Received: 15 October 2019 | Revised: 21 June 2020 | Accepted: 2 November 2020

DOI: 10.1111/sjos.12507

ORIGINAL ARTICLE

Scandinavian Journal of Statistics

Detecting early or late changes in linear models with heteroscedastic errors

Lajos Horváth[1] | **Curtis Miller**[1] | **Gregory Rice**[2]

[1]Department of Mathematics, University of Utah, Salt Lake City, Utah

[2]Department of Statistics and Actuarial Science, University of Waterloo Waterloo, Ontario, Canada

Correspondence
Gregory Rice, Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON N2L3G1, Canada.
Email: grice@uwaterloo.ca

Abstract

We construct and study a test to detect possible change points in the regression parameters of a linear model when the model errors and covariates may exhibit heteroscedasticity. Being based on a new trimming scheme for the CUSUM process introduced in Horváth et al. (2020), this test is particularly well suited to detect changes that might occur near the endpoints of the sample. A complete asymptotic theory for the test is developed under the null hypothesis of no change in the regression parameter, and consistency of the test is also established in the presence of a parameter change. Monte Carlo simulations show that our test is comparable to existing methods when the errors are homoscedastic. In contrast, existing methods developed for homoscedastic data are demonstrated to be ill-sized and poorly performing in the presence of heteroscedasticity, while the proposed test continues to perform well in heteroscedastic environments. These results are further demonstrated in a study of the linear connection between the price of crude oil and the U.S. dollar, and in detecting changes points in asset pricing models surrounding the COVID-19 pandemic.

KEYWORDS
change point, heteroscedastic data, linear model, Rényi statistic, Wiener process

MOS SUBJECT CLASSIFICATION
Primary 62E20, Secondary 62J05, 62P20

1 | INTRODUCTION AND MAIN RESULTS

In this paper, we consider the time varying linear model

$$y_t = \mathbf{x}_t^\top \boldsymbol{\beta}_t + \epsilon_t, \quad 1 \le t \le T, \tag{1}$$

where $\mathbf{x}_t \in \mathbb{R}^d$, and $E[\epsilon_t | \mathcal{F}_{X,t}] = 0$, with $\mathcal{F}_{X,t} = \sigma(\mathbf{X}_s, s \le t)$ denoting the σ-algebra generated by the covariates up to time t. Generally we are interested in testing the null hypothesis

$$H_0 : \ \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \ \dots \ = \boldsymbol{\beta}_T,$$

against the alternative

$$H_A : \ \text{there exists a } t^* \text{ such that } \ \boldsymbol{\beta}_1 = \ \dots \ \boldsymbol{\beta}_{t^*} \neq \boldsymbol{\beta}_{t^*+1} = \ \dots \ = \boldsymbol{\beta}_T.$$

Under the alternative the regression coefficients (loadings) change at a single time t^*. Testing H_0 against H_A is a classical problem that was initiated by Quandt, 1960, and since then the literature on this problem has developed at a steady pace. For example, Kim and Siegmund (1989) advocate the application of the maximally selected likelihood ratio test assuming independent and identically distributed normal errors in (1), and Gombay and Horváth (1994) show that the asymptotic distribution of the resulting test statistic satisfies a Darling–Erdős law. Ploberger and Krämer (1992) employ the least squares residuals in a CUSUM procedure to perform such a test, which is similar to Brown et al. (1975). Csörgő and Horváth (1997) survey several methods when the innovations are independent and identically distributed. Bai (1999) and Bai and Perron (1998) extend some of these detection techniques to dependent errors and multiple change points under the alternative. Hidalgo and Seo (2013) develop a Lagrange multiplier type test for this purpose. More recently, Horváth et al. (2020) use sequential averages of the sample residuals to develop efficient tests to detect early or late changes in the linear model parameters. For a survey on the change point problem from a time series point of view, we refer to Aue and Horváth (2013).

Generally in the literature to date it is assumed that the sequence $(\mathbf{x}_t, \epsilon_t)$'s is at least strictly stationary. Exceptions that do not necessarily focus on linear models include Gorecki et el. (2017), Harvey et al. (2016), Bardsley et al. (2017), Harris et al. (2017), and Wu and Zhou (2018), where nonstationarity is allowed in the observations that covers heteroscedasticity or changing volatilities. Commonly in applications, and as is the case with the data examples studied below, the assumption of heteroscedasticity is much more realistic. Further, available methods valid under heteroscedasticity have not been studied or optimized in terms of their ability to detect changes that might be near the beginning or end of the sample. Detecting such changes is often of interest when applying change point detection procedures retrospectively to a sample where a change is suspected to have occurred recently. In this paper, we aim to extend the residual based test of Horváth et al. (2020), which is based on a novel trimming scheme for the CUSUM process that is effective for end of sample change point detection, to the setting of heteroscedastic linear models.

In particular, we model $\mathbf{x}_t \in \mathbb{R}^d$ and ϵ_t as random variables that are interval stationary, that is to say that their distributions are allowed to change at up to M points,

$$1 < t_1 = \lfloor T\theta_1 \rfloor < t_2 = \lfloor T\theta_2 \rfloor < \dots < t_M = \lfloor T\theta_M \rfloor < T, \tag{2}$$

(we use the notation $t_0 = 0$ and $t_{M+1} = T$). We assume that the factors $\mathbf{x}_t$ and the errors ϵ_t on each interval of stationarity evolve according to Bernoulli shifts:

$$\epsilon_t = f_i(\eta_t, \eta_{t-1}, \eta_{t-2}, \dots), \quad \text{if } \ t_{i-1} < t \le t_i, 1 \le i \le M+1, \tag{3}$$

and

$$\mathbf{x}_t = \mathbf{g}_i(\eta_t, \eta_{t-1}, \eta_{t-2}, \dots), \quad \text{if } \ t_{i-1} < t \le t_i, 1 \le i \le M+1, \tag{4}$$

where the η_t are independent and identically distributed random elements of a measurable space $\mathcal{S}$. We do not require that $f_i \neq f_{i+1}$ and/or $\mathbf{g}_i \neq \mathbf{g}_{i+1}$, that is, $t_1, t_2, \dots, t_M$ are only possible times where the structure of the processes generating $\mathbf{x}_t$ and ϵ_t might change, and hence these sequences could exhibit fairly broad forms of nonstationarity. For example, $(\mathbf{x}_t, \epsilon_t)$ might evolve as a vector valued linear process whose representation changes at each point $t_1, t_2, \dots, t_M$.

Our method is based on sequentially comparing $\hat{\boldsymbol{\beta}}_{t,1}$ and $\hat{\boldsymbol{\beta}}_{t,2}$, where $\hat{\boldsymbol{\beta}}_{t,1}$ and $\hat{\boldsymbol{\beta}}_{t,2}$ are the least squares estimators for the regression coefficients computed from $\{(y_s, \mathbf{x}_s), 1 \le s \le t\}$ and $\{(y_s, \mathbf{x}_s), t < s \le T\}$, respectively. Let $\mathbf{Y}_{t,1} = (y_1, y_2, \dots, y_t)^\top$ and $\mathbf{Y}_{t,2} = (y_{t+1}, y_{t+2}, \dots, y_T)^\top$, and similarly

$$\mathbf{X}_{t,1} = \begin{pmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_t^\top \end{pmatrix} \quad \text{and} \quad \mathbf{X}_{t,2} = \begin{pmatrix} \mathbf{x}_{t+1}^\top \\ \mathbf{x}_{t+2}^\top \\ \vdots \\ \mathbf{x}_T^\top \end{pmatrix}.$$

Then we may express the least squares estimates before and after a candidate change point location t as

$$\hat{\boldsymbol{\beta}}_{t,1} = (\mathbf{X}_{t,1}^\top \mathbf{X}_{t,1})^{-1} \mathbf{X}_{t,1}^\top \mathbf{Y}_{t,1} \quad \text{and} \quad \hat{\boldsymbol{\beta}}_{t,2} = (\mathbf{X}_{t,2}^\top \mathbf{X}_{t,2})^{-1} \mathbf{X}_{t,2}^\top \mathbf{Y}_{t,2}.$$

Let $\mathbf{A}_i = \{a_i(k, \ell), 1 \le k, \ell \le d\}$, where

$$\lim_{T\to\infty} \frac{1}{t_i - t_{i-1}} \sum_{t=t_{i-1}+1}^{t_i} E x_t(k) x_t(\ell) = a_i(k, \ell), \quad 1 \le k, \ell \le d, 1 \le i \le M+1,$$

$\mathbf{x}_t = (x_t(1), x_t(2), \dots, x_t(d))^\top$. Below we make use of the following assumptions in order to quantify the asymptotic behavior of the test statistics proposed below. These basically entail and imply that the change point(s) in the parameters and volatility of the regressors must be well separated, and that the partial sample estimators of the regression parameters must satisfy a functional central limit theorem with a quantifiable rate of convergence.

Assumption 1. The constants θ_i in (2) satisfy $0 < \theta_1 < \theta_2 < \dots < \theta_M < 1$.

Assumption 2. The matrices $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_{M+1}$ are nonsingular.

Let $\|\cdot\|$ denote the Euclidean norm of vectors and matrices. The following assumption implies that the covariates and errors in (3) and (4) are generally weakly dependent with at least four finite moments.

Assumption 3. f_j and $\mathbf{g}_j$, $1 \le j \le M+1$ are nonrandom functionals defined on $\mathcal{S}^\infty$ with values in $\mathbb{R}$ and $\mathbb{R}^d$, with $\mathcal{S}$ being a measurable space. The sequences $\epsilon_i, \mathbf{x}_i, -\infty < i < \infty$ can be approximated with m-dependent sequences $\epsilon_{i,m}$ and $\mathbf{x}_{i,m}$ in the sense that with some $\kappa_1 > 4$, $\kappa_2 > 2$ and $c > 0$, $E|\epsilon_i|^{\kappa_1} < \infty$, $E\|\mathbf{x}_i\|^{\kappa_1} < \infty$,

$$(E|\epsilon_i - \epsilon_{i,m}|^{\kappa_1})^{1/\kappa_1} \le cm^{-\kappa_2}, \tag{5}$$

and

$$(E\|\mathbf{x}_i - \mathbf{x}_{i,m}\|^{\kappa_1})^{1/\kappa_1} \le cm^{-\kappa_2}, \tag{6}$$

where $\epsilon_{i,m} = f_j(\eta_i, \eta_{i-1}, \eta_{i-2}, \dots, \eta_{i-m+1}, \boldsymbol{\eta}^*_{i,m})$, $\mathbf{x}_{i,m} = \mathbf{g}_j(\eta_i, \eta_{i-1}, \eta_{i-2}, \dots, \eta_{i-m+1}, \boldsymbol{\eta}^*_{i,m})$, $t_{j-1} < i \le t_j$, $1 \le j \le M+1$, $\boldsymbol{\eta}^*_{i,m} = (\eta^*_{i,m,i-m}, \eta^*_{i,m,i-m-1}, \eta^*_{i,m,i-m-2}, \dots)$ and the $\eta^*_{i,m,n}$'s are independent copies of η_0, independent of $\{\eta_i, -\infty < i < \infty\}$.

This structural dependence condition encompasses the vast majority of time series models that are driven by independent identically distributed innovations sequences, under suitable conditions on the models implying the existence of stationary causal solutions. For example, Assumption 3 holds for series following ARMA and GARCH models under mild conditions. A thorough comparison of such dependence conditions with related mixing conditions can be found in Wu (2005).

Let

$$\mathbf{e}_t = \epsilon_t \mathbf{x}_t = (x_t(1)\epsilon_t, x_t(2)\epsilon_t, \dots, x_t(d)\epsilon_t)^\top,$$

and define the corresponding long run variances on the intervals of stationarity by

$$\mathbf{D}_i = \lim_{T\to\infty} \frac{1}{(t_i - t_{i-1})} E\left[\left(\sum_{t=t_{i-1}+1}^{t_i} \mathbf{e}_t\right)\left(\sum_{t=t_{i-1}+1}^{t_i} \mathbf{e}_t\right)^\top\right], \quad 1 \le i \le M+1.$$

Assumption 4. $\mathbf{D}_1$ and $\mathbf{D}_{M+1}$ are nonsingular matrices.

The requirement that only the matrices $\mathbf{D}_1$ and $\mathbf{D}_{M+1}$ are nonsingular arises due to the method that we propose below to estimate normalizing matrices for $\hat{\boldsymbol{\beta}}_{t,1}$ and $\hat{\boldsymbol{\beta}}_{t,2}$, which involve estimating them based on the data before and after candidate break points. The resulting estimates may then be considered as convex combinations of estimators that necessarily contain estimates for $\mathbf{D}_1$ and $\mathbf{D}_{M+1}$.

The long run covariance matrix of the normalized sum of the $\mathbf{e}_t$'s is time dependent so we need a time-dependent estimator as well. First we define the residuals as

$$\hat{\epsilon}_t = y_t - \mathbf{x}_t^\top \hat{\boldsymbol{\beta}}_{T,1}, \quad 1 \le t \le T.$$

The corresponding estimator for $\mathbf{e}_t = \epsilon_t \mathbf{x}_t$ is

$$\hat{\mathbf{e}}_t = \hat{\epsilon}_t \mathbf{x}_t = (x_t(1)\hat{\epsilon}_t, x_t(2)\hat{\epsilon}_t, \dots, x_t(d)\hat{\epsilon}_t)^\top.$$

The estimators for the long-run covariance matrices starting from the beginning and the end of the data we define as

$$\hat{\mathbf{Q}}_t(1) = \frac{1}{t}\sum_{s=1}^{t} \hat{\mathbf{e}}_s \hat{\mathbf{e}}_s^\top + \sum_{u=1}^{t-1} K\left(\frac{u}{h_t}\right) \frac{1}{t-u}\left(\sum_{s=1}^{t-u} \hat{\mathbf{e}}_s \hat{\mathbf{e}}_{s+u}^\top + \sum_{s=1}^{t-u} \hat{\mathbf{e}}_{s+u} \hat{\mathbf{e}}_s^\top\right),$$

and

$$\hat{\mathbf{Q}}_t(2) = \frac{1}{T-t}\sum_{s=t+1}^{T} \hat{\mathbf{e}}_s \hat{\mathbf{e}}_s^\top + \sum_{u=1}^{T-t-1} K\left(\frac{u}{h_{T-t}}\right) \frac{1}{T-t+u}\left(\sum_{s=t+1}^{T-u} \hat{\mathbf{e}}_s \hat{\mathbf{e}}_{s+u}^\top + \sum_{s=t+1}^{T-u} \hat{\mathbf{e}}_{s+u} \hat{\mathbf{e}}_s^\top\right).$$

