma}^2(t_i)\|_2 \mathbf{1}(A_i) = O(\pi_n),
\] (A.20)
\[
\max_{n - \lfloor nb_n + nc_n \rfloor \leq j \leq n} |\hat{\lambda}_{j,2}| \leq |\hat{\lambda}_{n - \lfloor nb_n + nc_n \rfloor - 1,1}| + \sum_{i=n - \lfloor nb_n + nc_n \rfloor}^{n} \frac{|e_{i+k}|}{\sigma^4(t_i)} |\sigma^2(t_i) - \hat{\sigma}^2(t_i)|
\]
\[
\leq \max_{\lfloor nb_n + nc_n \rfloor = s \leq n - \lfloor nb_n + nc_n \rfloor - 1} |\hat{\lambda}_{s,j,2}|
\]
\[
+ \sum_{i=1}^{\lfloor nb_n + nc_n \rfloor} \frac{|e_{i+k}|}{\sigma^4(t_i)} |\sigma^2(t_i) - \hat{\sigma}^2(t_i)|
\]
\[
= \max_{\lfloor nb_n + nc_n \rfloor = s \leq n - \lfloor nb_n + nc_n \rfloor - 1} |\hat{\lambda}_{s,j,2}| + O_p(\pi_n).
\] (A.21)
where the constants $\pi_n$ and $\pi_n$ are given by $\pi_n = nb_n c_n^2 + \sqrt{\pi_n b_n^2} + \sqrt{\pi_n c_n^2} + \frac{1}{\sqrt{b_nc_n}}$, $\pi_n = (nb_n + nc_n)(b_n^2 + c_n^2 + \frac{1}{\sqrt{b_n}} + \frac{1}{\sqrt{c_n}})$, respectively.

In order to prove a corresponding estimate for the remaining term

$$\max_{[nb_n + nc_n] = s \leq j \leq n - [nb_n + nc_n] - 1} |\bar{\lambda}_{s,j,2}|$$

in (A.21) we study the asymptotic behavior of the quantity $\hat{\sigma}^2(t) - \sigma^2(t)$. By similar arguments as given above, we have that

$$\max_{[nb_n + nc_n] = s \leq j \leq n - [nb_n + nc_n] - 1} |\bar{\lambda}_{s,j,2}| \leq \max_{[nb_n + nc_n] = s \leq j \leq n - [n\nu - n^{1-\nu'} - nb_n - nc_n]} |\bar{\lambda}_{s,j,2}|$$

$$+ \max_{[n\nu + n^{1-\nu'} + nb_n + nc_n] = s \leq j \leq n - [nb_n + nc_n]} |\bar{\lambda}_{j,2}| + O_p(\pi_n + n^{1-\nu'})$$

and by Corollary B.1 in Section B.1 it easily follows that

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

(A.22)

where $\mathbb{T}_n = [c_n, \tilde{t}_v - c_n - n^{-\nu'}] \cup [\tilde{t}_v + c_n + n^{-\nu'}, 1 - c_n]$. We now consider the decomposition

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

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 B.1 in Section B.1 we obtain

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

The triangle inequality and Proposition B.1 in Section B.2 imply

$$\left\| \sup_{t \in \mathbb{T}_n^\prime} \left[ \sum_{i=1}^n K_{cn}(t_i - t) [Q_{1,i} - \frac{2e_i}{nb_n} \sum_{j=1}^n e_j K_{bn}(t_i - t_j) - \mu_2 \hat{\mu}(t_i)b_n^2 e_i] \right] \right\|_4 = O(nb_n^3 c_n^{3/4} + b_n^{-1} c_n^{3/4}),$$

(A.23)

where we use the notation $\mathbb{T}_n^\prime = [b_n + c_n, \tilde{t}_v - b_n - c_n - n^{-\nu'}] \cup [\tilde{t}_v + b_n + c_n + n^{-\nu'}, 1 - b_n - c_n]$. Similar arguments as given in the calculation of $\max_{[nb_n] \leq i \leq n - [nb_n]} |A_{n,i}|$ in the proof of (A.1) and the summation by parts formula show

$$\left\| \sup_{t \in \mathbb{T}_n^\prime} \left[ \frac{1}{nb_n} \sum_{i=1}^n K_{cn}(t_i - t) e_i \sum_{j=1}^n e_j K_{bn}(t_i - t_j) \right] \right\|_2 = O(b_n^{-1}),$$

$$\left\| \sup_{t \in \mathbb{T}_n^\prime} \left[ \sum_{i=1}^n K_{cn}(t_i - t) \mu_2 \hat{\mu}(t_i)b_n^2 e_i \right] \right\|_2 = O(n^{1/2} b_n^2),$$
and (A.23) gives \( \| \sup_{t \in \mathbb{T}_n} \left| \sum_{i=1}^{n} K_{b_n}(t_i - t)Q_{1,i} \right| \|_2 = O(nb_n^{3/4}c_n^{-1} + b_n^{-1} + n^{1/2}b_n^2) \). On the other hand, note that

\[
\left\| \sup_{t \in \mathbb{T}_n} \left| \sum_{i=1}^{n} K_{c_n}(t_i - t)Q_{2,i} \right| \right\|_2 \leq R_{n,1} + R_{n,2}
\]

where

\[
R_{n,1} = \left\| \sup_{t \in \mathbb{T}_n} \left| \sum_{i=1}^{n} K_{c_n}(t_i - t) \left( \frac{1}{nb_n} \sum_{j=1}^{n} e_j K_{b_n}(t_i - t_j) + \frac{\mu_2 \hat{\mu}(t_i) b_n^2}{2} \right) \right| \right\|_2
\]

\[
R_{n,2} = \left\| \sup_{t \in \mathbb{T}_n} \left| \sum_{i=1}^{n} K_{c_n}(t_i - t) \left( \mu(t_i) - \hat{\mu}(t_i) + \frac{1}{nb_n} \sum_{j=1}^{n} e_j K_{b_n}(t_i - t_j) + \frac{\mu_2 \hat{\mu}(t_i) b_n^2}{2} \right) \right| \right\|_2 \times \left( \mu(t_i) - \hat{\mu}(t_i) - \frac{1}{nb_n} \sum_{j=1}^{n} e_j K_{b_n}(t_i - t_j) - \frac{\mu_2 \hat{\mu}(t_i) b_n^2}{2} \right) \right\|_2.
\]

Proposition B.1 in Section B.2 and similar calculations as given in the proof of (A.1) show that

\[
R_{n,1} = O\left( \frac{nb_n^3c_n^{3/4}}{\sqrt{n}b_n} + nb_n^{1/2}c_n^{1/2} \right) = O(\sqrt{nb_n^{5/2}c_n^{1/2}} + nb_n^{5/2}c_n^{1/2}).
\]

while a further application of Lemma B.1 in Section B.1 yields

\[
R_{n,2} = O\left( \frac{nb_n^3c_n^{3/4}}{\sqrt{n}b_n} + nb_n^{5/2}c_n^{1/2} \right) = O(\sqrt{nb_n^{5/2}c_n^{1/2}} + nb_n^{5/2}c_n^{1/2}).
\]

Consequently, combining the arguments in (A.22)-(A.24), it follows that

\[
\left\| \sup_{t \in \mathbb{T}_n} \left| \hat{\sigma}^2(t) - \sigma^2(t) - \frac{\mu_2 \hat{\sigma}^2(t)c_n^{3/2}}{2} - \frac{1}{nc_n} \sum_{i=1}^{n} K_{b_n}(t_i - t) (e_i^2 - \mathbb{E}(e_i^2)) \right| \right\|_2 = O(\pi_n),
\]

where \( \pi_n = c_n^3 + \frac{1}{nc_n} + b_n^3c_n^{-1/4} + \frac{1}{nb_n c_n} + \frac{b_n^2}{\sqrt{nc_n}} + c_n^{-1/2}b_n^{-1}n^{-1} + c_n^{-1/2}b_n^4 + b_n^{5/2}(nc_n)^{-1/2} + b_n^5c_n^{-1/2} \).

Let

\[
T_{n,1} = [\lfloor nb_n + nc_n \rfloor, nt_n - n^{-1} - nb_n - nc_n] \cap \mathbb{Z},
\]

\[
T_{n,2} = [nt_n + n^{-1} - nb_n - nc_n, n - \lfloor nb_n + nc_n \rfloor] \cap \mathbb{Z},
\]

defining \( \bar{W}_i = \frac{e_i e_i + k}{\sigma^2(t_i)} \) and \( Z_i' = e_i^2 - \mathbb{E} e_i^2 \), then it follows from (A.25) that

\[
\mathbb{E} \left( \max_{|nb_n + nc_n| = s,j \in T_{n,1}} \left| \bar{\lambda}_{s,j,2} + \sum_{i=|nb_n + nc_n|}^{j} \frac{\bar{W}_i (\sum_{j=1}^{n} K_{c_n}(t_j - t_i)Z_i' + \mu_2 \hat{\sigma}^2(t_i)nc_n^{3/2})}{\sigma^2(t_i)nc_n} \right| \right) \leq \sum_{i \in T_{n,1}} \frac{1}{\sigma^4(t_i)} \| e_i e_i + k \|_2 \left\| \hat{\sigma}^2(t_i) - \sigma^2(t_i) - \frac{1}{nc_n} \sum_{j=1}^{n} K_{c_n}(t_j - t_i)Z_i' - \mu_2 \hat{\sigma}^2(t_i)nc_n^{3/2}/2 \right\|_2 = O(n\pi_n).
\]

(A.26)
By the Cauchy-Schwarz inequality we obtain \( \| \tilde{W}_i - \tilde{W}^{(m)}_i \|_4 = O(\chi|i-m|) \), \( \| Z'_i - Z^{(m)}'_i \|_4 = O(\chi|i-m|) \), where \( Z^{(m)}_i = (e^{(m)}_i)^2 - E(e^{(m)}_i)^2 \), \( \tilde{W}^{(m)}_i = \frac{e^{(m)}_{i+k}}{\sigma(t_i)^2} \), and

\[
  e^{(m)}_i = G_j(t_i, F^{(m)}_i), \quad \text{if} \quad b_j < t_i \leq b_{j+1}.
\]

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

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

Then by a similar \( m \)-dependent approximating technique as given in the proof of (A.1) we get

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

Similarly, and more easily one obtains

\[
  \max_{j \in T_{n,1}} \left| \sum_{i=1}^{j} e_{i+k} e_i \mu_2 \sigma^2(t_i) c_{n}^2 / (2\sigma^4(t_i)) \right| = O_p(nc_n^2). \quad (A.27)
\]

Hence, it follows from (A.26) and (A.27) that

\[
  \max_{j \in T_{n,1}} |\bar{\lambda}_{j,2}| = O_p(n\pi_n + nc_n^2).
\]

Similarly,

\[
  \max_{j \in T_{n,2}} |\bar{\lambda}_{j,2}| = O_p(n\pi_n + nc_n^2),
\]

which implies, observing (A.19) - (A.21),

\[
  \max_{1 \leq j \leq n} |\bar{\Lambda}_j| = O_p(\pi_n + \pi_n + n\pi_n + nc_n^2).
\]

Combining this result with the estimates (A.13), (A.14) and (A.18), and by our choice of the bandwidths, we have that

\[
  \max_{1 \leq j \leq n} \left| \sum_{i=1}^{j} (\hat{W}_i - W_i) \right| = O_p(n\pi_n^2 + nb_n^3c_n^{-1/4} + b_n^{-1}c_n^{-1} + n^{1-\nu'})
\]

which establishes the estimate (A.12) and completes the proof of Theorem 3.1. \( \square \)
A.1.3 Proof of Theorem 3.2

We proof the theorem when there is an abrupt change of variance at $\tilde{t}_v$. The case that there is no abrupt change in variance is similar and easier. Recall that $\hat{W}_i^{(k)} = \frac{\hat{e}_i e_{i+k}}{\sigma(t_i) \sigma(t_{i+k})}$ and $W_i^{(k)} = \frac{e_i e_{i+k}}{\sigma(t_i) \sigma(t_{i+k})}$. We consider the corresponding partial sums $S_{j,m}^{(k)} = \sum_{r=j}^{j+m-1} W_r^{(k)}$, $\tilde{S}_{j,m} = \sum_{r=j}^{j+m-1} \tilde{W}_r^{(k)}$ and define $\Phi_{n,m}(t)$ in (3.14) and $\hat{S}_{j,m}(\phi_j, \sigma_j) = (S_{j,m}^{(n)}, \ldots, S_{j,m}^{(n)})^T$, $\hat{S}_n = \hat{S}_{1,n}$. Similarly, we define $S_{j,m}, \hat{S}_n, \tilde{S}_{j,m}$ and $\tilde{S}_n$ and the linear interpolation on the interval $[0, 1]$ by

$$\tilde{\Phi}_{m,n}(t) = \tilde{\Phi}_{[nt],m} + (nt - [nt])(\tilde{\Phi}_{[nt]+1,m} - \tilde{\Phi}_{[nt],m}).$$  \hspace{1cm} (A.28)

The assertion follows from the continuous mapping theorem if the weak convergence

$$\{\Phi_{m,n}(t)\}_{t \in [0,1]} \Rightarrow \{U(t)\}_{t \in [0,1]}$$

conditional on $\mathcal{F}_n$ can be established. For a proof of this statement define $(\Phi_{m,n}(t), \Phi_{m,n}(t))$ and $(\hat{\Phi}_{m,n}(t), \hat{\Phi}_{m,n}(t))$ by replacing $(\tilde{S}_{j,m}, \tilde{S}_n)$ in the definition of $\Phi_{m,n}$ and $\hat{\Phi}_{m,n}$ with $(S_{j,m}, S_n)$ and $(\tilde{S}_{j,m}, \tilde{S}_n)$, respectively. Note that similar arguments as given in the proof of Theorem 3 in Zhou (2013) show that $\{\Phi_{m,n}(t)\}_{t \in [0,1]} \Rightarrow \{U(t)\}_{t \in [0,1]}$. The assertion of Theorem 3.2 then follows from the estimate

$$\sup_{t \in [0,1]} |\Phi_{m,n}(t) - \Phi_{m,n}(t)| = O_p \left( \frac{m}{\sqrt{n}} \right) + \sqrt{m\delta_n},$$

where $\delta_n = (c_n^2 + \frac{1}{\sqrt{n}c_n} + b_n^2 + \frac{1}{\sqrt{n}b_n})c_n^{-1/4}\log n$. In order to prove (A.29) let $C$ denote a sufficiently large constant, which may vary from line to line in the following calculations, and consider the event

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

$$B_n = \left\{ \sup_{t \in [0, t_{u,n} - n^{-\nu'}, |t|_{t_{u,n} - n^{-\nu'}}]} |\sigma^2(t) - \sigma^2(t)| \leq C \left( c_n^2 + \frac{1}{\sqrt{n}c_n} + b_n^2 + \frac{1}{\sqrt{n}b_n}c_n^{-1/4} \right) \log n \right\},$$

where $\nu' \in (\frac{1}{2}, 1 - \frac{4}{l})$. By Lemma B.3 and Corollary B.1 of Section B.1, we have that $\lim_{n \to \infty} \mathbb{P}(A_n \cap B_n) = 1$. Then it is easy to see that

$$\|\Phi_{m,n} - \Phi_{m,n}\| = O \left( \frac{m^{1/2}}{n} + \frac{1}{\sqrt{n}} \right) = O \left( \frac{m}{\sqrt{n}} \right) + \sqrt{n\delta_n}.$$}

Write $E_n = A_n \cap B_n$. On the other hand, for $1 \leq j \leq n - m + 1$ and any fix lag-$k_i$, $1 \leq i \leq l$, the estimate (omit the superscript for short)

$$\mathbb{E}[(S_{j,m} - \hat{S}_{j,m})^2 1(E_n)] = \mathbb{E} \left\{ \sum_{r=j}^{j+m-1} \frac{(\hat{e}_i \hat{e}_{i+k} - e_i e_{i+k})}{\sigma(t_i) \sigma(t_{i+k})} 1(E_n) \right\}^2 = O(m^2\delta_n^2)$$

11
for \( j \not\in \left[\lfloor n_\nu - n^{1-\nu'} \rfloor - m - 1, \lfloor n_\nu + n^{1-\nu'} \rfloor + m + 1 \right] \), and
\[
E[(S_{j,m} - \hat{S}_{j,m})^2 \mathbf{1}(E_n)] \leq C m^2, \text{ for } j \in \left[\lfloor n_\nu - n^{1-\nu'} \rfloor - m - 1, \lfloor n_\nu + n^{1-\nu'} \rfloor + m + 1 \right]
\]
Similarly,
\[
E[(S_n - \hat{S}_n)^2 \mathbf{1}(E_n)] = O(m^2 + n^{2-2\nu'} + n^2 \delta_n^2).
\]
Note that
\[
\|\Phi_{n,m} - \hat{\Phi}_{n,m}\|^2 = \frac{1}{m(n - m + 1)} \sum_{s=1}^{n-m+1} \sum_{i=1}^{n} \left( S_{j,m}^{(r_s)} - \hat{S}_{j,m}^{(r_s)} - \frac{m}{n} (S_n^{(r_s)} - \hat{S}_n^{(r_s)}) \right)^2
\]
\[
= O(m \delta_n^2 + \frac{m}{n} + \frac{m}{n^{1-\nu'}}).
\]
An application of Doob’s inequality and Proposition B.3 in Section B.2 finally yields
\[
\max_{1 \leq i \leq n-m+1} |\Phi_{i,m} - \hat{\Phi}_{i,m}| = O_p \left( \frac{m}{\sqrt{n}} + \frac{\sqrt{m}}{n^{\nu'/2}} + \sqrt{m} \delta_n \right),
\]
\[
\max_{1 \leq i \leq n-m+1} |\hat{\Phi}_{i,m} - \Phi_{i,m}| = O_p \left( \frac{m}{\sqrt{n}} + \frac{\sqrt{m}}{n^{\nu'/2}} + \sqrt{m} \delta_n \right).
\]
The estimate (A.29) now follows from this result and definition (A.28) and an application of triangle inequality, which completes the proof of Theorem 3.2.

\section*{A.2 Proof of Lemma 4.1 - 4.2}

In order to simplify the notation define
\[
G_n^{(k)}(m) = S_n^{(k)} - \frac{m}{n} S_n^{(k)}, \quad \hat{G}_n^{(k)}(m) = \hat{S}_n^{(k)} - \frac{m}{n} \hat{S}_n^{(k)},
\]
where as before, \( S_n^{(k)} = \sum_{i=1}^{m} \frac{e_{i+1}}{\sigma(t_i)\sigma(t_{i+1})} \), \( \hat{S}_n^{(k)} = \sum_{i=1}^{m} \frac{\hat{e}_{i+1}}{\hat{\sigma}(t_i)\hat{\sigma}(t_{i+1})} \), Then it is easy to see that the estimator \( \hat{t}_n^{(k)} \) of the change point in the correlation function defined in (4.4) can be represented as
\[
\hat{t}_n^{(k)} = \frac{1}{n} \arg\max_{1 \leq m \leq n} (\hat{G}_n^{(k)}(m))^2.
\]

\subsection*{A.2.1 Proof of Lemma 4.1}

We fix a lag-\( r_s \) for some \( 1 \leq s \leq l \). Recall that under the null hypothesis (4.1), we have \( \rho_{u,r_s} = \rho_1^{(s)} \) for \( u \leq \lfloor nt_{r_s} \rfloor \) and \( \rho_{u,r_s} = \rho_2^{(s)} + \Delta_{r_s} = \rho_2^{(s)} \) for \( u > \lfloor nt_{r_s} \rfloor \), where \( \Delta_{r_s} \) is an unknown (without loss of generality) positive constant. We omit the superscript and subscript \( r_s \) in this proof. A simple calculation shows that
\[
f_n(m) := E G_n(m) = n(m(n)t(n) - m(n) \wedge t(n))\Delta,
\]
where we used the notation $m(n) = m/n$ and $t(n) = \lfloor nt \rfloor / n$. 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

$$\max_{1 \leq i \leq n} \left| S_i - \mathbb{E}(S_i) - \sum_{j=1}^{i} \kappa_s(t_j)V_j \right| = o_p(n^{1/4} \log n), \quad (A.31)$$

where $\kappa_s$ is the $s_{th}$ diagonal element of $\kappa$, which is defined in assumption (A5). Define $\Xi_j = \sum_{j=1}^{t} \kappa_s(t_i)V_i$. By the arguments given in the proof of Theorem 3.1, we have

$$\max_{1 \leq m \leq n} |G_n(m) - \hat{G}_n(m)| = O_p(\varrho_n), \quad (A.32)$$

where $\varrho_n = nc_n^2 + nb_n \epsilon_n^{-1/4} + b_n^{-1} \epsilon_n - n^{1-\nu'}$ and $\nu' \in (1/2, 1 - 4/9)$. Now a similar reasoning as given in the proof of Lemma 5 of Zhou and Wu (2010) and assumptions (A3) (A4) and (A5) yield that there exists a constant $C$ such that $\kappa_s^2(t) \leq C$ for all $t \in [0, 1]$. Then it is easy to see that $\|\Xi_n\|^2 = O(n)$. By Doob’s inequality, we have that

$$\max_{1 \leq j \leq n} |\Xi_j| = O_p(\sqrt{n}), \quad (A.33)$$

and observing (A.31) we obtain

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

Define $\hat{V}_n(m) = \hat{G}_n^2(m) - \hat{G}_n^2([nt])$, note that $\hat{V}_n([nt]) = 0$ and consider a constant $\beta \in (1/2, 2/3)$, such that $n^{1-\beta} / \varrho_n \rightarrow \infty$. By the choices of $b_n$ and $c_n$, there exists qualified $\beta$. Observing the definition (A.30) and the estimate (A.31), it follows that

$$\max_{1 \leq m \leq n} \left| G_n^2(m) - \left( f_n(m) + \Xi_m - \frac{m}{n} \Xi_n \right)^2 \right| = O_p(n^{5/4} \log n). \quad (A.34)$$