The test statistic we propose is the maximized quadratic form

$$\hat{Z}_T = \max_{a_T \le t \le T-b_T} \left((\hat{\boldsymbol{\beta}}_{t,1} - \hat{\boldsymbol{\beta}}_{t,2})^\top \hat{\mathbf{Q}}_{T,t}^{-1} (\hat{\boldsymbol{\beta}}_{t,1} - \hat{\boldsymbol{\beta}}_{t,2})\right)^{1/2}, \tag{7}$$

where

$$\hat{\mathbf{Q}}_{T,t} = \left(\frac{\mathbf{X}_{t,1}^\top \mathbf{X}_{t,1}}{t}\right)^{-1} \hat{\mathbf{Q}}_t(1) \left(\frac{\mathbf{X}_{t,1}^\top \mathbf{X}_{t,1}}{t}\right)^{-1} \mathbb{1}\{t \le T/2\} + \left(\frac{\mathbf{X}_{t,2}^\top \mathbf{X}_{t,2}}{T-t}\right)^{-1} \hat{\mathbf{Q}}_t(2) \mathbb{1}\{t > T/2\} \left(\frac{\mathbf{X}_{t,2}^\top \mathbf{X}_{t,2}}{T-t}\right)^{-1}.$$

and $\mathbb{1}$ denotes the indicator function. It straightforward to see that

$$\hat{\mathbf{Q}}_{T,t}^{-1} = \left(\frac{\mathbf{X}_{t,1}^\top \mathbf{X}_{t,1}}{t}\right) \hat{\mathbf{Q}}_t(1)^{-1} \left(\frac{\mathbf{X}_{t,1}^\top \mathbf{X}_{t,1}}{t}\right) \mathbb{1}\{t \le T/2\} + \left(\frac{\mathbf{X}_{t,2}^\top \mathbf{X}_{t,2}}{T-t}\right) \hat{\mathbf{Q}}_t(2)^{-1} \left(\frac{\mathbf{X}_{t,2}^\top \mathbf{X}_{t,2}}{T-t}\right) \mathbb{1}\{t > T/2\}.$$

The motivation to use $\hat{\mathbf{Q}}_{T,t}^{-1}$ in the normalization defining $\hat{Z}_T$ is to improve power when the change point might occur near the end of the sample, and guard against conditional heteroscedasticity: we use $\hat{\mathbf{Q}}_t(1)$ when evaluating for potential early changes, and $\hat{\mathbf{Q}}_t(2)$ in evaluating for changepoints that might be near the end of the sample. In order that these estimates are consistent, we also make the following two standard assumptions on the kernel function and bandwidth:

Assumption 5. (i) $K(0)=1$ (ii) $K(u)=0$ if $|u| \ge c$ with some $c>0$ (iii) K is Lipschitz continuous on the real line

Assumption 6. (i) $h_t \to \infty$ and $h_t = O(t^{1/2}(\log t)^{-(3+\zeta)})$ with some $\eta > 0$, as $t \to \infty$ (ii) $h_t = h_{\rho^{i-1}}, \rho^{i-1} \le t < \rho^i, 1 \le i < \infty$ with some $\rho > 1$.

The assumption that the kernel function has bounded support is made mainly to simplify the proofs, and could be replaced with the condition that it is square integrable. A form of Assumption 6(ii) already appeared in Berkes et al. (2005, 2006), where uniform convergence of the Bartlett estimator of the long run covariance is studied. Our final assumption describes the strength/size of the trimming of the CUSUM in defining $\hat{Z}_T$. One feature that separates this work from the past literature is that we allow the trimming parameters a_T and b_T to be small, that is, of lower order than the sample size T, in order to improve performance for end of sample changes.

Assumption 7. $a_T \to \infty, b_T \to \infty, a_T/T \to 0$ and $b_T/T \to 0$.

Let

$$r_T = \min(a_T, b_T),$$

and define

$$\lim_{T\to\infty} \frac{r_T}{a_T} = \gamma_1, \qquad \lim_{T\to\infty} \frac{r_T}{b_T} = \gamma_2.$$

The standard Wiener process in $\mathbb{R}^d$ is denoted by $\mathbf{W}$, i.e. $\mathbf{W}$ is a Gaussian process with $E\mathbf{W}(t) = \mathbf{0}$ and $E\mathbf{W}(t)\mathbf{W}(s)^\top$ is a diagonal matrix with $\min(t, s)$ in the diagonal. Let

$$\zeta = \max_{0\le u\le 1} \|\mathbf{W}(u)\|,$$

and define

$$\xi = \max\left(\gamma_1^{1/2}\zeta_1, \gamma_2^{1/2}\zeta_2\right),$$

where ζ_1, ζ_2 are independent random variables, distributed as ζ.

Theorem 1. *If H_0 holds and Assumptions 1, 2, 3, 4, 5, 6, and 7 are satisfied, then we have that*

$$r_T^{1/2}\hat{Z}_T \xrightarrow{\mathcal{D}} \xi.$$

Remark 1. If the weighted errors $\{\mathbf{e}_s, s \ge 0\}$ are uncorrelated, that is, $E\mathbf{e}_s\mathbf{e}_u^\top$ is the zero matrix, then we only need to keep the first terms in the definitions $\hat{\mathbf{Q}}_t(1)$ and $\hat{\mathbf{Q}}_t(2)$. For example, in many popular volatility models ϵ_t is independent of $\{\epsilon_s, s < t, \mathbf{x}_u, u \le t\}$, and the vectors $\{\mathbf{e}_s, s \ge 0\}$ are uncorrelated. Let

$$\overline{\mathbf{Q}}_t(1) = \frac{1}{t}\sum_{s=1}^{t} \hat{\mathbf{e}}_s\hat{\mathbf{e}}_s^\top, \quad \overline{\mathbf{Q}}_t(2) = \frac{1}{T-t}\sum_{s=t+1}^{T} \hat{\mathbf{e}}_s\hat{\mathbf{e}}_s^\top,$$

and

$$\overline{\mathbf{Q}}_{T,t} = \left(\frac{\mathbf{X}_{t,1}^\top\mathbf{X}_{t,1}}{t}\right)^{-1} \overline{\mathbf{Q}}_t(1)\left(\frac{\mathbf{X}_{t,1}^\top\mathbf{X}_{t,1}}{t}\right)^{-1} \mathbb{1}\{t \le T/2\} + \left(\frac{\mathbf{X}_{t,2}^\top\mathbf{X}_{t,2}}{T-t}\right)^{-1} \overline{\mathbf{Q}}_t(2)\left(\frac{\mathbf{X}_{t,2}^\top\mathbf{X}_{t,2}}{T-t}\right)^{-1} \mathbb{1}\{t > T/2\}.$$

For uncorrelated $\mathbf{e}_t$'s the test statistic is

$$\overline{Z}_T = \max_{a_T \le t \le T-b_T} \left((\hat{\boldsymbol{\beta}}_{t,1} - \hat{\boldsymbol{\beta}}_{t,2})^\top \overline{\mathbf{Q}}_{T,t}^{-1}(\hat{\boldsymbol{\beta}}_{t,1} - \hat{\boldsymbol{\beta}}_{t,2})\right)^{1/2},$$

and under H_0 and Assumptions 1, 2, 3, 4, and 7, we have

$$r_T^{1/2}\overline{Z}_T \quad \overset{\mathcal{D}}{\to} \quad \xi. \tag{8}$$

These results motivate one to reject H_0 when $r_T^{1/2}\hat{Z}_T$ exceeds the $1-\alpha$ critical value of ξ, which according to Theorem 1 is an asymptotically size α test of H_0.

Next we consider the consistency of our testing procedure under the alternative of exactly one change in the $\boldsymbol{\beta}$'s. Let $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(T)}$ denote the regression parameter before and after the change. To compare our result with existing literature, we consider the case when the change is not too late (cf. Assumption 9 in Theorem 2). Not too early changes can be handled similarly.

Theorem 2. *If Assumptions 1, 2, 3, 4, 5, 6, and 7 are satisfied,*

$$\limsup_{T\to\infty}\frac{t^*}{T} < 1, \tag{9}$$

$$\limsup_{T\to\infty}\frac{a_T}{t^*} < \infty, \tag{10}$$

and

$$r_T^{1/2}\min\left(\frac{t^*}{a_T},1\right)\|\boldsymbol{\beta}^{(1)}-\boldsymbol{\beta}^{(T)}\| \to \infty, \quad \text{as} \quad T\to\infty, \tag{11}$$

then we have that

$$r_T^{1/2}\hat{Z}_T \quad \overset{P}{\to} \quad \infty. \tag{12}$$

If $a_T \le t^*$, then we only require that the size of the change is larger than $r_T^{-1/2}$. Hence this test is expected to be able to detect relatively small changes, even if $\|\boldsymbol{\beta}^{(1)}-\boldsymbol{\beta}^{(T)}\| \to 0$. The residual-based approaches require additional conditions on the change $\boldsymbol{\beta}^{(1)}-\boldsymbol{\beta}^{(T)}$. For example, the Rényi-type statistic in Horváth et al. (2020) will not detect a change if $\boldsymbol{\beta}^{(1)}-\boldsymbol{\beta}^{(T)}$ and $\sum_{t=1}^{T}\mathbf{x}_t$ are orthogonal vectors. Further these results hold under the above general heteroscedasticity assumptions.

Remark 2. If the errors are uncorrelated, Assumptions 1, 2, 3, 4, 7 and (9), (10), (11) hold, then we also have consistency of the test statistic $\overline{Z}_T$, in particular we have that

$$r_T^{1/2}\overline{Z}_T \quad \overset{P}{\to} \quad \infty. \tag{13}$$

Since the proof of (13) goes along the lines of (12), it is omitted.

Remark 3. The test described above is consistent to detect the single change point alternative H_A. It is beyond the scope of the present work to describe the consistency of the proposed test statistic for detecting and estimating multiple change points; however, in the event that one is interested in doing so, we recommend applying binary segmentation. A single change point may be consistently estimated as

$$\hat{t}_T = \min\left\{t\in[a_t, T-b_T] : \hat{Z}_T = \left((\hat{\boldsymbol{\beta}}_{t,1}-\hat{\boldsymbol{\beta}}_{t,2})^\top \hat{\mathbf{Q}}_{T,t}^{-1}(\hat{\boldsymbol{\beta}}_{t,1}-\hat{\boldsymbol{\beta}}_{t,2})\right)^{1/2}\right\}.$$

Based on this estimate, the sample may be partitioned into two subsamples with indices $\{1, \dots, \hat{t}_T\}$ and $\{\hat{t}_T + 1, \dots, T\}$, and then one may apply the same detection and estimation procedure for a single change point to each subsample. An effective stopping criterion for this procedure applied to any particular interval is when one is not able to reject H_0 at a specified significance level. We illustrate the use of this procedure in the data example below.

Remark 4. In practice the user must select the trimming parameters a_T and b_T. Local power calculations presented below and in the proof of Theorem 2 suggest that the power is maximized when a_T and b_T are as large as possible so that $\{a_T, \dots, T - b_T\}$ still includes the potential change point(s). We compare a number of such choices in the simulation study below, ranging from $a_T = b_T = \log(T)$ to $a_T = b_T = T^{1/2}$, see Figure 4. Generally we suggest $a_T = b_T = T^{1/2}$ as a default setting.

Remark 5. We provide here a brief comparison in terms of local power of the proposed statistic to standard CUSUM statistics. The standard maximally selected CUSUM process analogous to the statistic $\hat{Z}_T$ is

$$P_T = \max_{d<t<T-d} \frac{t(T-t)}{T^{3/2}} \left((\hat{\boldsymbol{\beta}}_{t,1} - \hat{\boldsymbol{\beta}}_{t,2})^\top \overline{\mathbf{Q}}_{T,t}^{-1} (\hat{\boldsymbol{\beta}}_{t,1} - \hat{\boldsymbol{\beta}}_{t,2}) \right)^{1/2}. \tag{14}$$

It can be shown that under the conditions of Theorem 1 that

$$P_T \xrightarrow{\mathcal{D}} \sup_{0<t<1} \|\boldsymbol{\Delta}(t)\|,$$

where $\boldsymbol{\Delta}$ is a continuous Gaussian process satisfying $\boldsymbol{\Delta}(0) = \mathbf{0}$ and $\boldsymbol{\Delta}(1) = \mathbf{0}$. Under the conditions of Theorem 2,

$$P\left\{ \limsup_{T\to\infty} \max_{d<t<T-d} \|\overline{\mathbf{Q}}_{T,t}^{-1}\| < \infty \right\} = 1$$

and there is a positive definite matrix $\mathbf{Q}$ such that

$$\overline{\mathbf{Q}}_{T,t^*} \xrightarrow{P} \mathbf{Q}.$$

The difference between parameter estimates at a candidate change point t can be decomposed as

$$\hat{\boldsymbol{\beta}}_{t,1} - \hat{\boldsymbol{\beta}}_{t,2} = \mathcal{E}_{T,t} + \mathbf{z}_{t,T},$$

where $\mathbf{z}_{t,T}$ is a nonrandom drift term. Standard arguments give that

$$\mathbf{z}_{t,T} \approx \begin{cases} \frac{T-t^*}{T-t}(\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(T)}), & \text{if } d < t \le t^* \\ \frac{t^*}{t}(\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(T)}), & \text{if } t^* < t < T-d, \end{cases}$$

and further the error process $\mathcal{E}_{T,t}$ is the same under H_0 and H_A. The function $t(T-t)\,\|\mathbf{z}_{t,T}\|$ will hence reach its largest value around t^*. An early change point may be defined asymptotically as

$$\frac{t^*}{T} \to 0.$$

We define a local alternative as

$$\frac{t^*}{T^{1/2}}\|\boldsymbol{\beta}^{(1)}-\boldsymbol{\beta}^{(T)}\| \to \mathfrak{a} \in [0,\infty], \quad \text{as} \quad T \to \infty.$$

If $\mathfrak{a}=0$, then (14) still holds. If $0<\mathfrak{a}<\infty$, then on account of $\boldsymbol{\Delta}(0)=\mathbf{0}$ it can be shown that there is a constant $0<\mathfrak{a}^*<\infty$ such that $P_T \to \mathfrak{a}^*$ in probability and $P_T \to \infty$ in probability, if $\mathfrak{a}=\infty$. For the sake of simplicity, we assume in the definition of $\hat{Z}_T$ that $a_T=b_T=r_T$ and

$$\limsup_{T\to\infty}\frac{r_T}{t^*} \le 1, \tag{15}$$

that is, the early change comes after trimming parameter r_T. We observe similarly that the drift $\|\mathbf{z}_{T,t}\|$ reaches its largest value around t^* again. We consider then the analogous local alternative for $\hat{Z}_t$

$$r_T\|\boldsymbol{\beta}^{(1)}-\boldsymbol{\beta}^{(T)}\| \to \mathfrak{b} \in [0,\infty], \quad \text{as} \quad T \to \infty.$$

If $\mathfrak{b}=0$, then $a_T^{1/2}\hat{Z}_T$ still satisfies the limit result in Theorem 1. If $0<\mathfrak{b}<\infty$, then $a_T^{1/2}\hat{Z}_T$ converges in distribution to a nondegenerate random variable, differing from the limit in Theorem 1. In case of $\mathfrak{b}=\infty$, $a_T^{1/2}\hat{Z}_T \to \infty$ in probability. These then constitute the necessary and sufficient condition for consistency of $\hat{Z}_T$ under (15). In summary, $\hat{Z}_T$ can be expected to have superior power against early and small changes when compared to P_T, as the local alternative causing the normalized $\hat{Z}_T$ to be asymptotically constant forces $\|\boldsymbol{\beta}^{(1)}-\boldsymbol{\beta}^{(T)}\| \approx \mathfrak{b}r_T^{-1}$, in which case t^* must be of the order $T^{1/2}r_T >> r_T$ in order for P_T to be asymptotically constant.

2 | SIMULATIONS STUDY

We now present the results of a simulation study in which we aimed to both evaluate the finite sample properties of the test based on $\hat{Z}_T$ in both homoscedastic and heteroscedastic environments, and compare this test to a number of existing procedures to detect structural changes in linear regression models. In all simulations below we took the covariate dimension $d=2$ in (1). The data is generated as

$$y_t = \begin{cases} \beta_1^{(1)} + x_{t,2}\beta_2^{(1)} + \epsilon_t, & 1 \le t \le t^* \\ \beta_1^{(T)} + x_{t,2}\beta_2^{(T)} + \epsilon_t, & t^*+1 \le t \le T, \end{cases} \tag{16}$$

We chose $\boldsymbol{\beta}^{(1)}=(1,2)^\top$ and $\boldsymbol{\beta}^{(T)}=\boldsymbol{\beta}^{(1)}+\delta(1,-1)^\top$. We considered the values of $\delta \in [-2,2]$ (incrementing by 0.1), and with this setup $\delta=0$ gives H_0, while $\delta \ne 0$ gives alternatives of various strengths satisfying H_A. The data vector is $\mathbf{x}_t=(1,x_{t,2})$, and the $x_{t,2}$'s are taken to be independent identically distributed $N(1,1)$ random variables. The tests that we consider include the proposed test based on $\hat{Z}_T$, along with the Rényi-type statistic of Horváth et al. (2020) with trimming parameter $a_T=b_T=T^{1/2}$ computed from the residuals, the CUSUM statistic of Ploberger and Krämer (1992), Andrews (1993) Wald-type statistic, and the Hidalgo and Seo (2013) statistic. The Rényi-type statistic of Horváth et al. (2020) and the CUSUM statistic of Ploberger and Krämer (1992) are the univariate statistics computed from the estimated residuals of the regression model. The Andrews and Hidalgo–Seo statistics were tailored to the regression context as described in their papers, and are based on functionals of the process of estimated regression coefficients $\hat{\boldsymbol{\beta}}_{t,1} - \hat{\boldsymbol{\beta}}_{t,2}$, similar to the proposed test. For the Andrews (1993) Wald-type statistic, we chose to trim 10% of the sample at both ends, and p-values were computed using an empirical CDF of the limiting distribution of the statistic estimated from 1,000,000 realizations.

We note that these statistics do not all use the same long-run variance (LRV) estimation procedure. The Hidalgo–Seo statistic uses the periodogram matrix for LRV estimation. The statistics of the present paper, the Rényi-type statistics of Horváth et al. (2020) and the CUSUM statistic of Ploberger and Ploberger and Krämer (1992) use either the LRV estimation with kernel methods (correlated errors) or the sample variance (uncorrelated data). If LRV estimation is needed we use the quadratic spectral kernel, and the bandwidth was selected using the procedure advocated by Andrews (1991) and implemented in Aschersleben and Wagner (2016). Andrews (1993) statistic uses the procedure advocated in their paper allowing for correlated residuals. Notably in this case the LRV estimator is consistent only under the null hypothesis of no change. The varying bandwidths h_t are set to $h_t = \hat{m}_T t^{1/5}$, where the constant $\hat{m}_T$ is estimated from the whole sample using the method of Andrews (1991).

2.1 | Early change

We start by assuming that the change in the regression parameter occurs early. We concentrate on the case when the change is relatively close to the beginning of the sample, and take $t^* = \lfloor T^{3/5} \rfloor$.

We consider a number of settings for ϵ_t that follow both homoscedastic and heteroscedastic regimes. We use the following four data generating processes in the homoscedastic case: (HO-IID) independent and identically distributed $N(0, 1)$ random variables;

(HO-AR)$\{\epsilon_t\}$ is an AR(1) process with autocorrelation coefficient $\phi = 0.5$ and the innovations are i.i.d. normal random variables with zero mean and variance 0.5;

(HO-ARMA) $\{\epsilon_t\}$ is an ARMA(2, 2) process $\epsilon_t = \phi_1\epsilon_{t-1} + \phi_2\epsilon_{t-2} + \eta_t + \theta_1\eta_{t-1} + \theta_2\eta_{t-2}$ with $\phi_1 = 0.4$, $\phi_2 = -0.03$, $\theta_1 = 0.5$, $\theta_2 = 0.06$, and the variance of the i.i.d. mean zero normal noise process is chosen so that variance of ϵ_t is 1;

(HO-GARCH) GARCH(1, 1) process $\epsilon_t = \sigma_t\eta_t$, $\sigma_t^2 = \omega + \alpha\eta_{t-1}^2 + \beta\sigma_{t-1}^2$ with $\omega = 0.3$, $\alpha = 0.2$, and $\beta = 0.5$ and the innovations are i.i.d. standard normal random variables.

If ϵ_t is heteroscedastic, then we consider a couple scenarios. First, we take ϵ_t to follow one of the below four DGPs in the second half of the sample, while the sequence in the first half is generated according to the corresponding above homoscedastic DGP. (HE1-IID) after $t_1 = \lfloor T/2 \rfloor$ the errors are independent and identically distributed $N(0, 10)$ random variables;

(HE1-AR) in the second half of the sample the errors follow an AR(1) sequence with $\phi = 0.5$ and the innovations are i.i.d. normal random variables with zero mean and variance 5;

(HE1-ARMA) for the ARMA(2, 2) case, after $t_1 = \lfloor T/2 \rfloor$ the ϵ_t's have parameters $\phi_1 = 0.4$, $\phi_2 = 0.2$, $\theta_1 = -0.1$, $\theta_2 = -0.42$ and the variance of the i.i.d. mean zero normal noise process is chosen so that variance of ϵ_t is 10;

(HE1-GARCH) in the GARCH(1, 1) case, in the second half the errors have parameters $\omega = 1$, $\alpha = 0.7$, $\beta = 0.2$. In all cases, the variance of the errors changes from 1 to 10 in the middle of the data.

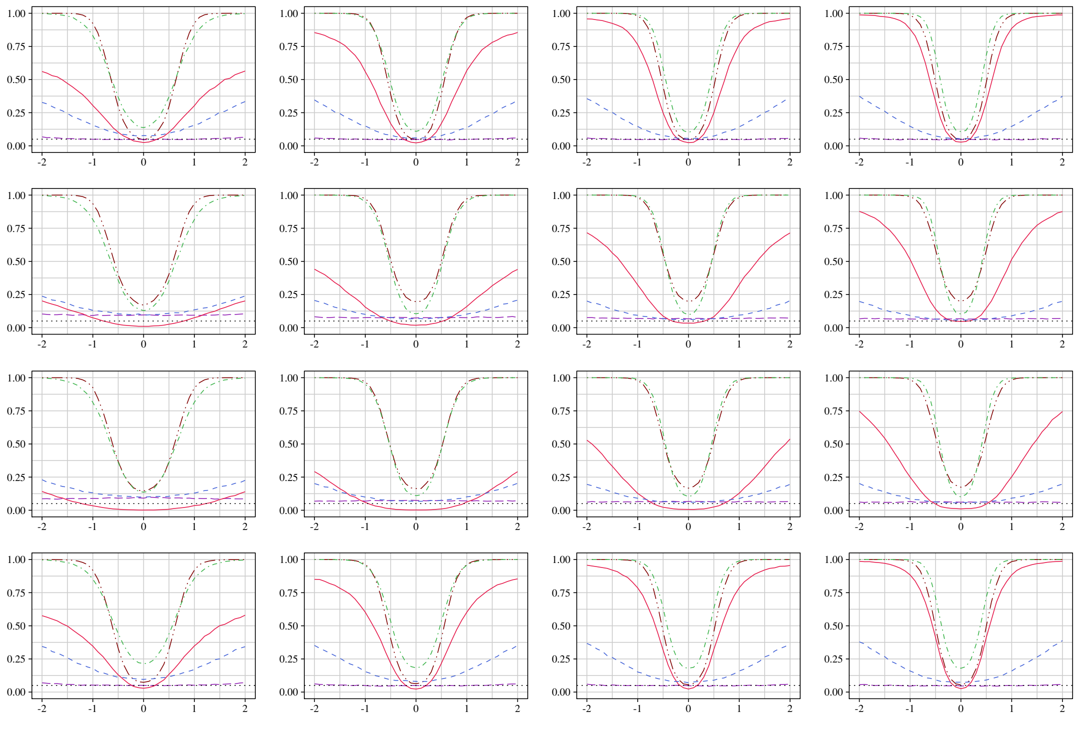


FIGURE 1 Power plots as a function of δ with nominal size $\alpha = 0.05$ for homoscedastic data when the change in the regression parameters occurs at $t^* = \lfloor T^{3/5} \rfloor$; columns from left to right correspond to the sample sizes 250, 500, 750, and 1000 respectively, the first row is (HO-IID), the second row is (HO-AR), the third row is (HO-ARMA), and the fourth row is (HO-GARCH). The power curves are displayed for the proposed method $\hat{Z}_T$ (——), Horváth et al. (2020) (_ _ _), Ploberger and Krämer (1992) (_ _ _), Andrews (1993) (_ . . _), Hidalgo and Seo (2013) (_ . _) [Colour figure can be viewed at wileyonlinelibrary.com]

The second heteroscedasticity scenario only changes the location of the change in variance, so that the latter three-quarters of the sample has a different variance than the first, so $t_1 = \lfloor T/4 \rfloor$. The magnitude and structure of the change is otherwise the same. In this scenario we care only to study the effect of the location of the variance change, so we restrict ourselves to the Normal DGP, labeled (HE2-IID).

For every combination of sample size, test statistic, δ, and data-generating process for ϵ_t, 20,000 independent tests of each type considered were conducted, and their approximate p-values were computed. H_0 is rejected for p-values less than $\alpha = 0.05$. The sample sizes considered were $T \in \{250, 500, 750, 1000\}$. These results are presented in terms of power plots in Figures 1 and 2.

These results can be summarized as follows. In the case of homoscedastic residuals, the best-performing test statistic in all plots in terms of power is either the Hidalgo–Seo statistic or Andrews' statistic. The Hidalgo-Seo statistic displayed considerable size inflation in many cases even for large sample sizes, while Andrews' statistic exhibited size inflation less frequently, although still often did. The residual-based statistics of Ploberger and Krämer (1992) and Horváth et al. (2020) did not demonstrate good power properties, though they generally exhibited good size. The CUSUM statistic is the worst performing while the univariate Rényi-type statistic of Horváth et al. (2020) sometimes exhibited decent power for large sample sizes. On balance, though, the statistic $\hat{Z}_T$ of (7) often exhibited strong power without suffering from size inflation,

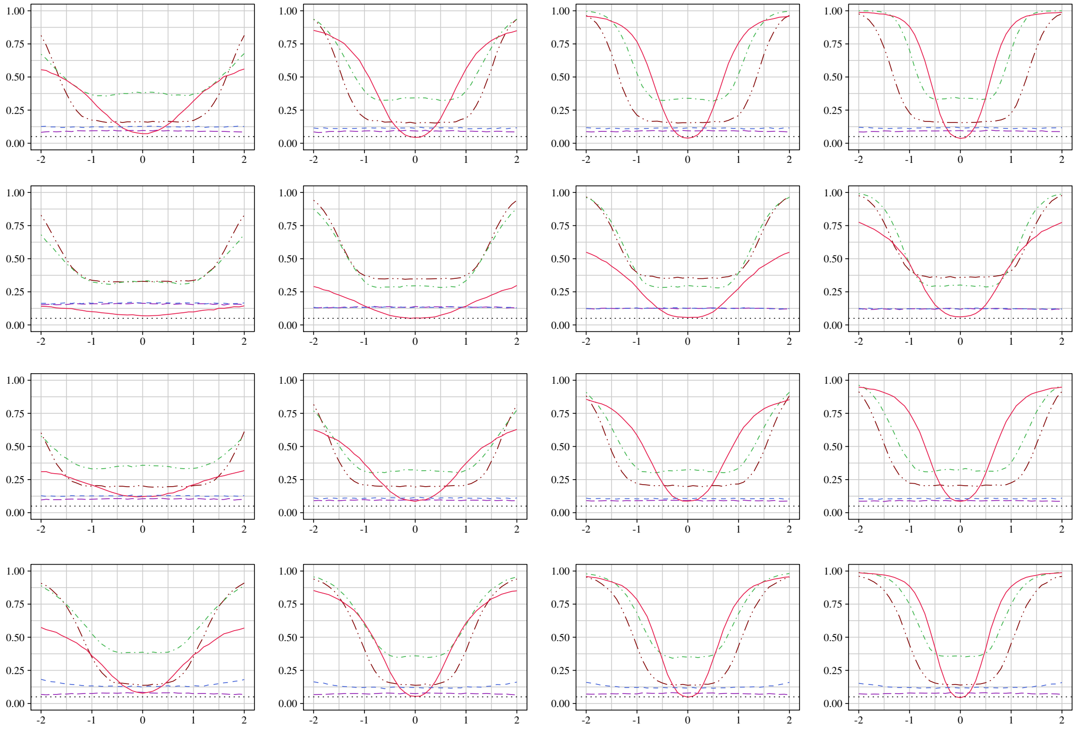


FIGURE 2 Power plots as a function of δ with nominal size $\alpha = 0.05$ for heteroscedastic data when the change in the regression parameters occurs at $t^* = \lfloor T^{3/5} \rfloor$; columns correspond to the sample sizes 250, 500, 750, and 1000 respectively, the first row is (HE1-IID), the second row is (HE1-AR), the third row is (HE1-ARMA), and the fourth row is (HE1-GARCH). The power curves are displayed for the proposed method $\hat{Z}_T$ (___), Horváth et al. (2020) (_ _ _), Ploberger and Krämer (1992) (_ _ _), Andrews (1993) (_ . . _), and Hidalgo and Seo (2013) (_ . _) [Colour figure can be viewed at wileyonlinelibrary.com]

in fact generally we observed that the test based on $\hat{Z}_T$ was slightly undersized. Although the statistic $\hat{Z}_T$ was designed for heteroscedastic data, it performs reasonably well in homoscedastic environments.