By (A.33), we have $\max_{1 \leq m \leq n} (\Xi_m - \frac{m}{n} \Xi_n)^2 = O_p(n)$, and together with (A.32) and (A.34) this yields

$$\max_{m \in \mathcal{M}_n} \hat{V}_n(m) = \max_{m \in \mathcal{M}_n} \left[ G_n^2(m) - G_n^2([nt]) \right] + O_p(n \varrho_n) = \max_{m \in \mathcal{M}_n} \left\{ f_n^2(m) - f_n^2([nt]) + 2(f_n(m) - f_n([nt]))\Xi_m + 2f_n([nt])(\Xi_m - \Xi_{[nt]}) - 2\frac{m}{n} f_n(m)\Xi_n + 2\frac{nt}{n} f_n([nt])\Xi_n \right\} + O_p(n \varrho_n + n^{5/4} \log n), \quad (A.35)$$

where the maxima are taken over the set

$$\mathcal{M}_n = \{ m \mid [nt] - [n^{1-\beta/2}] \leq m \leq [nt] - [n^{1-\beta}] \}.$$
Observing the definition of \( f_n(m) \) in (A.30) we have for some positive constant \( C \),

\[
\max_{m \in M_n} (f^2_n(m) - f^2_n([nt])) \leq -Cn^{2-\beta},
\]

and (A.33) implies

\[
\max_{m \in M_n} (f_n(m) - f_n([nt])) \Xi_m = O_p(n^{3/2-\beta/2} \log n),
\]

\[
\max_{m \in M_n} \left( \frac{m}{n} f_n(m) - \frac{[nt]}{n} f_n([nt]) \right) \Xi_n = O_p(n^{3/2-\beta/2} \log n).
\]

Using the representation \( \Xi_m - \Xi_{[nt]} = \sum_{i=m+1}^{[nt]} \sigma(t_i)V_i \) and similar arguments as in the derivation of (A.33) yields

\[
\max_{m \in M_n} (\Xi_m - \Xi_{[nt]}) = O_p(n^{1/2(1-\beta/2)} \log n).
\]

Consequently,

\[
\max_{m \in M_n} f_n([nt]) [\Xi_m - \Xi_{[nt]}] = O_p(n^{3/2-\beta/4} \log n).
\]

(A.36)

By our choice of \( \beta \), it now follows from (A.35) - (A.36) that

\[
\mathbb{P}\left( \limsup_{n \to \infty} \max_{m \in M_n} \hat{V}_n(m) = -\infty \right) = 1.
\]

(A.37)

On the other hand, similar arguments give the estimates

\[
\max_{1 \leq m \leq [nt]-[n^{1-\beta/2}]} [f^2_n(m) - f^2_n([nt])] \leq -Cn^{2-\beta/2},
\]

\[
\max_{1 \leq m \leq [nt]-[n^{1-\beta/2}]} f_n([nt]) [\Xi_m - \Xi_{[nt]}] = O_p(n^{3/2} \log n),
\]

\[
\max_{1 \leq m \leq [nt]-[n^{1-\beta/2}]} \left( \frac{m}{n} f_n(m) - \frac{[nt]}{n} f_n([nt]) \right) \Xi_n = O_p(n^{3/2} \log n),
\]

\[
\max_{1 \leq m \leq [nt]-[n^{1-\beta/2}]} [f_n(m) - f_n([nt])] \Xi_m = O_p(n^{3/2} \log n),
\]

and by our choice of \( \beta \) we obtain

\[
\mathbb{P}(\limsup_{n \to \infty} \max_{1 \leq m \leq [nt]-[n^{1-\beta/2}]} \hat{V}_n(m) = -\infty) = 1.
\]

Combined with (A.37) this gives

\[
\mathbb{P}(\limsup_{n \to \infty} \max_{1 \leq m \leq [nt]-[n^{1-\beta}]} \hat{V}_n(m) = -\infty) = 1,
\]

14
and it can be shown by similar arguments that
\[ \mathbb{P}(\limsup_{n \to \infty} \max_{[nt]+[n^{1-\beta}] \leq m \leq n} \hat{V}_n(m) = -\infty) = 1. \]
Consequently, it follows that
\[ \lim_{n \to \infty} \mathbb{P}(|nt | - [nt] | \leq n^{1-\beta}) = 1, \]
which proves (4.6) of Lemma 4.1. In the case where the variance has no jump at time \( t \), the result (4.5) follows from the fact that for any lag-\( r_s \), \( \hat{G}_n^{(r_s)}(m)/\sqrt{n} \) converges weakly to some Gaussian process \( \{U^{(s)}(u) - uU^{(s)}(1)\}_{u \in [0,1]} \), which implies \( \hat{t}^{(r_s)}_n \xrightarrow{D} \hat{T}^{(r_s)} = \arg\max_{u \in (0,1)} |U^{(s)}(u) - uU^{(s)}(1)| \), where the Gaussian process \( \{U^{(s)}(u)\}_{u \in [0,1]} \) is the \( s \)-th entry of the vector Gaussian process \( \{U(t)\}_{t \in [0,1]} \) which is defined in Theorem 3.1.

**A.2.2 Proof of Lemma 4.2**

We fix a lag-\( r_s \) for some \( 1 \leq s \leq l \) and then omit the superscript/subscript \( r_s \). Recall the definition of (3.8), the notation \( W_j = W_{j}^{(r_s)} = e_{s}e_{j+r_s}/\sqrt{\sigma_j \sigma_{j+r_s}} \), and denote the change point by \( t = t_{r_s} \). Finally define
\[
\Delta_{n,1} = \sum_{j=1}^{[nt]} W_j, \quad \Delta_{n,2} = \frac{1}{n - [nt]} \sum_{j= [nt]+1}^{n} W_j \quad (A.38)
\]
We first consider the situation of (4.6) in the main article, that is \( |\Delta| > 0 \). From the proof of Theorem 3.1 we have that
\[
\Delta_{n,1} - \mathbb{E}[\Delta_{n,1}] = O_p\left(\frac{1}{\sqrt{n}}\right), \quad \Delta_{n,2} - \mathbb{E}[\Delta_{n,2}] = O_p\left(\frac{1}{\sqrt{n}}\right).
\]
Since \( \Delta = \rho_2 - \rho_1 = \mathbb{E}[\Delta_{n,2}] - \mathbb{E}[\Delta_{n,1}] \) 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 \( A_n = \{|\hat{t}_n - t| \leq \frac{C}{\sqrt{n}}\} \). Then by Lemma 4.1, we have that \( \lim_{n \to \infty} \mathbb{P}(A_n) = 1 \). This yields
\[ (\Delta_{n,1} - \hat{\Delta}_{n,1})I(A_n) = I(A_n)(A_n + B_n + C_n), \]
where
\[
A_n = \sum_{j=1}^{[nt]} \frac{W_j}{[nt]} - \sum_{j=1}^{[nt]} \frac{W_j}{[nt]}, \quad B_n = \sum_{j=1}^{[nt]} \left( \frac{W_j}{[nt]} - \frac{\hat{W}_j}{[nt]} \right), \quad (A.39)
\]
\[
C_n = \sum_{j=1}^{[nt]} \left( \frac{\hat{W}_j}{[nt]} - \frac{\hat{W}_j}{[nt]} \right).
\]
It is easy to see that $I(\mathcal{A}_n) A_n = O_p(\frac{1}{\sqrt{n}})$. Using the same arguments as in the proof of Theorem 3.1, we obtain $I(\mathcal{A}_n) B_n = o_p(\sqrt{n}/n) = o_p(1/\sqrt{n})$ and

$$I(\mathcal{A}_n) \cdot C_n = I(\mathcal{A}_n) \left( \sum_{j=1}^{[n/\sqrt{n}]} W_j \frac{[nt] - [nt]}{[nt]} \right) \leq C \cdot I(\mathcal{A}_n) \sum_{j=1}^{[n/\sqrt{n}]} W_j \frac{1}{n\sqrt{n}}$$

$$= C \cdot I(\mathcal{A}_n) \left( \sum_{j=1}^{[nt]} W_j + o_p(\sqrt{n}) \right) \frac{1}{n\sqrt{n}}$$

$$= C I(\mathcal{A}_n) \cdot \left( \sum_{j=1}^{[nt]} W_j + \sum_{j=[nt]+1}^{[n/\sqrt{n}]} W_j I(t \leq \hat{t}_n) - \sum_{j=[nt]+1}^{[n/\sqrt{n}]} W_j I(t > \hat{t}_n) + o_p(\sqrt{n}) \right) \frac{1}{n\sqrt{n}}$$

$$= O_p\left( \frac{1}{\sqrt{n}} \right). \quad (A.40)$$

Combining (A.39) - (A.40) and using Proposition B.3 in Section B.2 shows $\Delta_{n,1} - \hat{\Delta}_{n,1} = O_p\left( \frac{1}{\sqrt{n}} \right)$. Similarly, we have $\Delta_{n,2} - \hat{\Delta}_{n,2} = O_p\left( \frac{1}{\sqrt{n}} \right)$, and the assertion of the lemma follows when $|\Delta| > 0$.