In the case of heteroscedastic errors, the residual-based statistics of Ploberger and Krämer (1992) and Horváth et al. (2020) effectively exhibited no power for our choices of parameters. The Hidalgo–Seo statistic was markedly oversized for heteroscedastic data for all sample sizes considered. Andrews' statistic did not display size problems to the same degree as the Hidalgo–Seo statistic, but still exhibited considerable size inflation. Further the power of each of these statistics was markedly reduced when compared to the homoscedastic settings. The tests based on the proposed statistic $\hat{Z}_T$ though exhibited good size, and similar power profiles as in the homoscedastic case, sometimes even beating the ill–sized Hidalgo–Seo and Andrews' statistics in terms of power. Figure 3 suggests, though, that the size problems experience by the Hidalgo-Seo and Andrews statistics could be due to the location of the change in variance, since causing the variance change to happen much earlier eliminates the size problems experienced by these statistics. However, they still do not attain the same level of power as observed in $\hat{Z}_T$.

The trimming parameter for $\hat{Z}_T$ used in the above simulations was $a_T = b_T = T^{1/2}$. We also explored alternative trimming parameters and show simulations results for those comparisons in Figure 4. The trimming parameter furthest from the location of the change point is $\log(T)$, the

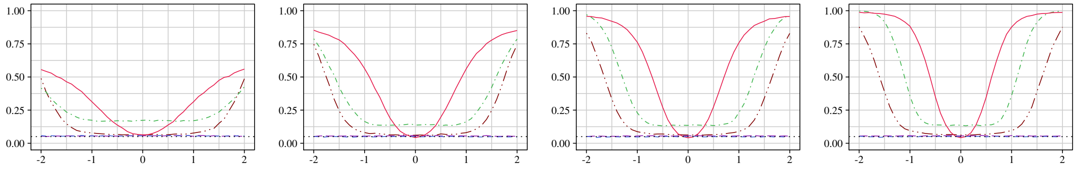


FIGURE 3 Power plots as a function of δ with nominal size $\alpha = 0.05$ for heteroscedastic data following process (HE2-IID) when the change in the regression parameters occurs at $t^* = \lfloor T^{3/5} \rfloor$; columns correspond to the sample sizes 250, 500, 750, and 1000, respectively. The power curves are displayed for the proposed method $\hat{Z}_T$ (___), Horváth et al. (2020) (_ _ _), Ploberger and Krämer (1992) (_ _ _), Andrews (1993) (_ . . _), and Hidalgo and Seo (2013) (_ . _) [Colour figure can be viewed at wileyonlinelibrary.com]

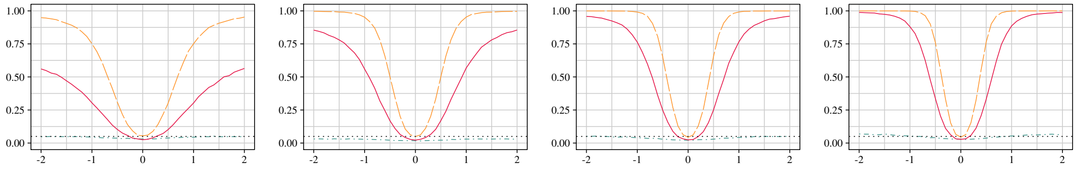


FIGURE 4 Power plots as a function of δ with nominal size $\alpha = 0.05$ for heteroscedastic data following process (HO-IID) when the change in the regression parameters occurs at $t^* = \lfloor T^{3/5} \rfloor$; columns from left to right correspond to the sample sizes 250, 500, 750, and 1000, respectively. The power curves are displayed for the proposed method when the trimming parameter is $T^{1/2}$ (___), $T^{3/5}$ (__ __), and $\log(T)$ (_ . _) [Colour figure can be viewed at wileyonlinelibrary.com]

closest is $T^{3/5}$, which coincides with the location of the change point, and the parameter generally used in this paper is $T^{1/2}$. We see that power increases as the trimming parameter gets closer to the location of the change point, as expected. Indeed, in this case the power of the proposed test exceeds that of the Andrews' statistic even in this homoscedastic setting. Taking the trimming parameter to be very small, such as $a_T = \log(T)$, results in markedly reduced power, as expected. For this reason we generally recommend $a_T = b_T = T^{1/2}$ as a default setting.

2.2 | Mid-sample change

Although the proposed statistic is tailored to end-of-sample change points, it is worthwhile to study it empirical properties for mid-sample changes as well.

We present simulation results taking $t^* = T/2$, that is, the change in linear model coefficients happens in the middle of the sample. We consider only (HO-IID) and (HE1-IID) DGPs, as we are investigating here only the effect of the location of the change point on the behavior of the investigated statistics. Note that when the DGP is (HE1-IID), the location of the variance change and the location of the potential change point equal each other. We plot our results in Figure 5.

In the homoscedastic case Andrews' statistic performed the best, having excellent size and power. $\hat{Z}_T$ also retained good power and size properties in this context. The Hidalgo–Seo statistic also has good power but exhibited some size inflation as before. In the heteroscedastic case, the Andrews and Hidalgo–Seo statistics still exhibited marked size inflation. In none of these settings did the residual-based methods perform well.

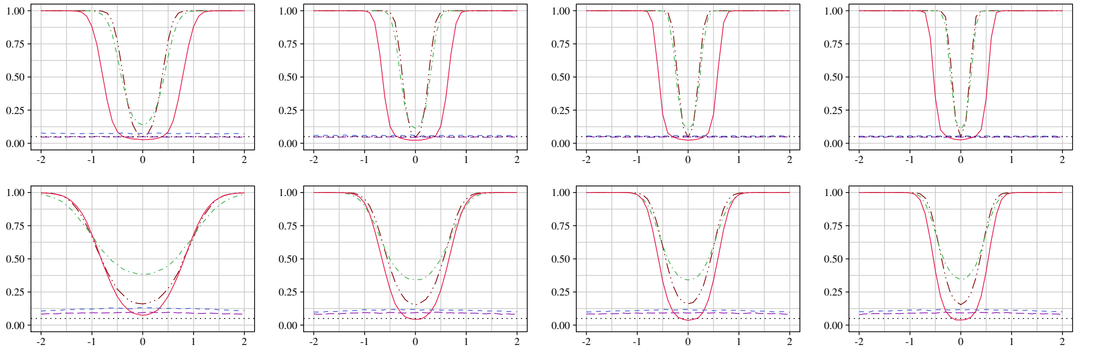


FIGURE 5 Power plots as a function of δ with nominal size $\alpha = 0.05$ when the change in the regression parameters occurs at $t^* = \lfloor T/2 \rfloor$; columns from left to right correspond to the sample sizes 250, 500, 750, and 1000 respectively, the first row is (HO-IID) and the second row is (HE1-IID). The power curves are displayed for the proposed method $\hat{Z}_T$ (——), Horváth et al. (2020) (_ _ _), Ploberger and Krämer (1992) (_ _ _) Andrews (1993) (_ . . _), and Hidalgo and Seo (2013) (_ . _) [Colour figure can be viewed at wileyonlinelibrary.com]

3 | DATA EXAMPLES

With the purpose of a demonstrating, and comparing to benchmarks, the proposed test with real data, we consider two application to analyze structural changes in the context of oil prices' relationship with the U.S. dollar, and asset price models surrounding the COVID-19 pandemic.

3.1 | The changing relationship between crude oil price and the U.S. dollar.

Commodities in general are priced in dollars, and thus commodity prices and exchange rates tend to be related. Krugman (1980, 1983) presents theoretical models suggesting how exchange rates interact with oil prices, giving conditions under which oil price increases will be associated either with appreciation or depreciation of the U.S. dollar (see also Beckmann et al., 2017, for theoretical results). Empirical work, such as Beckmann et al. (2017), Ji et al. (2019), and Arfaoui and Rejeb (2017), also suggest a relationship between exchange rates and oil prices, with Beckmann et al. (2017) observing that this relationship varies over time.

The datasets used in our example were obtained from the Federal Reserved Economic Database (FRED), the oil price series has FRED code MCOILWTICO and the dollar strength index has FRED code TWEXBPA. In Figures 6 and 7 we observe the inverse relationship between dollar strength and oil prices as found in other empirical studies (see Arfaoui & Rejeb, 2017; Beckmann et al., 2017; Ji et al., 2019). Additionally, we notice that at least one of the series (specifically, oil prices) appears to be heteroscedastic, with volatility appearing to increase after the 1990s. Also, there appears to be a change in the correlation between the series.

We consider (16) as a possible models for the data, where y_t is the price of oil in month t and $x_{t,2}$ is the strength of the dollar as measured by a trade weighted index, that is, $d = 2$, $\mathbf{x}_t = (1, x_{t,2})^\top$ in (1). This is a simplified version of the model appearing in Arfaoui and Rejeb (2017). We acknowledge here that this model is quite simplistic, and does not take into account potential cointegration between the variables. Figures 6 and 7 suggest that the volatility is time dependent and the model in (16) might be reasonable. We used Theorem 1 to test the

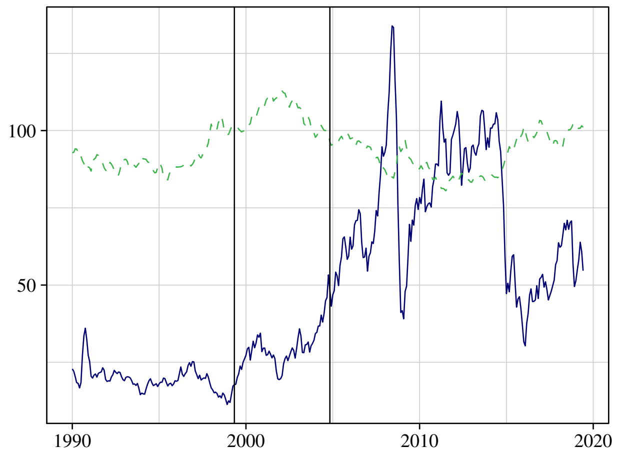


FIGURE 6 The segmented crude oil prices (——) and the strength of the U.S. dollar (– –) data where the vertical lines show the estimated times of changes in the coefficients. [Colour figure can be viewed at wileyonlinelibrary.com]

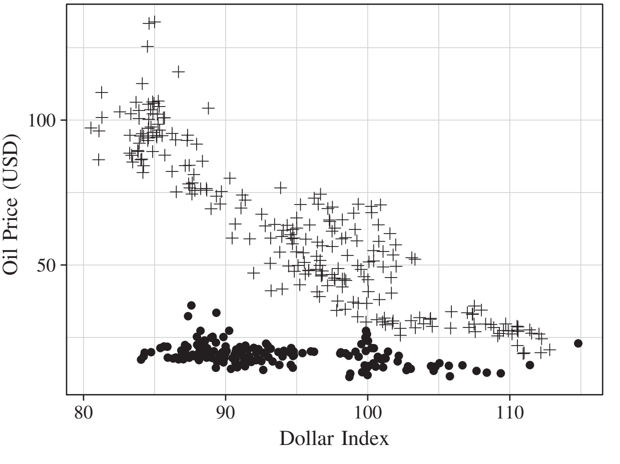


FIGURE 7 Monthly oil prices (in USD) are plotted against a trade-weighted index measuring the strength of the U.S. dollar. Months between January 1990 to February 2000 are plotted as dots, while months from March 2000 to January 2019 are plotted as crosses. Evident difference in correlation and variability are apparent in the two time periods

null hypothesis of no change in the change point regression model (16). The number of observations is $T = 354$ and covers the period January 1990–June 2019 (monthly data) and we chose $a_{354} = b_{354} = 28.2$. The Bartlett kernel was used to estimate the long-run covariances. The null hypothesis is rejected at the 0.05 significance level. In order to estimate the candidate change point, we used the location of the largest value of the test statistic, that is,

$$\hat{t}_T = \min\left\{ t \in [a_t, T - b_T] \, : \, \hat{Z}_T = \left((\hat{\boldsymbol{\beta}}_{t,1} - \hat{\boldsymbol{\beta}}_{t,2})^\top \hat{\mathbf{Q}}_{T,t}^{-1} (\hat{\boldsymbol{\beta}}_{t,1} - \hat{\boldsymbol{\beta}}_{t,2}) \right)^{1/2} \right\}.$$

In our example $\hat{t}_{354} = 178$ and corresponds to October 2004. Applying a binary segmentation to detect/estimate further change points, at $\hat{t}_{354} = 178$ we cut the sample into two subsets and test each for further changes. We do not detect a change in the second subset (October 2004–June 2019). We do detect a change in the first subset (January 1990–September 2004) at $\hat{t}_{177} = 113$, which corresponds to May 1999. Due to small sample sizes we do not look for further changes in the remaining subsets. Thus the original sample is segmented into three subsets and the vertical lines in Figure 6 show the estimated times of changes. Our findings together with the estimates for the $\boldsymbol{\beta}$'s and the sample variances of the error terms computed from the residuals are provided

TABLE 1 Segmentation of the data where a linear connection is assumed between crude oil prices and the strength of the U.S. dollar

	January 1990–April 1999	May 1999 – September 2004	October 2004 – June 2019
γ_1	55.783	77.849	344.278
γ_2	−0.398	−0.464	−2.962
σ_ϵ^2	11.672	30.275	138.421

in Table 1. It is clear that not only the regression coefficients changed but the variances of the errors are time varying.

To demonstrate that our method requires only a few observations before or after a change point to detect it, and exhibits comparatively improved end-of-sample change point detection performance, we consider applying change point tests to samples containing the estimated change point $\hat{t}_{177} = 113$, corresponding to May 1999, near the sample end point. We considered an initial sample of size 120 spanning the period from January 1990 to January, 2000, which contains the point May 1999 near the end, and then calculated approximate p values for tests of H_0 in the aforementioned regression model based on the proposed test along with several benchmarks. This process was then repeated after adding individual observations to the end point of the sample, thereby producing a sequence of p values corresponding to the end point of the growing sample. The negative logarithm of these p values are plotted in Figure 8 as a function of the sample end date. Crossing of the dotted line in Figure 8 for a particular test implies the rejection of the null hypothesis of no change in the regression model at the significance level 5%. Here we see that a change point is detected at the 5% level starting in the period January 1990–August, 2000 when $\hat{Z}_T$ is used to detect a change point. The CUSUM statistic of Ploberger and Krämer (1992) is never able to reject the null hypothesis at this level. The statistics of Hidalgo and Seo (2013) and Horváth et al. (2020) in this case appear to be exhibiting the same oversized effect that was observed in the simulation study as a result of heteroscedasticity of the covariates; they reject the null hypothesis initially, but as the number of observations start increasing they begin to give contradictory and inconsistent results. Andrews' statistic always rejects the null hypothesis with miniscule p values and appears to be unreliable; we cannot even display it in the figure.