For the case that $|\Delta| = 0$, define the following two functions of $u$, $0 \leq u \leq 1$,

$$\hat{\Delta}_{n,1}(u) = \frac{1}{[nu]} \sum_{j=1}^{[nu]} \hat{\sigma}_j^2(t_j), \quad \hat{\Delta}_{n,2}(u) = \frac{1}{n - [nu]} \sum_{j=[nu]+1}^{n} \hat{\sigma}_j^2(t_j).$$

Then using Doob’s inequality and similar arguments to that in Theorem 3.1, we have that

$$\max_{\frac{1}{\log n} \leq u \leq \frac{1}{\log n}} |\hat{\Delta}_{n,1}(u)| = O_p\left( \frac{\log n}{\sqrt{n}} \right), \quad \max_{\frac{1}{\log n} \leq u \leq \frac{1}{\log n}} |\hat{\Delta}_{n,2}(u)| = O_p\left( \frac{\log n}{\sqrt{n}} \right). \quad (A.41)$$

Recall the definition of $\hat{T}_n = \arg\max_{x \in (0,1)} |\hat{U}(x) - x\hat{U}(1)|$ in the proof of Lemma 4.1, where $\hat{U}(x) := U^{(rs)}(x)$ for short. Write $\tilde{U}(x) = U(x) - x\hat{U}(1), \mathcal{W}_n = [0, \frac{1}{\log n}] \cup [1 - \frac{1}{\log n}, 1], \tilde{\mathcal{W}}_n = [\frac{1}{\log n}, 1 - \frac{1}{\log n}]$. Then by observing the variance structure, we can see that

$$\lim_{n \to \infty} \mathbb{P}\left\{ \max_{x \in \tilde{\mathcal{W}}_n} \tilde{U}(x) \geq \max_{x \in \mathcal{W}_n} \hat{U}(x) \right\} = 0,$$

which shows that the event $\tilde{\mathcal{W}}_n := \{ \hat{t}_n \in \tilde{\mathcal{W}}_n \}$ satisfies that $\lim_{n \to \infty} \mathbb{P}(\tilde{\mathcal{W}}_n) = 1$. By (A.41) and Proposition B.3, we have that

$$\hat{\Delta}_{n,1} = O_p\left( \frac{\log n}{\sqrt{n}} \right), \quad \hat{\Delta}_{n,2} = O_p\left( \frac{\log n}{\sqrt{n}} \right), \quad (A.42)$$

which finishes the proof.
A.3 Proof of Theorem 4.1 and 4.2
A.3.1 Proof of Theorem 4.1

We consider the non-observable analogue
\[ T^{(k),r}_n = \frac{3}{t_k^2(1-t_k)^2} \int_0^1 (U^{(k)}_n)^2(s) ds. \]

of the statistic \( T^{(k),r}_n \) defined in (4.7), where the process \( U^{(k)}_n \) is given by
\[ U^{(k)}_n(s) = \frac{1}{n} \sum_{j=1}^{[ns]} \frac{e_j e_{j+k}}{\sigma(t_j) \sigma(t_{j+k})} - \frac{s}{n} \sum_{j=1}^{n} \frac{e_j e_{j+k}}{\sigma(t_j) \sigma(t_{j+k})}. \]

It follows from the proof of Theorem 3.1 and Lemma 4.1 that, for any fixed lag \( r_k \) \((1 \leq k \leq l)\), we have that
\[ \{ \sqrt{n}(U^{(r_k)}_n(u) + (u \wedge t - ut) \Delta_{r_k}) \}_{u \in [0,1], k \in [1,l]\cap \mathbb{Z}} \Rightarrow \{ U^{(k)}_n(u) - u U^{(k)}_n(1) \}_{u \in [0,1], k \in [1,l]\cap \mathbb{Z}} \quad (A.43) \]
whenever \( \Delta_{r_k} \neq 0 \). 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 \( \{ \sqrt{n}(T^{(r_k),r}_n - \Delta_{r_k}^2) \}_{k \in [1,l]\cap \mathbb{Z}} \Rightarrow \{ Z^{(r_k)}(\Delta_{r_k}) \}_{k \in [1,l]\cap \mathbb{Z}} \), where the random variable \( Z^{(r_k)} \) is defined in Theorem 4.1. By the proof of Theorem 3.1, we have that for \( 1 \leq k \leq l \), and constant \( v' \) satisfies \( v' \in (\frac{1}{2}, 1 - \frac{4}{t}) \)
\[ \sup_{0 \leq s \leq 1} n|U^{(r_k)}_n(s) - \hat{\Delta}_{r_k}^{(k)}(s)| = O_p(nc^2_n + nb^3_n c^{-1/4}_n + b^{-1}_n c^{-1}_n + n^{1-v'}), \]

From (A.43) it follows that \( \int_0^1 |U^{(r_k)}_n(s)| ds = O_p(1) \). Consequently, we have that for \( 1 \leq k \leq l \),
\[ n^{1/2} \int_0^1 [(U^{(r_k)}_n)^2(s) - (\hat{\Delta}_{r_k}^{(k)})^2(s)] ds \leq \]
\[ \sup_{0 \leq s \leq 1} n^{1/2}|U^{(r_k)}_n(s) - \hat{\Delta}_{r_k}^{(k)}(s)| \int_0^1 |U^{(r_k)}_n(s) + \hat{\Delta}_{r_k}^{(k)}(s)| ds \]
\[ \leq 2n^{1/2} \sup_{0 \leq s \leq 1} |U^{(r_k)}_n(s) - \hat{\Delta}_{r_k}^{(k)}(s)| \int_0^1 |U^{(r_k)}_n(s)| ds + n^{1/2} \sup_{0 \leq s \leq 1} |U^{(r_k)}_n(s) - \hat{\Delta}_{r_k}^{(k)}(s)| \]
\[ = O_p(n^{1/2}c^2_n + n^{1/2}b^3_n c^{-1/4}_n + n^{-1/2}b^{-1}_n c^{-1}_n + n^{1/2-v'}), \]

which completes the proof. \( \square \)
A.3.2 Proof of Theorem 4.2

For lag-$k$, recall the definition of $\hat{\Delta}^{(k)}$, $\hat{\Phi}_{i,m}^{A,(k)}$ in (4.11), (4.12), (4.13) and define

$$ A^{(k)}_j = \frac{e_j e_{j+k}}{\sigma(t_j)\sigma(t_{j+k})} - \Delta_k 1(j \geq \lfloor nt_k \rfloor), $$

$$ \Phi_{i,m}^{A,(k)} = \frac{1}{\sqrt{m(n-m+1)}} \sum_{j=1}^{n-m+1} (S^{A,(k)}_{j,m} - \frac{m}{n} S^{A,(k)}_n) R_j, $$

where $S^{A,(k)}_{j,m} = \sum_{r=j}^{j+m-1} A^{(k)}_r$, $S^{A,(k)}_n = \sum_{r=1}^{n} A^{(k)}_r$. We introduce the processes

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

$$ \hat{\Phi}_{m,n}^{A,(k)}(s) = \hat{\Phi}_{[ns],m}^{A,(k)} + (ns - \lfloor ns \rfloor) (\hat{\Phi}_{[ns]+1,m}^{A,(k)} - \hat{\Phi}_{[ns],m}^{A,(k)}), $$

and note that by Zhou (2013), $\{\Phi_{m,n}^{A,(k)}(s) \}_{s \in [0,1]} \Rightarrow U(s)$ conditional on $\mathcal{F}_n$, where $\Phi_{m,n}^{A,(k)}(s) = (\Phi_{m,n}^{A,(r_1)}(s), ..., \Phi_{m,n}^{A,(r_1)}(s))^T$. The assertion of Theorem 4.2 is therefore a consequence of the estimate

$$ \max_{1 \leq u \leq l} \sup_{s \in (0,1)} |\Phi_{m,n}^{A,(r_u)}(s) - \Phi_{m,n}^{A,(r_u)}(s)| = O_p \left( \frac{m}{\sqrt{n}} \right) + \left( \frac{m \log n}{\sqrt{n}} \right)^{1/2} + \sqrt{m \delta_n}, \tag{A.44} $$

To see this, note that for any fixed lag-$r_u$ ($1 \leq u \leq l$)

$$ \frac{1}{m(n-m+1)} \sum_{j=1}^{n-m+1} (S^{A,(r_u)}_{j,m} - S^{A,(r_u)}_n)^2 \leq 2(I + II), $$

where

$$ I = \frac{1}{m(n-m+1)} \sum_{j=1}^{n-m+1} \left( \sum_{r=j}^{j+m-1} \frac{e_r e_{r+r_u}}{\sigma(t_r)\sigma(t_{r+r_u})} - \frac{\hat{e}_r \hat{e}_{r+r_u}}{\hat{\sigma}^2(t_r)} \right)^2, $$

$$ II = \frac{1}{m(n-m+1)} \sum_{j=1}^{n-m+1} \left( \sum_{r=j}^{j+m-1} (\Delta_{r_u} 1(r \geq \lfloor nt_{r_u} \rfloor) - \hat{\Delta}_{n}^{(r_u)} 1(r \geq \lfloor nt_n^{(r_u)} \rfloor)) \right)^2 \leq 2(II_1 + II_2), $$

$$ II_1 = \frac{1}{m(n-m+1)} \sum_{j=1}^{n-m+1} \left( \sum_{r=j}^{j+m-1} (\Delta_{r_u} - \hat{\Delta}_{n}^{(r_u)})) 1(r \geq \lfloor nt_{r_u} \rfloor) \right)^2, $$

$$ II_2 = \frac{1}{m(n-m+1)} \sum_{j=1}^{n-m+1} \left( \sum_{r=j}^{j+m-1} \Delta_{r_u} 1(r \geq \lfloor nt_{r_u} \rfloor) - 1(r \geq \lfloor nt_n^{(r_u)} \rfloor) \right)^2. $$