We performed an additional simulation experiment to study the behavior of the statistics studied in this paper in contexts tailored to what we observed in this oil price versus U.S. dollar data. To simulate similar data using model (16), we let $T = 130$, $x_{t,2}$ be an ARMA(1, 1) sequence with autocorrelation parameter $\phi = 0.2$ and moving average parameter $\theta = 0.4$, with the innovations as i.i.d. normal random variables with mean zero and variance 1.323. We modelled the errors ϵ_t being i.i.d. normal random variables $N\left(0, \tau_t^2\right)$, with $\tau_t^2 = 9$ when $t \leq 108$ and $\tau_t^2 = 144$ for $t > 108$, i.e. $t_1 = 108$. We let the time of change in the simulations occur at $t^* = 114$. Let $\boldsymbol{\beta}^{(1)} = (55.783, -0.398)^\top$, and after the change be $\boldsymbol{\beta}^{(T)} = \boldsymbol{\beta}^{(1)} + \delta((77.849, -0.464)^\top - (55.783, -0.398)^\top)$, where $\delta \in [0, 2]$. Note that $\delta = 0$ corresponds to the null hypothesis of no change. When computing the LRV we used the Bartlett kernel and set the bandwidth parameter to $1.3T^{1/2}$. The trimming parameters were set to $a_T = b_T = 1.5T^{1/2}$ in the definition of $\hat{Z}_T$ and in the Rényi-type statistic of Horváth et al. (2020). Figure 9 shows that all tests are over sized except $\hat{Z}_T$. Due to the size inflation those tests, not surprisingly, have higher power. Even if the sample size T is increased, the size inflation will remain but $\hat{Z}_T$ stays correctly sized under the null hypothesis and gains power.

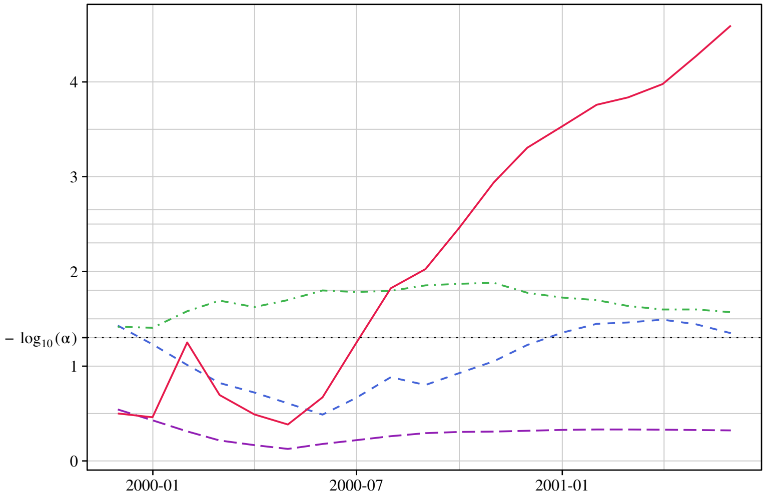
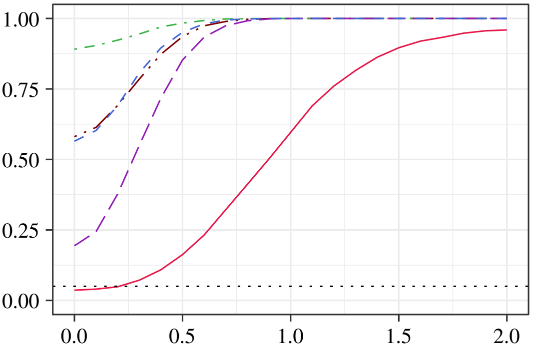


FIGURE 8 $-\log_{10}(p)$-values for $\hat{Z}_T$ (——), Horváth et al. (2020) (_ _ _), Ploberger and Krämer (1992) (_ _ _), and Hidalgo and Seo (2013) (_ . _). The vertical direction can be interpreted as the number of zeros before the first significant digit of the p value. Andrews' statistic is not displayed as it always rejected the null hypothesis with very small p values [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 9 Empirical power of the statistic $\hat{Z}_{130}$ (——), Horváth et al. (2020) (_ _ _), Ploberger and Krämer (1992) (_ _ _) Andrews (1993) (_ . . _), and Hidalgo and Seo (2013) (_ . _). The dotted line corresponds to $\alpha = 0.05$, the significance level of the test [Colour figure can be viewed at wileyonlinelibrary.com]



3.2 | The COVID-19 panedemic's effect on factor-based asset pricing models

In the first quarter of 2020 a global coronavirus pandemic emerged, with massive economic implications strongly affecting people's day-to-day lives. The result was massive economic disruption due to the direct health effects of the virus and also the response to the virus, including quarantines and social distancing policies closing businesses and keeping people in their homes. Setting aside the public health consequences of the pandemic, financial markets witnessed turmoil perhaps even exceeding the 2008 financial crisis. Markets declined in record amounts.

Due to social distancing measures, many adopted video conferencing technology to allow for remote communication. Many chose to use the platform Zoom, provided by Zoom Video Communications. Zoom debuted for public trading only about a year before the crisis Novet (2019) under ticker symbol ZM, but Zoom's popularity yielded excellent performance by ZM near the beginning of the crisis in contrast to general market trends.

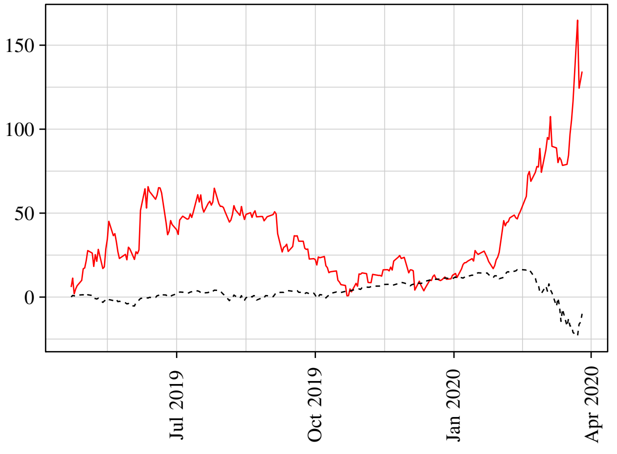


FIGURE 10 The return of SPY (- - -) and ZM (——) on an investment made on April 21, 2019, as a percentage. [Colour figure can be viewed at wileyonlinelibrary.com]

In this paper we consider the one-factor model:

$$R_t - R_{Ft} = \alpha + \beta(R_{Mt} - R_{Ft}) + \epsilon_t, \tag{17}$$

where R_t is the return of the security on day t, R_{Mt} is the market return, and R_{Ft} is the risk-free rate of return. The parameters α and β correspond to the risk/return metrics commonly quoted in the finance industry, with α being excess return over the market and β a measure of the security's sensitivity to market movements. Here the security is ZM, the market return is measured using the return on the ETF SPY, which is intended to mirror the S & P 500, and the risk-free rate is proxied using the return of three month U.S. Treasury bills. We downloaded the ZM and SPY data from the NASDAQ website and the daily interest rate of the Treasury bills from the U.S. Treasury website. The time frame extends from April 21, 2019, to March 26, 2020. We show the behavior of SPY and ZM over this period in Figure 10.

We believe that the crisis prompted a change in the relationship between ZM and the market. Furthermore, we believe that the two metrics α and β moved in different directions due to the crisis, with α increasing and β dropping below zero. While we cannot necessarily a priori identify a day and a single specific event that would prompt the change due to the rolling nature of the crisis, a change should occur sometime in February or March as the virus spread to more countries and concern surrounding it grew.

We applied the tests in this paper to this dataset as they were used in the simulations. The trimming parameter for the new statistic is the square root function. When applied to the whole dataset, the statistic takes its maximal value at February 21, 2020. The null hypothesis of no change was rejected with a minuscule p value. Prior to this day, we estimate that $\alpha = 0.001$ and $\beta = 1.680$; after the change, these parameters shifted to $\alpha = 0.010$ and $\beta = -0.380$. (Note that the prechange sample size is 220 days and the postchange sample size is just 14 days. We cannot be confident on the actual parameter values, just on the existence of a change.) Prior to this date our method did not detect a change point at the 0.05 significance level, so we believe this is the only change point in the dataset.

We compared our test's behavior to that of other tests applied to this dataset. As our test is intended to detect changes occurring near the ends of the sample well, we use the procedure described in Horváth et al. (2020), where we expand the window of days over which the statistics are computed at the right end point of the sample while plotting the $-\log_{10}(p)$-values of the tests, noting when tests start rejecting the null hypothesis. We show these results in Figure 11.

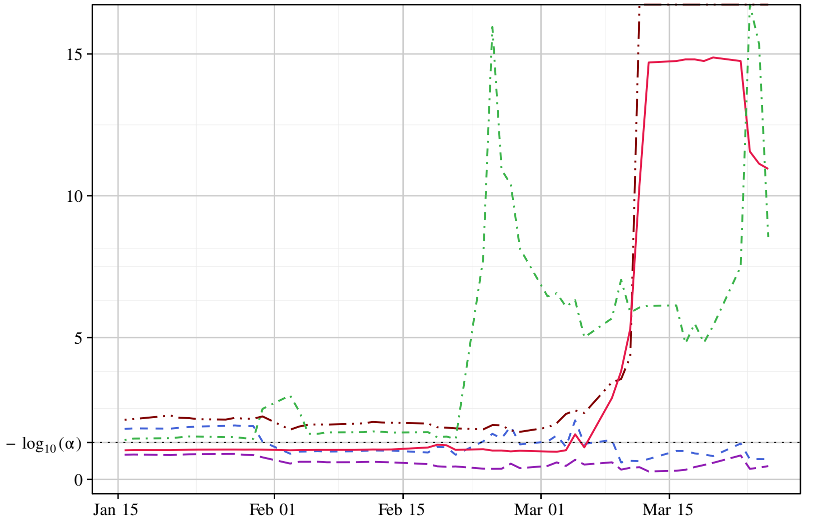


FIGURE 11 $-\log_{10}(p)$-values of test statistics on an expanding window of data. The y-axis can be interpreted as counting the number of zeros before the first significant digit of the p value. The statistics plotted are $\hat{Z}_T$ (——), Horváth et al. (2020) (– – –), Ploberger and Krämer (1992) (– – –) Andrews (1993) (– · · –), and Hidalgo and Seo (2013) (– · –) [Colour figure can be viewed at wileyonlinelibrary.com]

Three statistics clearly detect the change: the statistic introduced in this paper and the statistics proposed in Andrews (1993) and Hidalgo and Seo (2013), all of which are essentially using the regression coefficients rather than the residuals for detecting a change. The CUSUM and residual-based Rényi-type statistic struggle and do not seem to detect a change when one clearly occurred. While the Hidalgo–Seo and Andrews' statistics are able to detect a change quite early, it is unclear whether this is too early, as these tests have rather poor size under heteroscedasticity. The proposed statistic appears to not only not reject the null hypothesis when it should not, but also clearly rejects the null hypothesis, when it should.

4 | PRELIMINARY RESULTS

We assume that H_0 holds and the common regression parameter is denoted by $\boldsymbol{\beta}$. It is easy to see that

$$\hat{\boldsymbol{\beta}}_{t,1} = \boldsymbol{\beta} + (\mathbf{X}_{t,1}^\top \mathbf{X}_{t,1})^{-1}\mathbf{X}_{t,1}^\top \mathbf{E}_{t,1} \quad \text{and} \quad \hat{\boldsymbol{\beta}}_{t,2} = \boldsymbol{\beta} + (\mathbf{X}_{t,2}^\top \mathbf{X}_{t,2})^{-1}\mathbf{X}_{t,2}^\top \mathbf{E}_{t,2},$$

where $\mathbf{E}_{t,1} = (\epsilon_1, \epsilon_2, \dots, \epsilon_t)^\top$ and $\mathbf{E}_{t,2} = (\epsilon_{t+1}, \epsilon_{t+2}, \dots, \epsilon_T)^\top$.

Lemma 1. *If Assumptions 1, 2, and 3 are satisfied, we have that*

$$\max_{t_1 \le t \le T} \|(\mathbf{X}_{t,1}^\top \mathbf{X}_{t,1})^{-1}\| = O_P(1/T), \tag{18}$$

$$\max_{1 \le t \le t_M} \|(\mathbf{X}_{t,2}^\top \mathbf{X}_{t,2})^{-1}\| = O_P(1/T), \tag{19}$$

and

$$\max_{1\le t\le t_1}\|\mathbf{X}_{t,1}^\top\mathbf{X}_{t,1}-t\mathbf{A}_1\|/t^{1/2+\delta}=O_P(1), \tag{20}$$

$$\max_{t_M\le t<T}\|\mathbf{X}_{t,2}^\top\mathbf{X}_{t,2}-(T-t)\mathbf{A}_{M+1}\|/(T-t)^{1/2+\delta}=O_P(1) \tag{21}$$

for any $\delta>0$*, as* $T\to\infty$.

Proof. Writing $\mathbf{g}_i=(g_{i,1},g_{i,2},\dots,g_{i,d})^\top$, we have that $x_t(r)=g_{i,r}(\eta_t,\eta_{t-1},\dots)$ and $x_t(\ell)=g_{i,\ell}(\eta_t,\eta_{t-1},\dots)$, if $t_{i-1}<t\le t_i$. Let

$$x_t^*(j,m)=g_{i,j}(\eta_t,\eta_{t-1},\dots,\eta_{t-m},\eta^*_{t,t-m-1},\eta^*_{t,t-m-2},\dots),\quad \text{if}\ \ t_{i-1}<t\le t_i.$$

By the Cauchy–Schwartz inequality we have with $\kappa=\kappa_1-4$ that

$$\begin{aligned}
&E|x_t(r)x_t(\ell)-x_t^*(r,m)x_t^*(\ell,m)|^{2+\kappa/2}\\
&\quad\le 2^{2+\kappa/2}(E(|x_t(r)-x_t^*(r,m)||x_t^*(\ell,m)|)^{2+\kappa/2}+E(|x_t(\ell)-x_t^*(\ell,m)||x_t(r)|)^{2+\kappa/2})\\
&\quad\le 2^{2+\kappa/2}((E|x_t(r)-x_t^*(r,m)|^{4+\kappa})^{1/2}(E|x_t^*(\ell,m)|^{4+\kappa})^{1/2}\\
&\quad+(E|x_t(\ell)-x_t^*(\ell,m)|^{4+\kappa})^{1/2}(E|x_t(r)|^{4+\kappa})^{1/2})
\end{aligned}$$

so by Assumption 3 for all t and $1\le r,\ell\le d$

$$\sum_{m=1}^{\infty}E|x_t(r)x_t(\ell)-x_t^*(r,m)x_t^*(\ell,m)|^{2+\kappa/2}<\infty. \tag{22}$$