By the proof of Theorem 3.2, we have $I = O_p(m\delta_n^2 + \frac{m^2}{n} + \frac{m}{n^{\nu'}})$, where $\nu'$ and $\delta_n$ is defined in the proof of Theorem 3.2. First consider the case that $\Delta_{r_u} > 0$. Similarly by Lemma 4.2 and
Proposition B.3, \( II_1 = O_p\left(\frac{m\log^2 n}{n}\right) \). Let \( \mathcal{W}_n = \{|\hat{r}_{ru}\rangle - t_{ru}| \leq n^{-\alpha'}\} \) for some \( \alpha' \in (1/2, \nu') \). Then \( \lim_{n \to \infty} P(\mathcal{W}_n) = 1 \) and an application of Proposition B.3 shows that \( II_2 = O_p\left(\frac{m\log n}{\sqrt{n}}\right) \). So
\[
\frac{1}{m(n - m + 1)} \sum_{j=1}^{n-m+1} (\hat{S}_{j,m}^{A_{ru}} - S_{j,m}^{A_{ru}})^2 = O_p\left(\frac{m^2}{n} + \frac{m\log n}{\sqrt{n}}\right),
\]
Similarly
\[
\frac{m}{n^2(n - m + 1)} \sum_{j=1}^{n-m+1} (\hat{S}_n^{A_{ru}} - S_n^{A_{ru}})^2 = O_p\left(\frac{m^2}{n} + \frac{m\log n}{\sqrt{n}}\right),
\]
By a similar argument as given in the proof of Theorem 3.2 and an application of Doob’s inequality we can show
\[
\sup_{s \in (0,1)} |\Phi_{m,n}^{A_{ru}}(s) - \Phi_{m,n}^{A_{ru}}(s)| = O_p\left(\frac{m}{\sqrt{n}} + \left(\frac{m\log n}{\sqrt{n}}\right)^{1/2} + \sqrt{m\delta_n}\right), \quad (A.45)
\]
When \( \Delta_{ru} = 0 \) it follows from (A.42) that \( II = O_p\left(\frac{m\log^2 n}{n}\right) \). Similarly (A.45) holds. Thus (A.44) holds, which finishes the proof.

\[\square\]

A.4 Proof of Algorithm 4.1

Proof. For any lag-\( r_s \), if \( \Delta_{ru} = 0 \), then the type 1 error is protected since \( \hat{T}_{ru}^{(r)} = O_P(1/n) \) and \( M_{ru}^{(r)} \) is symmetric. Otherwise, the algorithm is valid in view of Lemma 4.2 and Proposition B.3.

\[\square\]

B More technical details

B.1 Uniform bounds for nonparametric estimates

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

Lemma B.1. If assumptions (A1)-(A3) are satisfied and \( b_n \to 0, nb_n \to \infty \), we have
\[
\sup_{t \in \mathbb{T}_n} \left| \hat{\mu}_{b_n}(b_n) - \mu(b_n) - \frac{\mu_2\hat{\mu}(t)}{2} b_n^2 - \frac{1}{nb_n} \sum_{i=1}^{n} e_i K_{b_n}(t_1 - t) \right| = O(b_n^3 + \frac{1}{nb_n}),
\]
where \( \mathbb{T}_n = [b_n, 1 - b_n] \).
Proof. With the notations
\[ S_{n,l}(t) = \frac{1}{nb_n} \sum_{i=1}^{n} \left( \frac{t_i - t}{b_n} \right)^l K_{b_n}(t_i - t), \]
\[ R_{n,l}(t) = \frac{1}{nb_n} \sum_{i=1}^{n} Y_i \left( \frac{t_i - t}{b_n} \right)^l K_{b_n}(t_i - t), \]

\((l = 0, 1, \ldots)\) we obtain the representation
\[
\begin{bmatrix}
\hat{\mu}_{b_n}(t) \\
\frac{1}{nb_n} \sum_{i=1}^{n} e_i Y_i \\
\frac{1}{nb_n} \sum_{i=1}^{n} e_i \nu_1(t_i - t)
\end{bmatrix} = S_{n,0}(t) \begin{bmatrix}
\hat{\mu}_{b_n}(t) \\
\frac{1}{nb_n} \sum_{i=1}^{n} e_i Y_i \\
\frac{1}{nb_n} \sum_{i=1}^{n} e_i \nu_1(t_i - t)
\end{bmatrix} = S_{n,0}(t) - \frac{1}{nb_n} \sum_{i=1}^{n} e_i \nu_1(t_i - t)
\]
\[
\begin{bmatrix}
\hat{\mu}_{b_n}(t) \\
\frac{1}{nb_n} \sum_{i=1}^{n} e_i Y_i \\
\frac{1}{nb_n} \sum_{i=1}^{n} e_i \nu_1(t_i - t)
\end{bmatrix} = S_{n,0}(t) \begin{bmatrix}
\hat{\mu}_{b_n}(t) \\
\frac{1}{nb_n} \sum_{i=1}^{n} e_i Y_i \\
\frac{1}{nb_n} \sum_{i=1}^{n} e_i \nu_1(t_i - t)
\end{bmatrix} = S_{n,0}(t) - \frac{1}{nb_n} \sum_{i=1}^{n} e_i \nu_1(t_i - t)
\]

for the local linear estimate \(\hat{\mu}_{b_n}\), where the last identity defines the \(2 \times 2\) matrix \(S_n(t)\) and the vector \(R_n(t)\) in an obvious manner. By elementary calculation and a Taylor expansion we have
\[
S_n(t) \begin{bmatrix}
\hat{\mu}_{b_n}(t) - \mu(t) \\
\frac{1}{nb_n} \sum_{i=1}^{n} e_i K_{b_n}(t_i - t) + \frac{1}{2} \bar{\mu}(t) \mu_2 b_n^2 \\
\frac{1}{nb_n} \sum_{i=1}^{n} e_i K_{b_n}(t_i - t) (\frac{t_i - t}{b_n})
\end{bmatrix} = O(b_n^3 + \frac{1}{nb_n})
\]
uniformly with respect to \(t \in \Xi_n\). 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 \(t \in \Xi_n\), which yields
\[
\sup_{t \in \Xi_n} \left| \hat{\mu}_{b_n}(t) - \mu(t) - \frac{1}{nb_n} \sum_{i=1}^{n} e_i K_{b_n}(t_i - t) \right| = O(b_n^3 + \frac{1}{nb_n}).
\]

Therefore the lemma follows from the definition of the estimate \(\hat{\mu}_{b_n}\) in (3.4). \(\square\)

**Lemma B.2.** Assume that the conditions of Lemma B.1 hold, then
\[
\sup_{t \in \Xi'_n} \left| c(t)(\hat{\mu}_{b_n}(t) - \mu(t)) \right| = \frac{1}{nb_n} \sum_{i=1}^{n} \left[ c(t_i) - \mu(t_i) - \mu_2 \bar{\mu}(t_i) (\frac{t_i - t}{b_n}) \right] e_i K_{b_n}(t_i - t) + \frac{b_n^2}{2} \bar{\mu}(t_i) (\nu_2(t_i) - \nu_1(t_i) \nu_3(t_i)) = O(b_n^3 + \frac{1}{nb_n}),
\]
where \(\Xi'_n = [0, b_n] \cup (1 - b_n, 1]\), \(\nu_{j,b_n}(t) = \int_{t/b_n}^{(1-t)/b_n} x^j K(x) dx\) and \(c(t) = \nu_0(t) \nu_2(t) - \nu_1^2(t) \).\]

Proof. For any \(t \in [0, b_n] \cup (1 - b_n, 1]\), using (B.1), we obtain
\[
S_n(t) \begin{bmatrix}
\hat{\mu}_{b_n}(t) - \mu(t) \\
\frac{1}{nb_n} \sum_{i=1}^{n} e_i Y_i - \mu(t) (t_i - t) \\
\frac{1}{nb_n} \sum_{i=1}^{n} e_i \nu_1(t_i - t)
\end{bmatrix} = S_n(t) - \frac{1}{nb_n} \sum_{i=1}^{n} e_i \nu_1(t_i - t)
\]
\[
S_n(t) \begin{bmatrix}
\hat{\mu}_{b_n}(t) - \mu(t) \\
\frac{1}{nb_n} \sum_{i=1}^{n} e_i Y_i - \mu(t) (t_i - t) \\
\frac{1}{nb_n} \sum_{i=1}^{n} e_i \nu_1(t_i - t)
\end{bmatrix} = S_n(t) - \frac{1}{nb_n} \sum_{i=1}^{n} e_i \nu_1(t_i - t)
\]
and a Taylor expansion yields

\[
S_n(t) \left[ \frac{1}{nb_n} \sum_{i=1}^{n} e_i K_{b_n}(t_i - t) + \frac{b_n^2}{2} \nu_{2,b_n}(t) \bar{\mu}(t) \right] = \left[ \frac{1}{nb_n} \sum_{i=1}^{n} e_i K_{b_n}(t_i - t) \left( \frac{t_i - t}{b_n} \right) + \frac{b_n^2}{2} \nu_{3,b_n}(t) \bar{\mu}(t) \right] + O\left( b_n^3 + \frac{1}{nb_n} \right)
\]

(B.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

\[
S_n(t) = \left[ \nu_{0,b_n}(t) \nu_{1,b_n}(t) \right] + O\left( \frac{1}{nb_n} \right).
\]

(B.3)

Therefore, combining (B.2) and (B.3), it follows that

\[
c(t)(\hat{\mu}_{b_n}(t) - \mu(t)) = \frac{1}{nb_n} \sum_{i=1}^{n} \left[ \nu_{2,b_n}(t_i - t) \left( \frac{t_i - t}{b_n} \right) \right] e_i K_{b_n}(t_i - t) + \frac{b_n^2}{2} \bar{\mu}(t) \left( \nu_{2,b_n}(t_i - t) - \nu_{1,b_n}(t_i - t) \nu_{3,b_n}(t) \right) + O\left( b_n^3 + \frac{1}{nb_n} \right)
\]

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 \( \mu \) in the \( \| \cdot \|_4 \)-norm.

**Lemma B.3.** Assume that assumptions (A1)-(A4) are satisfied and that \( nb_n^3 \to \infty \), \( nb_n^6 \to 0 \), then

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

(B.4)

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

(B.5)

**Proof.** Observing the stochastic expansion in Lemma B.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 projection operator \( \mathcal{P}_i \) we note that

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

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

\[
\left\| \sum_{i=1}^{n} \mathcal{P}_{i-k} e_i K_{b_n}(t_i - t) \right\|_4^2 \leq C \sum_{i=1}^{n} \left\| \left( \mathcal{P}_{i-k} e_i K_{b_n}(t_i - t) \right) \right\|_4^2,
\]
and condition (A4) implies $\| \sum_{i=1}^{n} P_{i-k} e_i K_{b_n}(t_i - t) \|_4 = O(\sqrt{n b_n} \chi_k)$, uniformly with respect to $t \in [0,1]$. This yields

$$\sup_{t \in [0,1]} \left\| \sum_{i=1}^{n} e_i K_{b_n}(t_i - t) \right\|_4 = O(\sqrt{n b_n}). \tag{B.6}$$

Similar arguments show $\sup_{t \in [0,1]} \| \frac{\partial}{\partial t} \sum_{i=1}^{n} e_i K_{b_n}(t_i - t) \|_4 = O(\sqrt{n b_n} b_n^{-1})$. By Proposition B.1 in Section B.2 it follows that

$$\| \sup_{t \in [0,1]} \sum_{i=1}^{n} e_i K_{b_n}(t_i - t)/(nb_n) \|_4 = O((nb_n)^{-1/2} b_n^{-1/4}), \tag{B.7}$$

and by Lemma B.1 we obtain

$$\left\| \sup_{t \in T_n} (\hat{\mu}_{b_n}(t) - \mu(t))^2 - \left( \frac{1}{nb_n} \sum_{i=1}^{n} e_i K_{b_n}(t_i - t) + \frac{\mu_2 \hat{\mu}(t)}{2} b_n^2 \right)^2 \right\|_2 = O\left( \frac{\chi_n}{\sqrt{n b_n} b_n^{1/4}} + \frac{1}{n b_n} \right), \tag{B.8}$$

where $\chi_n = b_n^2 + 1/n b_n$. Hence

$$\| \sup_{t \in T_n} (\hat{\mu}_{b_n}(t) - \mu(t))^2 \|_2 = O\left( \frac{1}{n b_n^{3/2}} + \frac{1}{n b_n^2} \right).$$

By similar arguments and Lemma B.2 it follows that

$$\| \sup_{t \in T_n} (\hat{\mu}_{b_n}(t) - \mu(t))^2 \|_2 = O\left( \frac{1}{n b_n^{3/2}} + \frac{1}{n b_n^2} \right),$$

and a combination of the last two estimates gives (B.5). On the other hand, Lemma B.1, (B.6) and similar but easier arguments as given in the derivation of (B.8) show that

$$\sup_{t \in T_n} \| (\hat{\mu}_{b_n}(t) - \mu(t))^2 \|_2 = O\left( \frac{1}{\sqrt{n b_n}} + \frac{b_n^2}{n} \right),$$

which proves the remaining estimate (B.4). \qed

The following results give a uniform bound for the $p$-mean of $\hat{\sigma}^2(t) - \sigma^2(t)$, where $\hat{\sigma}^2(\cdot)$ is the variance estimator defined above (3.7) in the main article.

**Lemma B.4.** Suppose that Assumptions (A1)-(A4) are satisfied, $c_n \to 0$, $nc_n \to \infty$, and i) The variance function $\sigma^2$ is strictly positive, twice differentiable with a Lipschitz continuous second derivative $\sigma''$. Then the estimate $\hat{\sigma}^2(t,k) = \hat{\sigma}_{c_n,b_n}(t,k)$ defined above (3.7) in the main article satisfies

$$\max_{k \in [n^{-\alpha}, n^{-\beta}]} \sup_{t \in T_{c_n}} \left| \sigma^2(t,k) - \sigma(t) - \frac{\mu \sigma^2(t) c_n^2}{2} - \frac{1}{nc_n} \sum_{i=1}^{n} (\hat{e}_i^2 - \mathbb{E}e_i^2) K_{c_n}(t_i - t) \right| = O\left( c_n^3 + \frac{1}{nc_n} \right), \tag{B.9}$$

22
\[
\max_{k \in [n\zeta], n - [n\zeta]} \sup_{t \in \mathcal{T}_{k,n}'} \left| c(t, k) - \sigma^2(t) - \frac{1}{nc_n} \sum_{i=1}^{n} \left[ \nu_{2,c_n}(t, k) - \nu_{1,c_n}(t, k) \left( \frac{t_i - t}{c_n} \right) \right] \right| = O\left( c_n^3 + \frac{1}{nc_n} \right),
\]
 \hspace{1cm} (B.10)

\[
\max_{k \in [n\zeta], n - [n\zeta]} \sup_{t \in \mathcal{T}_{k,n}'} \left| c(t, k) - \sigma^2(t) - \frac{1}{nc_n} \sum_{i=1}^{n} \left[ \nu_{2,c_n}(t, k) - \nu_{1,c_n}(t, k) \nu_{3,c_n}(t, k) \right] \right| = O\left( c_n^3 + \frac{1}{nc_n} \right),
\]
 \hspace{1cm} (B.11)

where

\[
\mathcal{T}_{k,n} = [c_n, t_k - c_n] \cup [t_k + c_n, 1 - c_n], \quad \mathcal{T}_{k,-n} = [0, c_n] \cup (t_k - c_n, t_k),
\]

\[
\mathcal{T}_{k,n}' = [t_k, t_k + c_n] \cup (1 - c_n, 1],
\]

\[
\nu_{j,c_n}(t, k) = \int_{t/c_n}^{(t/c_n + 1)/c_n} x^2 K(x) dx,
\]

\[
\nu_{j,c_n}(t, k) = \int_{(t_k - c_n)/c_n}^{(1 - c_n)/c_n} x^2 K(x) dx, \quad c(t, k) = \nu_{0,c_n}(t, k) + \nu_{2,c_n}(t, k) - \nu_{1,c_n}(t, k),
\]

(ii) If there is an abrupt change of variance happened at time \( t_\nu \), then a similar result of (B.9) holds as follows:

\[
\max_{k \in [n\zeta], n - [n\zeta]} \sup_{t \in \mathcal{T}_{k,n}'} \left| \hat{\sigma}^2(t, k) - \sigma^2(t) - \frac{\mu_2 \hat{\sigma}^2(t)}{2} - \frac{1}{nc_n} \sum_{i=1}^{n} (\hat{e}_i^2 - \mathbb{E}e_i^2) K_{c_n}(t_i - t) \right| = O\left( c_n^3 + \frac{1}{nc_n} \right),
\]

\[
\max_{k \in [n\zeta], n - [n\zeta]} \sup_{t \in \mathcal{T}_{k,n}'} \left| \hat{\sigma}^2(t, k) - \sigma^2(t) - \frac{\mu_2 \hat{\sigma}^2(t)}{2} - \frac{1}{nc_n} \sum_{i=1}^{n} (\hat{e}_i^2 - \mathbb{E}e_i^2) K_{c_n}(t_i - t) \right| = O\left( c_n^3 + \frac{1}{nc_n} \right),
\]

where \( \mathcal{T}_{t_\nu,n} = [c_n, t_k - c_n] \cup [t_\nu + c_n, 1 - c_n], \mathcal{T}_{t_\nu,n}' = [c_n, t_\nu - c_n] \cup [t_k + c_n, 1 - c_n]. \) The similar results hold for (B.10) and (B.11), with \( t \in [0, c_n] \cup [t_k - c_n, t_k] \) uniformly for \( t_k \leq t_\nu \) and \( t \in [t_k, t_k + c_n] \cup [1 - c_n, 1] \) uniformly for \( t_k \geq t_\nu \), respectively.

**Proof.** For a proof of part (i) we only show (B.11). The remaining results can be proved similarly. For any \( k \in [n\zeta], n - [n\zeta] \) and \( t \in \mathcal{T}_{k,n}' \), following the argument given in the proof of Lemma B.1, we have that

\[
S_n(t, k) \left[ (\hat{\sigma}^2(t, k) - \sigma^2(t)) \right] = \left[ \frac{1}{nc_n} \sum_{i=1}^{n} (\hat{e}_i^2 - \sigma^2(t) - \hat{\sigma}^2(t)(t_i - t)) K_{c_n}(t_i - t) \right],
\]

\[
= \left[ \frac{1}{nc_n} \sum_{i=1}^{n} (\hat{e}_i^2 - \sigma^2(t) - \hat{\sigma}^2(t)(t_i - t))(t_i - t) K_{c_n}(t_i - t) \right].
\]

23
where $S_n(t, k^+)$ is defined as follows:

$$S_n(t, k^+) = \left[ \begin{array}{cc} \nu_{0,cn}(t, k^+) & \nu_{1,cn}(t, k^+) \\ \nu_{1,cn}(t, k^+) & \nu_{2,cn}(t, k^+) \end{array} \right] + O\left( \frac{1}{nc_n} \right).$$

The lemma now follows by the same arguments as given in the proof of Lemma B.1 and Lemma B.2, and the fact that the remaining order in the notations $O(\cdot)$ is independent of $k$. Part (ii) can be shown similarly to (i) observing the definition of $\hat{\sigma}^2(t)$, the conditions on the bandwidths $c_n$ and $b_n$ and the fact that $\sigma^2(t)$ varies smoothly before and after $\tilde{t}_v$. □

**Corollary B.1.** Suppose that the conditions of Lemma B.4 hold. Let $\nu' \in (\frac{1}{2}, 1 - \frac{4}{i})$ where $i$ is defined in condition (A3) and condition (A4). Recall that $\hat{\sigma}^2(t) = \hat{\sigma}^2(t, \lfloor nt_n^* \rfloor)$, $\tilde{k} = \lfloor nt_v \rfloor$

(i) If the variance changes smoothly in the interval $(0, 1)$, then

$$\sup_{t \in [c_n, t_n^* - c_n] \cup [t_n^* + c_n, 1 - c_n]} \left| \hat{\sigma}^2(t) - \sigma^2(t) - \frac{\mu_3 \hat{\sigma}^2(t)c_n^2}{2} - \frac{1}{nc_n} \sum_{i=1}^{n} (\hat{e}_i^2 - \mathbb{E}(e_i^2))K_{cn}(t - t_i) \right| = O\left( c_n^3 + \frac{1}{nc_n} \right),$$

(B.12)

$$\sup_{t \in [0, c_n] \cup (t_n^* - c_n, t_n^*)} \left| c(t, k^*)(\hat{\sigma}^2(t) - \sigma^2(t)) - \frac{1}{nc_n} \sum_{i=1}^{n} \left[ \nu_{2,cn}(t, k^*) - \nu_{1,cn}(t, k^*) \left( \frac{t_i - t}{c_n} \right) \right] \right| \times \left[ \hat{e}_i^2 - \mathbb{E}(e_i^2) \right] K_{cn}(t_i - t) + \frac{c_n^2}{2} \hat{\sigma}^2(t) (\nu_{2,cn}(t, k^*) - \nu_{1,cn}(t, k^*) \nu_{3,cn}(t, k^*)) \right| = O\left( c_n^3 + \frac{1}{nc_n} \right),$$