We write $\mathbf{A}_i=\{a_i(r,\ell),1\le r,\ell\le d\}$. Using Aue et al. (2014) we can define Gaussian processes $\Gamma_{i,r,\ell}(u)=\Gamma_{T,i,r,\ell}(u)$ and constants $\sigma^2_{i,r,\ell}$, $1\le k,\ell\le d, 1\le i\le M+1$ such that $E\Gamma_{i,r,\ell}(u)=0$, $\Gamma_{i,r,\ell}(u)\Gamma_{i,r,\ell}(v)=\sigma^2_{i,r,\ell}\min(u,v)$ and

$$\max_{1\le u\le t_i-t_{i-1}}u^{-1/2+\zeta}\left|\sum_{s=t_{i-1}+1}^{t_{i-1}+u}(x_s(r)x_s(\ell)-a_i(r,\ell))-\Gamma_{i,r,\ell}(u)\right|=O_P(1), \tag{23}$$

with some $\zeta>0$. Since $\Gamma_{i,r,\ell}(u)/\sigma_{i,r,\ell}$ is a Wiener process, by the scale transformation of the Wiener process we have that

$$\max_{1\le i\le M+1}T^{-1/2}\max_{1\le u\le t_i-t_{i-1}}|\Gamma_{i,r,\ell}(u)|=O_P(1).$$

Thus we conclude

$$\mathbf{X}_{t,1}^\top\mathbf{X}_{t,1}=(t_1-t_0)\mathbf{A}_1+(t_2-t_1)\mathbf{A}_2+\dots+(t-t_{i-1})\mathbf{A}_i+\mathbf{R}_t,\quad \text{if}\ \ t_{i-1}<t\le t_i,$$

and

$$\max_{1\le t\le T}\|\mathbf{R}_t\|=O_P(T^{1/2}).$$

Using now Assumptions 1 and 2 we obtain (18). By the law of the iterated logarithms we have that

$$\max_{1\le u<\infty} |\Gamma_{i,r,\ell}(u)|/(u\log\log(u+5))^{1/2} = O_P(1),$$

hence (20) follows from (23). The proof of lemma 5.4 in Aue et al. (2014) (cf. also their Theorem B.1) is based on a blocking argument and therefore we can define Gaussian processes $\overline{\Gamma}_{i,r,\ell}(u) = \Gamma_{T,i,r,\ell}(u)$, $1 \le k, \ell \le d, 1 \le i \le M+1$ such that $E\overline{\Gamma}_{i,r,\ell}(u) = 0$, $\overline{\Gamma}_{i,r,\ell}(u)\overline{\Gamma}_{i,r,\ell}(v) = \sigma^2_{i,r,\ell}\min(u,v)$ and

$$\max_{1\le u\le t_i-t_{i-1}} u^{-1/2+\zeta}\left|\sum_{s=t_{i-1}+u}^{t_i}(x_s(r)x_s(\ell)-a_i(r,\ell))-\overline{\Gamma}_{i,r,\ell}(u)\right| = O_P(1), \tag{24}$$

with some $\zeta > 0$. Replacing (23) with (24), the results in (20) and (21) can be derived along the lines of (18) and (19). ■

Let $\epsilon^*_{t,m} = f_i(\eta_t, \eta_{t-1}, \dots, \eta_{t-m}, \eta^*_{t,t-m-1}, \eta^*_{t,t-m-2}, \dots)$, $\mathbf{e}^*_{t,m} = \mathbf{x}^*_{t,m}\epsilon^*_{t,m}$ and

$$\mathbf{x}^*_{t,m} = (x^*_t(1,m), x^*_t(2,m), \dots, x^*_t(d,m))^\top, \quad t_{i-1} < t \le t_i, 1 \le i \le M+1,$$

where the $x_t(r,m)$'s are defined in Lemma 1.

Lemma 2. *If Assumptions 1,2,3 are satisfied, we have that*

$$\max_{1\le t\le T}\left\|\sum_{s=1}^{t}\mathbf{e}_s\right\| = O_P(T^{1/2}). \tag{25}$$

For each T there are two independent Gaussian processes $\mathbf{\Gamma}^{(1)}_T$ and $\mathbf{\Gamma}^{(2)}_T$, $E\mathbf{\Gamma}^{(1)}_T(u) = \mathbf{0}$, $E\mathbf{\Gamma}^{(1)}_T(u)\mathbf{\Gamma}^{(1)}(v) = \mathbf{D}_1\min(u,v)$, $E\mathbf{\Gamma}^{(2)}_T(u) = \mathbf{0}$, $E\mathbf{\Gamma}^{(2)}_T(u)\mathbf{\Gamma}^{(2)}(v) = \mathbf{D}_{M+1}\min(u,v)$, and

$$\max_{1\le t\le t_1} t^{-1/2+\zeta}\left\|\sum_{s=1}^{t}\mathbf{e}_s - \mathbf{\Gamma}^{(1)}_T(t)\right\| = O_P(1), \tag{26}$$

$$\max_{1\le t\le k_{M+1}} t^{-1/2+\zeta}\left\|\sum_{s=T-t}^{T}\mathbf{e}_s - \mathbf{\Gamma}^{(2)}_T(t)\right\| = O_P(1). \tag{27}$$

with some $\zeta > 0$.

Proof. Following the proof of Lemma 1 one can easily show that

$$\sum_{m=1}^{\infty} E\|\mathbf{e}_t - \mathbf{e}^*_{t,m}\|^{2+\kappa/2} < \infty.$$

So using again Aue et al. (2014) we obtain immediately (25). The approximations in (26) and (27) follow from lemma 5.4 in Aue et al. (2014). The independence of the approximating Gaussian processes is a consequence of the blocking method used in their proof. ■

5 | PROPERTIES OF $\hat{\mathbf{Q}}_t(1)$ AND $\hat{\mathbf{Q}}_t(2)$

The asymptotic properties of $\hat{\mathbf{Q}}_{T,t}$ will be derived from Lemma 3. To state the result we need to introduce further notation.

Assumption 8. Let

$$\alpha_i = \mathcal{A}_j(\eta_i, \eta_{i-1}, \eta_{i-2}, \ldots), \text{ if } t_{j-1} < i \le t_j, 1 \le j \le M+1,$$

and

$$\beta_i = \mathcal{B}_j(\eta_i, \eta_{i-1}, \eta_{i-2}, \ldots), \text{ if } t_{j-1} < t \le t_j, 1 \le j \le M+1,$$

where $\mathcal{A}_j$ and $\mathcal{B}_j$ are nonrandom functionals defined on $\mathcal{S}^\infty$ with values in $\mathbb{R}$ and $\mathcal{S}$ is a measurable space. Also, $\eta_i = \eta_i(s, \omega)$ is jointly measurable in (s, ω), $-\infty < i < \infty$. The sequences $\alpha_i, \beta_i, -\infty < i < \infty$ can be approximated with m-dependent sequences $\alpha_{i,m}$ and $\beta_{i,m}$ in the sense that with some $\kappa_1 > 4$, $\kappa_2 > 2$ and $c > 0$ $E|\alpha_i|^{\kappa_1} < \infty, E|\beta_i|^{\kappa_1} < \infty$,

$$(E|\alpha_i - \alpha_{i,m}|^{\kappa_1})^{1/\kappa_1} \le cm^{-\kappa_2}, \tag{28}$$

and

$$(E|\beta_i - \beta_{i,m}|^{\kappa_1})^{1/\kappa_1} \le cm^{-\kappa_2}, \tag{29}$$

where $\alpha_{i,m} = \mathcal{A}_j(\eta_i, \eta_{i-1}, \eta_{i-2}, \ldots, \eta_{i-m+1}, \boldsymbol{\eta}^*_{i,m})$, $\beta_{i,m} = \mathcal{B}_j(\eta_i, \eta_{i-1}, \eta_{i-2}, \ldots, \eta_{i-m+1}, \boldsymbol{\eta}^*_{i,m})$, $t_{j-1} < i \le t_j, 1 \le j \le M+1$, $\boldsymbol{\eta}^*_{i,m} = \left(\eta^*_{i,m,i-m}, \eta^*_{i,m,i-m-1}, \eta^*_{i,m,i-m-2}, \ldots\right)$ and the $\eta^*_{i,m,n}$'s are independent copies of η_0, independent of $\{\eta_i, -\infty < i < \infty\}$.

Let

$$U_t = \frac{1}{t}\sum_{s=1}^{t} \alpha_s\beta_s \quad \text{and} \quad V_t = \sum_{u=1}^{t-1} K\left(\frac{u}{h_t}\right)\frac{1}{t-u}\sum_{s=1}^{t-u}\alpha_s\beta_{s+u},$$

and similarly

$$U^*_t = \frac{1}{T-t}\sum_{s=t+1}^{T} \alpha_s\beta_s \quad \text{and} \quad V^*_t = \sum_{u=t+1}^{T} K\left(\frac{u}{h_{T-t}}\right)\frac{1}{T-t-u}\sum_{s=t+1}^{T-u}\alpha_s\beta_{s+u}.$$

Lemma 3. *If Assumptions 5, 6, 7, and 8 hold, then we have that*

$$\max_{a_T \le t \le T}\left|U_t - \frac{1}{t}\sum_{s=1}^{t} E\alpha_s\beta_s\right| = o_P(1), \tag{30}$$

$$\max_{1 \le t \le T-b_T}\left|U^*_t - \frac{1}{T-t}\sum_{s=t+1}^{T} E\alpha_s\beta_s\right| = o_P(1), \tag{31}$$

$$\max_{a_T \le t \le T}|V_t - EV_t| = o_P(1), \tag{32}$$

and

$$\max_{1\le t\le T-b_T}|V_t^*-EV_t^*|=o_P(1). \tag{33}$$

Proof. We note

$$\mathrm{var}\left(\sum_{s=t_1}^{t_2}\alpha_s\beta_s\right)=\sum_{s=t_1}^{t_2}\sum_{z=t_1}^{t_2}E\{(\alpha_s\beta_s-E\alpha_s\beta_s)(\alpha_z\beta_z-E\alpha_z\beta_z)\}.$$

Using Assumption 8 there is a constant C_1 such that for all s

$$\sum_{z=1}^{\infty}|E\{(\alpha_s\beta_s-E\alpha_s\beta_s)(\alpha_z\beta_z-E\alpha_z\beta_z)|\le C_1,$$

and therefore

$$\mathrm{var}\left(\sum_{s=t_1}^{t_2}\alpha_s\beta_s\right)\le C_1(t_2-t_1+1). \tag{34}$$

Hence for all $\delta>0$ we have by Menshov's inequality (Billingsley, 1968, p. 102) that

$$\begin{aligned}
P\Big\{\max_{a_T\le t\le T}|U_t-EU_t)|\ge\delta\Big\}
&\le P\Big\{\max_{\log(a_T)\le i\le\log T}\max_{e^i\le t\le e^{i+1}}|U_t-EU_t)|\ge\delta\Big\}\\
&\le\sum_{i=\log(a_T)}^{\infty}P\Big\{\max_{e^i\le t\le e^{i+1}}|U_t-EU_t|\ge\delta\Big\}\\
&\le\sum_{i=\log(a_T)}^{\infty}P\left\{\max_{1\le t\le e^{i+1}}\left|\sum_{s=1}^{t}\alpha_s\beta_s-tEU_t\right|\ge\delta e^i\right\}\\
&\le\sum_{i=\log(a_T)}^{\infty}\frac{1}{\delta^2}e^{-2i}E\max_{1\le t\le e^{i+1}}\left|\sum_{s=1}^{t}\alpha_s\beta_s-tEU_t\right|^2\\
&\le C_2\sum_{i=\log(a_T)}^{\infty}\frac{1}{\delta^2}e^{-2i}e^ii^2,
\end{aligned} \tag{35}$$

with some constant C_2. Hence the proof of (30) is complete. Since the proof of (30) was based on bounds for moments, and the proof of theorem B.1 in Aue et al. (2014) is based on a blocking argument, we can obtain the result in (31) similarly, although now we are working backwards in time. Let

$$V_{t,1}=\sum_{u=1}^{t-1}K\left(\frac{u}{h_t}\right)\frac{1}{t}\sum_{s=1}^{t-u}\alpha_s\beta_{s+u}.$$

It is easy to see that by Assumption 5 there is a constant C_3 such that

$$
\begin{aligned}
&\mathrm{var}(V_t - V_{t,1}) \\
&= \sum_{u=1}^{t-1}\sum_{v=1}^{t-1} K\left(\frac{u}{h_t}\right) K\left(\frac{v}{h_t}\right) \frac{u}{t(t-u)} \frac{v}{t(t-v)} \sum_{s=1}^{t-u}\sum_{z=1}^{t-v} E[(\alpha_s\beta_{s+u} - E\alpha_s\beta_{s+u})(\alpha_z\beta_{z+v} - E\alpha_z\beta_{z+v})] \\
&\le C_3 \frac{h_t^2}{t^4} \sum_{u=1}^{ch_t}\sum_{v=1}^{ch_t} \left| \sum_{s=1}^{t-u}\sum_{z=1}^{t-v} E[(\alpha_s\beta_{s+u} - E\alpha_s\beta_{s+u})(\alpha_z\beta_{z+v} - E\alpha_z\beta_{z+v})] \right| .
\end{aligned}
$$

Using again Assumption 8 there is a constant C_4 such that for all t, u, and v

$$
\sum_{z=1}^{\infty} |E[(\alpha_s\beta_{s+u} - E\alpha_s\beta_{s+u})(\alpha_z\beta_{z+v} - E\alpha_z\beta_{z+v})]| \le C_4,
$$

and therefore

$$
\mathrm{var}(V_t - V_{t,1}) \le C_5 \frac{h_t^4}{t^3},
$$

with some constant C_5. For all $\delta > 0$, we have by the Chebyshev inequality and Assumption 6 that

$$
\begin{aligned}
&P\left\{ \max_{a_T \le t \le T} |V_t - EV_t - (V_{t,1} - EV_{t,1})| \ge \delta \right\} \\
&\le \sum_{t=a_T}^{\infty} P\left\{ |V_t - EV_t - (V_{t,1} - EV_{t,1})| \ge \delta \right\} \\
&\le \frac{C_5}{\delta^2} \sum_{t=a_T}^{\infty} \frac{h_t^4}{t^3} \\
&\le \frac{C_6}{\delta^2} \sum_{t=a_T}^{\infty} \frac{(t^{1/2}(\log t)^{-(3+\zeta)})^4}{t^3} \to 0, \quad \text{as} \quad T \to \infty,
\end{aligned}
$$

with some C_6. Let

$$
V_{t,2} = \sum_{u=1}^{t-1} K\left(\frac{u}{h_t}\right) \frac{1}{t} \sum_{s=1}^{t} \alpha_s\beta_{s+u}.
$$

Using the previous arguments one can show that with some constant C_7

$$
\mathrm{var}(V_{t,1} - V_{t,2}) \le C_7 \frac{h_t^2}{t^2},
$$

and therefore by Assumption 6

$$
\max_{a_t \le t \le T} |V_{t,1} - EV_{t,1} - (V_{t,2} - EV_{t,2})| \xrightarrow{P} 0, \quad \text{as} \quad t \to \infty.
$$

Next we write

$$
V_{t,2} - EV_{t,2} = \frac{1}{t} \sum_{s=1}^{t} \sum_{u=1}^{ch_t} K\left(\frac{u}{h_t}\right) (\alpha_s\beta_{s+u} - E\alpha_s\beta_{s+u}).
$$