(B.13)

$$\sup_{t \in (t_n^*, t_n^* + c_n) \cup (1 - c_n, 1]} \left| c(t, k^*)(\hat{\sigma}^2(t) - \sigma^2(t)) - \frac{1}{nc_n} \sum_{i=1}^{n} \left[ \nu_{2,cn}(t, k^*) - \nu_{1,cn}(t, k^*) \left( \frac{t_i - t}{c_n} \right) \right] \right| \times \left[ \hat{e}_i^2 - \mathbb{E}(e_i^2) \right] K_{cn}(t_i - t) + \frac{c_n^2}{2} \hat{\sigma}^2(t) (\nu_{2,cn}(t, k^*) - \nu_{1,cn}(t, k^*) \nu_{3,cn}(t, k^*)) \right| = O\left( c_n^3 + \frac{1}{nc_n} \right).$$

(B.14)
(ii) If the variance has an abrupt change point, then
\[
\sup_{t \in [c_n, \tilde{t}_v - c_n + n^{-\nu'}] \cup [\tilde{t}_v + c_n + n^{-\nu'}, 1 - c_n]}
\left| \hat{\sigma}^2(t) - \sigma^2(t) - \frac{\mu_2 \tilde{\sigma}^2(t) e_n^2}{2} - \frac{1}{nc_n} \sum_{i=1}^{n} (\hat{e}_i^2 - \mathbb{E}e_i^2) K_n(t_i - t) \right|
= O(c_n^3 + \frac{1}{nc_n}),
\]
\[
\sup_{t \in [0, c_n] \cup [t_n - c_n, t_n]} \left| c(t, k^*) (\hat{\sigma}^2(t) - \sigma^2(t)) - \frac{c_n^2}{2} \tilde{\sigma}^2(t) \nu^2_{2,n}(t, k^*) - \nu_1, c_n(t, k^*) \nu_3, c_n(t, k^*) \right|
= O(c_n^3 + \frac{1}{nc_n}).
\]
\[
\sup_{t \in [t_n + n^{-\nu'}, \tilde{t}_v + c_n + n^{-\nu'}] \cup (1 - c_n, 1]}
\left| c(t, k^+) (\hat{\sigma}^2(t) - \sigma^2(t)) - \frac{c_n^2}{2} \tilde{\sigma}^2(t) \nu^2_{2,n}(t, k^+) - \nu_1, c_n(t, k^+) \nu_3, c_n(t, k^+) \right|
= O(c_n^3 + \frac{1}{nc_n}).
\]
\[
\sup_{t \in (\tilde{t}_v + n^{-\nu'}, \tilde{t}_v + n^{-\nu'})} \left| \hat{\sigma}^2(t) - \sigma^2(t) \right| = O_p(1).
\]

**Proof.** Part (i) follows directly from Lemma B.4. Part (ii) follows from Lemma B.4 and the fact that \( \tilde{t}_v - t_n^* = o_p(n^{-\nu'}). \)

**Corollary B.2.** Suppose that the conditions of Lemma B.4 hold with \( \nu \geq 8. \)

(i) If there is no abrupt change in variance, then
\[
\sup_{t \in [0, 1]} \left\| \hat{\sigma}^2(t) - \sigma^2(t) \right\|_4 = O\left( c_n^2 + \frac{1}{\sqrt{nc_n}} + b_n^2 + \frac{1}{\sqrt{nb_n}} \right),
\]  
(B.15)
\[
\left\| \sup_{t \in (0, 1)} \left| \hat{\sigma}^2(t) - \sigma^2(t) \right| \right\|_4 = O\left( c_n^2 + \left( \frac{1}{\sqrt{nc_n}} + b_n^2 + \frac{1}{\sqrt{nb_n}} \right)c_n^{-1/4} \right). \]  
(B.16)

(ii) If there is an abrupt change in variance, then
\[
\sup_{t \in \left[ 0, t_n - n^{-\nu'} \right] \cup \left[ t_n + n^{-\nu'}, t_n \right]} \left\| \hat{\sigma}^2(t) - \sigma^2(t) \right\|_4 = O\left( c_n^2 + \frac{1}{\sqrt{nc_n}} + b_n^2 + \frac{1}{\sqrt{nb_n}} \right),
\]
\[
\left\| \sup_{t \in \left[ 0, t_n - n^{-\nu'} \right] \cup \left[ t_n + n^{-\nu'}, 1 \right]} \left| \hat{\sigma}^2(t) - \sigma^2(t) \right| \right\|_4 = O\left( c_n^2 + \left( \frac{1}{\sqrt{nc_n}} + b_n^2 + \frac{1}{\sqrt{nb_n}} \right)c_n^{-1/4} \right), \]
where \( \nu' \in \left( \frac{1}{2}, 1 - \frac{4}{\gamma} \right) \) for \( \nu \) defined in (A3) and (A4).

**Proof.** By Corollary B.1 (i), when there is no abrupt changes in variance, we have that
\[
\sup_{t \in [0, 1]} \left\| \hat{\sigma}^2(t) - \sigma^2(t) \right\|_4 \leq \sup_{t \in [0, 1]} \left\| (g_1(t, F_i)) \right\|_4 + \sup_{t \in [0, c_n] \cup [t_n^*, c_n]} \left\| (g_2(t, F_i)) \right\|_4 + \sup_{t \in [t_n^*, t_n^* + c_n] \cup [1 - c_n, 1]} \left\| (g_3(t, F_i)) \right\|_4 \]  
(B.17)
\[
\sup_{t \in [0, 1]} |\hat{\sigma}^2(t) - \sigma^2(t)| \leq \sup_{t \in [0, 1]} (g_1(t, F_i)) + \sup_{t \in [0, c_n]} (g_2(t, F_i)) + \sup_{t \in [t^*_n, c_n + t^*_n]} (g_3(t, F_i)) \tag{B.18}
\]

where \(g_i(t, F_i), i = 1, 2, 3\) is the approximation terms of (B.12)–(B.14), respectively. Noting that

\[
\sup_{t \in [0, 1]} \left\| \frac{1}{nc_n} \sum_{i=1}^n (e_i^2 - \mathbb{E} e_i^2) K_{c_n}(t_i - t) \right\|_4 = O\left( \frac{1}{\sqrt{nc_n}} \right).
\]

By the proof of Lemma B.3, we obtain (note that \(\iota \geq 8\))

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

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

\[
\sup_{t \in [0, 1]} \left\| \frac{1}{nc_n} \sum_{i=1}^n (e_i^2 - \hat{e}_i^2) K_{c_n}(t_i - t) \right\|_4 = O\left( b_n^2 + \frac{1}{\sqrt{nb_n}} \right).
\]

Combining with (B.17), we have shown (B.15).

Recalling for \(j = 0, 1, 2, 3\), \(\nu_{j,c_n}(t, k+) = \int_{(t_k-t)/c_n}^{(1-t)/c_n} x^j K(x) dx := \tilde{\nu}_{j,c_n}(t, t_k)\), then elementary calculations shows that

\[
\frac{\partial}{\partial t} \tilde{\nu}_{j,c_n}(t, t_k) = O(c_n^{-1}), \quad \frac{\partial}{\partial t_k} \tilde{\nu}_{j,c_n}(t, t_k) = O(c_n^{-1}), \quad \frac{\partial^2}{\partial t \partial t_k} \tilde{\nu}_{j,c_n}(t, t_k) = O(c_n^{-2}).
\]

Similar results hold for \(\nu_{j,c_n}(t, k-)\). Then (B.16) follows from (B.18) and Proposition B.2.

ii) follows from similarly arguments and the assertion ii) of Corollary B.1. \(\square\)

**Corollary B.3.** Suppose the conditions of Lemma B.4 hold, with \(\iota \geq 8\). Then

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

**Proof.** The lemma follows from Proposition B.1 in Section B.2, the triangle inequality and simple calculations. Note that the first assumption of Proposition B.1 is satisfied by the arguments in Corollary B.2. The second assumption regarding the derivative can be shown by similar arguments as given in (B.6) and (B.7). \(\square\)

### B.2 Three additional technical results

**Proposition B.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 \(t \in [0, 1]\), \(\|\Upsilon_n(t)\|_p = O(m_n), \|\bar{\Upsilon}_n(t)\|_p = O(l_n)\), 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 $\| \sup_{t \in [0,1]} |\mathcal{Y}_n(t)| \|_2 = O(\sqrt{mn \ln})$.

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

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

where $Z_{in} = \sup_{\tau_i - b_n \leq t \leq \tau_i} |\mathcal{Y}_n(t) - \mathcal{Y}_n(\tau_i)|$. Observing the inequalities

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

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

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

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

Proposition B.2. Let $\{\mathcal{Y}_n(x,y)\}_{x,y \in [0,1]}$ be a sequence of stochastic processes with differentiable paths. Assume that for some $p \geq 1$ and any $x, y \in [0,1]$, $\|\mathcal{Y}_n(x,y)\|_p = O(m_n)$, $\|\partial_x \mathcal{Y}_n(x,y)\|_p = O(l_1 n)$, $\|\partial_y \mathcal{Y}_n(x,y)\|_p = O(l_2 n)$, $\|\partial^2_{x,y} \mathcal{Y}_n(x,y)\|_p = O(l_3 n)$ where $m_n, l_n$ are sequences of real numbers, $m_n = O(l_n)$, let $c_n \rightarrow 0$. If $l_1 n < l_2 n$, then

$$\left\| \max_{x \in [0,1], y \in ([x-c_n]\wedge 0, [x+c_n]\wedge 1]} |\mathcal{Y}_n(x,y)| \right\|_p = O(m_n (c'_n)^{-2/p} c_n^{1/p}),$$

where $c'_n = \frac{l_1 n}{l_2 n} + \frac{m_n}{l_2 n} + \left(\frac{m_n}{l_2 n}\right)^{1/2}$.