Using again Assumptions 5 and 8 we conclude that for every constant h_*

$$\begin{aligned}
&\operatorname{var}\left(\sum_{s=t_1}^{t_2}\sum_{u=1}^{ch_*}K\left(\frac{u}{h_*}\right)(\alpha_s\beta_{s+u}-E\alpha_s\beta_{s+u})\right)\\
&=\sum_{s=t_1}^{t_2}\sum_{z=t_1}^{t_2}\sum_{u=1}^{ch_*}\sum_{v=1}^{ch_*}K\left(\frac{u}{h_*}\right)K\left(\frac{v}{h_*}\right)E[(\alpha_s\beta_{s+u}-E\alpha_s\beta_{s+u})(\alpha_z\beta_{z+v}-E\alpha_z\beta_{z+v})]\\
&\le C_8h_*^2(t_2-t_1+1).
\end{aligned} \tag{36}$$

Proceeding along the lines of (35) we get for all $\delta > 0$

$$\begin{aligned}
&P\left\{\max_{a_T\le t\le T}|V_{t,2}-EV_{t,2})|\ge\delta\right\}\\
&\quad\le P\left\{\max_{\log_\rho(a_T)\le i\le\log_\rho T}\max_{\rho^i\le t<\rho^{i+1}}|V_{t,2}-EV_{t,2})|\ge\delta\right\}\\
&\quad\le\sum_{i=\log_\rho(a_T)}^{\infty}P\left\{\max_{\rho^i\le t<\rho^{i+1}}|V_{t,2}-EV_{t,2}|\ge\delta\right\}\\
&\quad\le\sum_{i=\log_\rho(a_T)}^{\infty}P\left\{\max_{\rho^i\le t<\rho^{i+1}}\left|\sum_{s=1}^{t}\sum_{u=1}^{ch_t}K\left(\frac{u}{h_t}\right)(\alpha_s\beta_{s+u}-E\alpha_s\beta_{s+u})\right|\ge\delta\rho^i\right\}.
\end{aligned}$$

Using (36) and the assumption that h_t is constant on $[\rho^{i-1}, \rho^i)$, we get

$$\begin{aligned}
&\max_{\rho^i\le t<\rho^{i+1}}\left|\sum_{s=1}^{t}\sum_{u=1}^{ch_t}K\left(\frac{u}{h_t}\right)(\alpha_s\beta_{s+u}-E\alpha_s\beta_{s+u})\right|\\
&\quad=\max_{\rho^i\le t<\rho^{i+1}}\left|\sum_{s=1}^{t}\sum_{u=1}^{ch_{\rho^{i-1}}}K\left(\frac{u}{h_{\rho^{i-1}}}\right)(\alpha_s\beta_{s+u}-E\alpha_s\beta_{s+u})\right|\\
&\quad\le\max_{1\le t<\rho^{i+1}}\left|\sum_{s=1}^{t}\sum_{u=1}^{ch_{\rho^{i-1}}}K\left(\frac{u}{h_{\rho^{i-1}}}\right)(\alpha_s\beta_{s+u}-E\alpha_s\beta_{s+u})\right|.
\end{aligned}$$

By (36) and Menshov's inequality (Billingsley, 1968, p. 102) we conclude that with some constant C_9

$$E\left(\max_{1\le t<\rho^{i+1}}\left|\sum_{s=1}^{t}\sum_{u=1}^{ch_{\rho^{i-1}}}K\left(\frac{u}{h_{\rho^{i-1}}}\right)(\alpha_s\beta_{s+u}-E\alpha_s\beta_{s+u})\right|\right)^2\le C_9h^2_{\rho^{i-1}}\rho^i i^2,$$

and therefore by Assumption 6

$$\begin{aligned}
&\sum_{i=\log_\rho(a_T)}^{\infty}P\left\{\max_{\rho^i\le t<\rho^{i+1}}\left|\sum_{s=1}^{t}\sum_{u=1}^{ch_t}K\left(\frac{u}{h_t}\right)(\alpha_s\beta_{s+u}-E\alpha_s\beta_{s+u})\right|\ge\delta\rho^i\right\}\\
&\quad\le\sum_{i=\log_\rho(a_T)}^{\infty}\frac{1}{\delta^2}\rho^{-2i}C_8h^2_{\rho^{i-1}}\rho^i i^2
\end{aligned}$$

$$= \frac{C_8}{\delta^2} \sum_{i=\log_\rho(a_T)}^{\infty} h^2_{\rho^{i-1}} \rho^{-i} i^2$$
$$\leq \frac{C_{10}}{\delta^2} \sum_{i=\log_\rho(a_T)}^{\infty} \rho^i (\log \rho^i)^{-(3+\zeta)} \rho^{-i} i^2 \to 0, \quad \text{as} \quad T \to \infty.$$

Hence the proof of (32) is complete. Similar arguments give (33). ■

We recall that $\mathbf{e}_t = \mathbf{x}_t \epsilon_t$ and define

$$\mathbf{Q}_{t,1} = \frac{1}{t} \sum_{s=1}^{t} \mathbf{e}_s \mathbf{e}_s^\top + \sum_{u=1}^{t-1} K\left(\frac{u}{h_t}\right) \frac{1}{t-u} \left(\sum_{s=1}^{t-u} \mathbf{e}_s \mathbf{e}_{s+u}^\top + \sum_{s=1}^{t-u} \mathbf{e}_{s+u} \mathbf{e}_s^\top \right),$$

and

$$\mathbf{Q}_{t,2} = \frac{1}{T-t} \sum_{s=t+1}^{T} \mathbf{e}_s \mathbf{e}_s^\top + \sum_{u=1}^{T-t-1} K\left(\frac{u}{h_{T-t}}\right) \frac{1}{T-t+u} \left(\sum_{s=t+1}^{T-u} \mathbf{e}_s \mathbf{e}_{s+u}^\top + \sum_{s=t+1}^{T-u} \mathbf{e}_{s+u} \mathbf{e}_s^\top \right).$$

Lemma 4. *If Assumptions 1, 2, 3, 4, 5, and 6 are satisfied we have that*

$$\max_{d \leq t \leq T} \|\hat{\mathbf{Q}}_t(1) - \mathbf{Q}_{t,1}\| = o_P(1), \tag{37}$$

and

$$\max_{1 \leq t \leq T-d} \|\hat{\mathbf{Q}}_t(2) - \mathbf{Q}_{t,2}\| = o_P(1). \tag{38}$$

Proof. It is easy to see that

$$\hat{\epsilon}_t = \mathbf{x}_t^\top (\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}_{T,1}) + \epsilon_t, 1 \leq t \leq T,$$

where $\boldsymbol{\beta}_0$ denotes the common value of the regression parameter under H_0 and

$$\hat{\mathbf{e}}_t = \mathbf{x}_t \mathbf{x}_t^\top (\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}_{T,1}) + \mathbf{e}_t, 1 \leq t \leq T.$$

It follows from Lemmas 1 and 2 that

$$\|\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}_{T,1}\| = O_P(T^{-1/2}),$$

and therefore

$$\left\| \sum_{s=1}^{t} \mathbf{x}_s \mathbf{x}_s^\top (\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}_{T,1})(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}_{T,1})^\top \mathbf{x}_s^\top \mathbf{x}_s \right\| = O_P\left(\frac{1}{T}\right) \sum_{s=1}^{t} \|\mathbf{x}_s\|^4.$$

Assumption 3 yields (cf. lemma 5.4 in Aue et al., 2014) that

$$\max_{1 \leq t \leq T} \frac{1}{t} \sum_{s=1}^{t} \|\mathbf{x}_s\|^4 = O_P(1),$$

and therefore

$$\max_{1\le t\le T}\frac{1}{t}\left\|\sum_{s=1}^{t}\mathbf{x}_s\mathbf{x}_s^\top(\boldsymbol{\beta}_0-\hat{\boldsymbol{\beta}}_{T,1})(\boldsymbol{\beta}_0-\hat{\boldsymbol{\beta}}_{T,1})^\top\mathbf{x}_s^\top\mathbf{x}_s\right\| = O_P\left(\frac{1}{T}\right).$$

Similarly, Assumption 3 yields

$$\max_{1\le t\le T}\frac{1}{t}\left\|\sum_{s=1}^{t}\mathbf{e}_s\mathbf{e}_s^\top\right\| = O_P(1).$$

Hence

$$\begin{aligned}
&\max_{1\le t\le T}\frac{1}{t}\left\|\sum_{s=1}^{t}\mathbf{x}_s\mathbf{x}_s^\top(\boldsymbol{\beta}_0-\hat{\boldsymbol{\beta}}_{T,1})\mathbf{e}_s^\top\right\| \\
&\le \max_{1\le t\le 1}\frac{1}{t}\left\|\sum_{s=1}^{t}\mathbf{x}_s\mathbf{x}_s^\top(\boldsymbol{\beta}_0-\hat{\boldsymbol{\beta}}_{T,1})(\mathbf{x}_s\mathbf{x}_s^\top(\boldsymbol{\beta}_0-\hat{\boldsymbol{\beta}}_{T,1}))^\top\right\|\left\|\sum_{s=1}^{t}\mathbf{e}_s\mathbf{e}_s^\top\right\| \\
&\le \max_{1\le t\le T}\left\|\frac{1}{t}\sum_{s=1}^{t}\mathbf{x}_s\mathbf{x}_s^\top(\boldsymbol{\beta}_0-\hat{\boldsymbol{\beta}}_{T,1})(\mathbf{x}_s\mathbf{x}_s^\top(\boldsymbol{\beta}_0-\hat{\boldsymbol{\beta}}_{T,1}))^\top\right\|\max_{1\le t\le T}\left\|\frac{1}{t}\sum_{s=1}^{t}\mathbf{e}_s\mathbf{e}_s^\top\right\| \\
&= O_P(T^{-1/2}).
\end{aligned}$$

The result in (37) is now proven by Assumptions 5 and 6. Similar arguments give (38). ■

Lemma 5. *If Assumptions 1, 2, 3, 4, 5, 6, and 7 are satisfied we have that*

$$\max_{t_1\le t\le T}\|\hat{\mathbf{Q}}_t(1)^{-1}\| = O_P(1)$$

$$\max_{1\le t\le t_M}\|\hat{\mathbf{Q}}_t(2)^{-1}\| = O_P(1).$$

Proof. According to Lemma 4 we need to prove only that

$$\max_{t_1\le t\le T}\|\mathbf{Q}_{t,1}^{-1}\| = O_P(1), \tag{39}$$

and

$$\max_{1\le t\le t_M}\|\mathbf{Q}_{t,2}^{-1}\| = O_P(1). \tag{40}$$

We prove only (39) since similar arguments give (40). Using Lemma 3 for each coordinate of the matrix $\mathbf{Q}_{t,1}$ we get that

$$\max_{a_T\le t\le T}\|\mathbf{Q}_{t,1} - E\mathbf{Q}_{t,1}\| = o_P(1).$$

Let

$$\mathbf{U}_{t,1} = \sum_{j=1}^{i-1}(\theta_j-\theta_{j-1})\mathbf{D}_j + \frac{t-t_{i-1}}{T}\mathbf{D}_i,\quad t_{i-1}<t\le t_i, 1\le i\le M+1.$$

$\left(\sum_{\emptyset}=0\right)$. Assumptions 5 and 6 yield that

$$\max_{a_T\le t\le T}\|E\mathbf{Q}_{t,1}-\mathbf{U}_{t,1}\|\to 0.$$

If $\mathbf{v}\in\mathbb{R}^d$ is an eigenvector of $\mathbf{D}_1$, then $\mathbf{v}^\top\mathbf{U}_{t,1}\mathbf{v}\ge\theta_1\mathbf{v}^\top\mathbf{D}_1\mathbf{v}$ for all $t_1\le t\le T$. This completes the proof of (39). ∎

Lemma 6. *If Assumptions 1, 2, 3, 4, 5, 6, and 7 are satisfied, then we have that*

$$\max_{a_T\le t\le t_1}\|\hat{\mathbf{Q}}_t(1)^{-1}-\mathbf{D}_1^{-1}\|=o_P(1),$$

and

$$\max_{t_M\le t\le T-b_T}\|\hat{\mathbf{Q}}_t(2)^{-1}-\mathbf{D}_{M+1}^{-1}\|=o_P(1).$$

Proof. Following the proof of Lemma 5 one can verify that

$$\max_{a_T\le t\le t_1}\|\hat{\mathbf{Q}}_t(1)-\mathbf{D}_1\|=o_P(1),$$

and

$$\max_{t_M\le t\le T-b_T}\|\hat{\mathbf{Q}}_t(2)-\mathbf{D}_{M+1}\|=o_P(1),$$

which imply Lemma 6. ∎

6 | PROOFS OF THEOREMS 1, 2, AND REMARK 1

Proof of Theorem 1. We write

$$\hat{Z}_T^2=\max_{1\le i\le 5}\hat{Z}_{T,i},$$

where

$$\begin{aligned}
\hat{Z}_{T,1}&=\max_{a_T\le t\le c_T}(\hat{\boldsymbol{\beta}}_{t,1}-\hat{\boldsymbol{\beta}}_{t,2})^\top\hat{\mathbf{Q}}_{T,t}^{-1}(\hat{\boldsymbol{\beta}}_{t,1}-\hat{\boldsymbol{\beta}}_{t,2}),\\
\hat{Z}_{T,2}&=\max_{c_T\le t\le t_1}(\hat{\boldsymbol{\beta}}_{t,1}-\hat{\boldsymbol{\beta}}_{t,2})^\top\hat{\mathbf{Q}}_{T,t}^{-1}(\hat{\boldsymbol{\beta}}_{t,1}-\hat{\boldsymbol{\beta}}_{t,2}),\\
\hat{Z}_{T,3}&=\max_{t_1\le t\le t_M}(\hat{\boldsymbol{\beta}}_{t,1}-\hat{\boldsymbol{\beta}}_{t,2})^\top\hat{\mathbf{Q}}_{T,t}^{-1}(\hat{\boldsymbol{\beta}}_{t,1}-\hat{\boldsymbol{\beta}}_{t,2}),\\
\hat{Z}_{T,4}&=\max_{t_M\le t\le T-d_T}(\hat{\boldsymbol{\beta}}_{t,1}-\hat{\boldsymbol{\beta}}_{t,2})^\top\hat{\mathbf{Q}}_{T,t}^{-1}(\hat{\boldsymbol{\beta}}_{t,1}-\hat{\boldsymbol{\beta}}_{t,2}),\\
\hat{Z}_{T,5}&=\max_{T-d_T\le t\le T-b_T}(\hat{\boldsymbol{\beta}}_{t,1}-\hat{\boldsymbol{\beta}}_{t,2})^\top\hat{\mathbf{Q}}_{T,t}^{-1}(\hat{\boldsymbol{\beta}}_{t,1}-\hat{\boldsymbol{\beta}}_{t,2}),
\end{aligned}$$

$$c_T/a_T\to\infty,\ c_T/T\to 0;\text{ and }\ d_T/b_T\to\infty,\ d_T/T\to 0.$$

Using Lemmas 1 and 5 we conclude

$$\hat{Z}_{T,3} = O_P(1) \max_{t_1 \le t \le t_M} \|\hat{\boldsymbol{\beta}}_{t,1} - \hat{\boldsymbol{\beta}}_{t,2}\|^2.$$