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

$$\sup_{t \in (0,1)} |\mathcal{Y}_n(t)| \leq \max_{0 \leq i \leq \tilde{b}_n + 1, 0 \leq j \leq (i+1) |c_n/b_n|} |\mathcal{Y}_n(\tau_i, \tau_j)| + \max_{0 \leq i \leq \tilde{b}_n + 1, 0 \leq j \leq (i+1) |c_n/b_n|} Z_{i,j},$$

where $Z_{i,j} = \sup_{\tau_i - b_n < t_1 < \tau_i, \tau_j - b_n < t_2 < \tau_j} |\mathcal{Y}_n(t_1, t_2) - \mathcal{Y}_n(\tau_i, \tau_j)|$. Observing the inequalities

$$\|Z_{i,j}\|_p \leq \int_{\tau_i - b_n}^{\tau_i} \|\partial_x \mathcal{Y}_n(x,y)\|_p dx + \int_{\tau_i - b_n}^{\tau_i} \|\partial_y \mathcal{Y}_n(x,y)\|_p dy +$$

$$\int_{\tau_i - b_n}^{\tau_i} \int_{\tau_i - b_n}^{\tau_i} \|\partial^2_{x,y} \mathcal{Y}_n(x,y)\|_p dxdy = O(b_n^2 l_3 n + b_n l_1 n + b_n l_2 n) \quad (B.19)$$

27
and \( \max_{0 \leq i \leq b_n + 1, 0 \leq j \leq (i + \lceil cn/b_n \rceil) \wedge 1} Z_{ij}^p \leq \sum_{i=1}^{\tilde{b}_n + 1} \sum_{j=0 \vee (i - \lceil cn/b_n \rceil)}^{(i + \lceil cn/b_n \rceil) \wedge 1} Z_{i,j}^p \), we have

\[
\| \max_{1 \leq i \leq \tilde{b}_n + 1} Z_{in} \|_p = O(l_{1,n} b_n^{(p-1)/p} (\frac{c_n}{b_n})^{1/p} + l_{2,n} b_n^{(p-1)/2} (\frac{c_n}{b_n})^{1/2} + l_{3,n} b_n^{(p-1)/p} (\frac{c_n}{b_n})^{1/p}).
\]

Similarly, we obtain the estimate

\[
\| \max_{0 \leq i \leq \tilde{b}_n + 1, 0 \leq j \leq \tilde{b}_n + 1} |\gamma_n(t_i, t_j)| \|_p = O_p(b_n^{-2/p} c_n^{1/p} m_n),
\]

and picking \( b_n = c'_n \) proves the assertion. \( \square \)

**Proposition B.3.** Suppose \( A_n \) are sets such that \( \mathbb{P}(A_n) \to 0 \) as \( n \to \infty \), and \( X_n I(\tilde{A}_n) = O_p(1) \). 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 \( \mathbb{P}(|X_n| I(\tilde{A}_n) \geq M/2) \leq \epsilon/2 \) for \( n \geq N \). Then

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

for all \( n \geq N \). \( \square \)

## C Additional Simulation Result

### C.1 Simulation Results under Stationarity

We consider

\[(I^*) \quad G(t, \mathcal{F}_t) = H(t, \mathcal{F}_t) \sqrt{c(t)}/2 \text{ for } t \leq 0.5, \text{ and } G(t, \mathcal{F}_t) = H_1(t, \mathcal{F}_t) \sqrt{c(t)}/2 \text{ for } t > 0.5,
\]

where \( c(t) = 1 - (t - 0.5)^2 \), \( H(t, \mathcal{F}_t) = 0.2 H(t, \mathcal{F}_{t-1}) + \varepsilon_i \) for \( t \leq 0.5 \), \( H_1(t, \mathcal{F}_t) = (0.2 - \lambda) H_1(t, \mathcal{F}_{t-1}) + \varepsilon_i \) for \( t > 0.5 \) and random variables \( \{\varepsilon_i, i \in \mathbb{Z}\} \) are i.i.d. \( N(0, 1) \).

We use Algorithm 3.1 in page 13 of the main article to test change points in the lag-1 correlation of model \( I^* \), and compare the results with the following algorithm C.1 tailored to model \( I^* \) under the stationary assumption.

Let \( B(t) \) be a standard Brownian motion and \( \kappa^2 \) be the long run variance of \( \frac{\varepsilon_i \varepsilon_{i+1}}{\sigma^2} \). Note that under stationarity, the variance \( \sigma^2(t) \) and long run variance \( \kappa^2(t) \) are now time invariant.
Algorithm C.1.

[1] Calculate the statistic \( \hat{T}_n \) defined in (3.3) of the main article with \( \hat{e}_i \hat{e}_{i+1} \hat{\sigma}_2(t_i) \) replaced by \( \hat{L}_i = \hat{e}_i \hat{e}_{i+1} \hat{\sigma}_2^2 \), where \( \hat{e}_i \) is obtained by local linear estimation with bandwidth selected by GCV, and \( \hat{\sigma}_2 = \frac{\sum_{i=1}^{n} \hat{e}_i^2}{n} \).

[2] Estimate the long run variance by \( \hat{\kappa}_2^2 = \frac{\sum_{i=m+1}^{n-m} \hat{Q}_i^2}{(2m+1)(n-2m)} \), where \( m \) is selected by GCV.

[3] Reject the null hypothesis at nominal level \( \alpha \) if \( \hat{T}_n / \sqrt{n} > \hat{\kappa}_{M^1-\alpha} \), where \( M^1-\alpha \) is the \((1-\alpha)\)th quantile of the max_{0\leq t\leq 1} |B(t) - tB(1)|.

The simulated results are presented in Figure C.1. It shows that our method has decent power. It is slightly more powerful than Algorithm C.1 designed for stationary processes.
Table C.1: Simulated type I error of the test (3.15) for a change point in the lag-1 correlation for model II and of the test (4.15) for a relevant change in the lag-1 correlation for model III using a Biweight kernel (at the boundary point of the null). The last column represents the simulated Type I error if the bandwidth is $b_n$ selected by GCV.

| $b_n$  | 0.075 | 0.1  | 0.125 | 0.15 | 0.175 | 0.2  | 0.225 | GCV |
|--------|-------|------|-------|------|-------|------|-------|-----|
| II     | 5%    | 4.5  | 5.4   | 4.4  | 4.25  | 3.25 | 2.65  | 3.55 | 3.6 |
|        | 10%   | 10.6 | 10.75 | 9.6  | 9.6   | 9.4  | 7.5   | 8.4  | 8.9 |
| III    | 5%    | 6.1  | 5.85  | 5.95 | 6.2   | 6.3  | 4.95  | 6.05 | 5.5 |
|        | 10%   | 10.6 | 10.2  | 10.45| 10.55 | 11.65| 9.1   | 10.5 | 9.4 |

C.2 The Impact of Different Kernel Functions

Under piecewise local stationarity, Lemma B.1 in the supplementary material and Proposition 5 of Zhou (2013) imply that for a given kernel function $K$, the optimal bandwidth is given by

$$b_{optimal}^n = \left( \frac{\phi_0 \int_0^1 \tilde{\kappa}^2(t) dt}{\mu_2 \int_0^1 |\mu''(t)| dt} \right)^{1/5} n^{1/5},$$

where $\tilde{\kappa}^2(t)$ is the long run variance of $e_i$, $\mu_2 = \int x^2 K(x) dx$ and $\phi_0 = \int K^2(x) dx$. Fan and Yao (2003) pointed out that the performance of procedures with bandwidth is not very sensitive with respect to the choice of different kernel functions. We have confirmed these observations in further simulations. Exemplarily we show in Table C.1 the simulated type I error of the test (3.15) for a change point in the lag-1 correlation for model II and of the test (4.15) for a relevant change in the lag-1 correlation for model III using a Biweight kernel instead of the Epanechnikov kernel (for this kernel the corresponding results can be found in Table 1 of the main document). We do not observe any significant differences between the results obtained for the two different kernels and the simulated type I error rates are quite close to their nominal levels.

C.3 Performance under Different Sample Sizes

In this section we investigate the performance of the new tests for different sample sizes in more detail. In Table C.2 and C.3 we display the simulated type I error of the test (3.15) for a change point in the lag-1 correlation for model II and of the test (4.15) for a relevant change in the lag-1
Table C.2: Simulated Type I error of the test (3.15) for a change point in the lag-1 correlation for model II and the sample sizes 300 and 800. The last column represents the simulated Type I error if the bandwidth is $b_n$ selected by GCV.

| $b_n$ | 0.075 | 0.1 | 0.125 | 0.15 | 0.175 | 0.2 | 0.225 | GCV  |
|-------|-------|-----|-------|-----|-------|-----|-------|------|
| 300   |       |     |       |     |       |     |       |      |
| 5%    | 4.95  | 4.5 | 4.4   | 4.45| 3.8   | 4.1 | 3.6   | 5.25 |
| 10%   | 11.25 | 10.65| 10.15 | 10.5| 8.6   | 9.5 | 8.8   | 10.3 |
| 800   |       |     |       |     |       |     |       |      |
| 5%    | 3.9   | 3.25| 4.5   | 3.8 | 4.4   | 3.6 | 3.4   | 4.6  |
| 10%   | 9.3   | 9.3 | 9.6   | 9.2 | 9.1   | 8.95| 7.55  | 9.95 |

Table C.3: Simulated Type I error of the test (4.15) for a relevant change in the lag-1 correlation (at the boundary point of the null) for model III and the sample sizes 300 and 800. The last column represents the simulated Type I error if the bandwidth is $b_n$ selected by GCV.

| $b_n$ | 0.075 | 0.1  | 0.125 | 0.15 | 0.175 | 0.2  | 0.225 | GCV  |
|-------|-------|------|-------|-----|-------|------|-------|------|
| 300   |       |      |       |     |       |      |       |      |
| 5%    | 5.2   | 6.4  | 6.25  | 5.35| 5.7   | 4.55 | 5.05  | 4.95 |
| 10%   | 11.2  | 11.3 | 11.4  | 9.9 | 11.05 | 9.05 | 9.4   | 9.65 |
| 800   |       |      |       |     |       |      |       |      |
| 5%    | 5.7   | 5.7  | 6.75  | 5.6 | 6.7   | 4.9  | 4.9   | 5.45 |
| 10%   | 10.8  | 10.35| 10.8  | 9.8 | 10.85 | 8.45 | 8.7   | 10.65|
correlation (at the boundary point of the null) for model III and the sample sizes 300 and 800, respectively. Again we observe a reasonable approximation of the nominal level.

References

Fan, J. and Yao, Q (2003). Nonlinear Time Series, Nonparametric and Parametric Methods. Springer.

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.