Lemma 2 yields that

$$\max_{t_1 \le t \le t_M} \left\| \frac{1}{t} \sum_{s=1}^{t} \mathbf{e}_s \right\| = O_P(T^{-1/2}) \quad \text{and} \quad \max_{t_1 \le t \le t_M} \left\| \frac{1}{T-t} \sum_{s=t+1}^{T} \mathbf{e}_s \right\| = O_P(T^{-1/2}),$$

so by Lemma 1 we have

$$\max_{t_1 \le t \le t_M} \|\hat{\boldsymbol{\beta}}_{t,1} - \boldsymbol{\beta}_0\| = O_P(T^{-1/2}),$$

and

$$\max_{t_1 \le t \le t_M} \|\hat{\boldsymbol{\beta}}_{t,2} - \boldsymbol{\beta}_0\| = O_P(T^{-1/2}),$$

where $\boldsymbol{\beta}_0$ denotes the common value of the regression coefficient under H_0. Hence

$$r_T \hat{Z}_{T,3} = o_P(r_T/T) = o_P(1), \tag{41}$$

on account of Assumption 7. We note that by (20)

$$\max_{c_T \le t \le t_1} \|\hat{\boldsymbol{\beta}}_{t,1} - \boldsymbol{\beta}_0\| = O_P(1) \max_{c_T \le t \le t_1} \left\| \frac{1}{t} \sum_{s=1}^{t} \mathbf{e}_s \right\|,$$

and by (26) we obtain that

$$\max_{c_T \le t \le t_1} \left\| \frac{1}{t} \sum_{s-1}^{t} \mathbf{e}_s \right\| = O_P\left(1/c_T^{1/2}\right),$$

resulting in

$$\max_{c_T \le t \le t_1} \|\hat{\boldsymbol{\beta}}_{t,1} - \boldsymbol{\beta}_0\| = O_P\left(1/c_T^{1/2}\right). \tag{42}$$

Similarly,

$$\max_{c_T \le t \le t_1} \|\hat{\boldsymbol{\beta}}_{t,2} - \boldsymbol{\beta}_0\| = O_P\left(1/c_T^{1/2}\right). \tag{43}$$

Thus the first part of Lemma 6 yields

$$r_T \hat{Z}_{T,2} = O_P(r_T/c_T) = o_P(1). \tag{44}$$

Similar arguments give that

$$r_T \hat{Z}_{T,4} = o_P(1). \tag{45}$$

Next we show that

$$\left| \hat{Z}_{T,1} - \max_{a_T \le u \le c_T} \frac{1}{u^2} (\mathbf{\Gamma}^{(1)}(u))^\top \mathbf{D}_1^{-1} \mathbf{\Gamma}^{(1)}(u) \right| = o_P(1/a_T). \tag{46}$$

According to Assumption 7 and Lemma 2 we need to show only that

$$\left| \max_{a_T \le t \le c_T} (\hat{\boldsymbol{\beta}}_{t,1} - \boldsymbol{\beta}_0)^\top \frac{\mathbf{X}_{t,1}^\top \mathbf{X}_{t,1}}{t} \hat{\mathbf{Q}}_{t,1}^{-1} \frac{\mathbf{X}_{t,1}^\top \mathbf{X}_{t,1}}{t} (\hat{\boldsymbol{\beta}}_{t,1} - \boldsymbol{\beta}_0) - \max_{a_T \le u \le c_T} \frac{1}{u^2} (\mathbf{\Gamma}^{(1)}(u))^\top \mathbf{D}_1^{-1} \mathbf{\Gamma}^{(1)}(u) \right| .$$
$$= o_P(1/a_T).$$

It follows from the definition of $\hat{\boldsymbol{\beta}}_{t,1}$ that

$$(\hat{\boldsymbol{\beta}}_{t,1} - \boldsymbol{\beta}_0)^\top \frac{\mathbf{X}_{t,1}^\top \mathbf{X}_{t,1}}{t} \hat{\mathbf{Q}}_{t,1}^{-1} \frac{\mathbf{X}_{t,1}^\top \mathbf{X}_{t,1}}{t} (\hat{\boldsymbol{\beta}}_{t,1} - \boldsymbol{\beta}_0) = \left(\frac{1}{t} \sum_{s=1}^{t} \mathbf{e}_s \right)^\top \hat{\mathbf{Q}}_{t,1}^{-1} \left(\frac{1}{t} \sum_{s=1}^{t} \mathbf{e}_s \right),$$

and Lemma 6 yields

$$\max_{a_T \le t \le c_T} \left(\frac{1}{t} \sum_{s=1}^{t} \mathbf{e}_s \right)^\top \hat{\mathbf{Q}}_{t,1}^{-1} \left(\frac{1}{t} \sum_{s=1}^{t} \mathbf{e}_s \right)$$
$$= \max_{a_T \le t \le c_T} \left(\frac{1}{t} \sum_{s=1}^{t} \mathbf{e}_s \right)^\top \mathbf{D}_1^{-1} \left(\frac{1}{t} \sum_{s=1}^{t} \mathbf{e}_s \right) + o_P(1) \max_{a_T \le t \le c_T} \left\| \frac{1}{t} \sum_{s=1}^{t} \mathbf{e}_s \right\|^2 .$$

Using Lemma 2 we conclude that

$$a_T \max_{a_T \le t \le c_T} \left\| \frac{1}{t} \sum_{s=1}^{t} \mathbf{e}_s \right\|^2 = a_T \max_{a_T \le u \le c_T} \left\| \frac{1}{u} \mathbf{\Gamma}^{(1)}(u) \right\|^2 (1 + o_P(1)).$$

Using the calculations presented in the proof of lemma B.2 on p. 20 of the supplementary materials to Horvath et al. (2020), we obtain that

$$a_T \max_{a_T \le u \le c_T} \left\| \frac{1}{u} \mathbf{\Gamma}^{(1)}(u) \right\|^2 = O_P(1),$$

and therefore

$$\max_{a_T \le t \le c_T} (\hat{\boldsymbol{\beta}}_{t,1} - \boldsymbol{\beta}_0)^\top \frac{\mathbf{X}_{t,1}^\top \mathbf{X}_{t,1}}{t} \hat{\mathbf{Q}}_{t,1}^{-1} \frac{\mathbf{X}_{t,1}^\top \mathbf{X}_{t,1}}{t} (\hat{\boldsymbol{\beta}}_{t,1} - \boldsymbol{\beta}_0)$$
$$= \max_{a_T \le t \le c_T} \left(\frac{1}{t} \sum_{s=1}^{t} \mathbf{e}_s \right)^\top \mathbf{D}_1^{-1} \left(\frac{1}{t} \sum_{s=1}^{t} \mathbf{e}_s \right) + o_P(1/a_T).$$

Applying again Lemma 2 we get that

$$a_T \max_{a_T \le t \le c_T} \left(\frac{1}{t} \sum_{s=1}^{t} \mathbf{e}_s \right)^\top \mathbf{D}_1^{-1} \left(\frac{1}{t} \sum_{s=1}^{t} \mathbf{e}_s \right)$$
$$= a_T \max_{a_T \le u \le c_T} \left(\frac{1}{u} \boldsymbol{\Gamma}_T^{(1)}(u) \right)^\top \mathbf{D}_1^{-1} \left(\frac{1}{u} \boldsymbol{\Gamma}_T^{(1)}(u) \right) + o_P(1),$$

completing the proof of (46). Similar arguments give

$$\left| \hat{Z}_{T,5} - \max_{b_T \le u \le d_T} \left(\frac{1}{u} \boldsymbol{\Gamma}^{(2)}(u) \right)^\top \mathbf{D}_{M+1}^{-1} \left(\frac{1}{u} \boldsymbol{\Gamma}^{(2)}(u) \right) \right| = o_P(1/b_T). \tag{47}$$

Observing that

$$\left(a_T \max_{a_T \le u \le c_T} \left(\frac{1}{u} \boldsymbol{\Gamma}_T^{(1)}(u) \right)^\top \mathbf{D}_1^{-1} \left(\frac{1}{u} \boldsymbol{\Gamma}_T^{(1)}(u) \right), \max_{b_T \le u \le d_T} \left(\frac{1}{u} \boldsymbol{\Gamma}^{(2)}(u) \right)^\top \mathbf{D}_{M+1}^{-1} \left(\frac{1}{u} \boldsymbol{\Gamma}^{(2)}(u) \right) \right)$$
$$\stackrel{\mathcal{D}}{=} \left(a_T \max_{a_T \le u \le c_T} \left\| \frac{1}{u} \mathbf{W}^{(1)}(u) \right\|^2, b_T \max_{b_T \le u \le d_T} \left\| \frac{1}{u} \mathbf{W}^{(2)}(u) \right\|^2 \right),$$

where $\{\mathbf{W}^{(1)}(u), u \ge 0\}$ and $\{\mathbf{W}^{(2)}(u), u \ge 0\}$ are independent Wiener processes in $\mathbb{R}^d$, the result now follows as in lemma B.2 of the supplementary materials to Horvath et al. (2020). ■

Proof. The proof goes along the lines of Theorem 1 but the arguments are simpler since $\overline{\mathbf{Q}}_{T,t}$ only contains the first term of $\hat{\mathbf{Q}}_{T,t}$. Hence the details are omitted. ■

Proof of Theorem 2. First we assume that $a_T \le t^*$. We note that

$$\hat{\boldsymbol{\beta}}_{t^*,1} = \boldsymbol{\beta}^{(1)} + (\mathbf{X}_{t^*,1}^\top \mathbf{X}_{t^*,1})^{-1} \mathbf{X}_{t^*,1}^\top \mathbf{E}_{t^*,1} \quad \text{and} \quad \hat{\boldsymbol{\beta}}_{t^*,2} = \boldsymbol{\beta}^{(T)} + (\mathbf{X}_{t^*,2}^\top \mathbf{X}_{t^*,2})^{-1} \mathbf{X}_{t^*,2}^\top \mathbf{E}_{t^*,2}.$$

According to the proof of Theorem 1

$$r_T^{1/2} \| (\mathbf{X}_{t^*,1}^\top \mathbf{X}_{t^*,1})^{-1} \mathbf{X}_{t^*,1}^\top \mathbf{E}_{t^*,1} \| = O_P(1),$$

and

$$r_T^{1/2} \| (\mathbf{X}_{t^*,2}^\top \mathbf{X}_{t^*,2})^{-1} \mathbf{X}_{t^*,2}^\top \mathbf{E}_{t^*,2} \| = O_P(1).$$

Next we note that by the arguments used in Section 5

$$\| \hat{\mathbf{Q}}_{T,t^*}^{-1} \| = O_P(1).$$

Hence the first part of Theorem 2 is proven.

Next we assume that $t^* < a_T$. By the definition of $\hat{\boldsymbol{\beta}}_{a_T,1}$ and

$$\hat{\boldsymbol{\beta}}_{a_T,1} = (\mathbf{X}_{a_T,1}^\top \mathbf{X}_{a_T,1})^{-1} \left(\sum_{s=1}^{t^*} \mathbf{x}_s \mathbf{x}_s^\top \boldsymbol{\beta}^{(1)} + \sum_{s=t^*+1}^{a_T} \mathbf{x}_s \mathbf{x}_s^\top \boldsymbol{\beta}^{(T)} \right) + (\mathbf{X}_{a_T,1}^\top \mathbf{X}_{a_T,1})^{-1} \mathbf{X}_{a_T,1}^\top \mathbf{E}_{a_T,1},$$

and

$$\hat{\boldsymbol{\beta}}_{a_T,2} = \boldsymbol{\beta}^{(T)} + (\mathbf{X}_{a_T,2}^\top \mathbf{X}_{a_T,2})^{-1} \mathbf{X}_{a_T,2}^\top \mathbf{E}_{a_T,2}.$$

According to Theorem 1

$$r_T^{1/2}\|(\mathbf{X}_{a_T,1}^\top \mathbf{X}_{a_T,1})^{-1}\mathbf{X}_{a_T,1}^\top \mathbf{E}_{a_T,1}\| = O_P(1),$$

and

$$r_T^{1/2}\|(\mathbf{X}_{a_T,2}^\top \mathbf{X}_{a_T,2})^{-1}\mathbf{X}_{a_T,2}^\top \mathbf{E}_{a_T,2}\| = O_P(1).$$

Using the results of Section 4 we conclude that

$$\left\| (\mathbf{X}_{a_T,1}^\top \mathbf{X}_{a_T,1})^{-1}\left(\sum_{s=1}^{t^*}\mathbf{x}_s\mathbf{x}_s^\top \boldsymbol{\beta}_1 + \sum_{s=t^*+1}^{a_T}\mathbf{x}_s\mathbf{x}_s^\top \boldsymbol{\beta}_2\right) - \left(\frac{t^*}{a_T}\boldsymbol{\beta}^{(1)} + \frac{a_T - t^*}{a_T}\boldsymbol{\beta}^{(T)}\right)\right\|$$
$$= o_P\left(\frac{t^*}{a_T}\right) + o_P\left(\frac{a_T - t^*}{a_T}\right) = o_P\left(\frac{t^*}{a_T}\right),$$

on account of (10). Since by the results in Section 4 we have that

$$\|\hat{\mathbf{Q}}_{T,a_T}^{-1}\| = O_P(1),$$

the result is proven when $t^* < a_T$. ■

ACKNOWLEDGEMENTS

We wish to thank Mr. John Clements for suggesting the crude oil/U.S. dollar dataset. We also wish to thank three anonymous referees as well as the editor and associate editor for many helpful and constructive comments that helped us improve this paper significantly. The Third author was partially supported by the Natural Science and Engineering Research Council of Canada's Discovery grant, RGPIN 50503-10477.

ORCID

Gregory Rice https://orcid.org/0000-0002-1565-7828

REFERENCES

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, *59*, 817–858.

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

Arfaoui, M., & Rejeb, A. B. (2017). Oil, gold, US dollar and stock market interdependencies: A global analytical insight. *European Journal of Management and Business Economics*, *26*, 278–293.

Aschersleben, P., & Wagner, M. 2016. *cointReg: Parameter estimation and inference in a cointegrating regression*, R package version 0.2.0.

Aue, A., Hörmann, S., Horváth, L., & Hušková, M. (2014). Dependent functional linear models with applications to monitoring structural change. *Statistica Sinica*, *24*, 1043–1073.

Aue, A., & Horváth, L. (2013). Structural breaks in time series. *Journal of Time Series Analysis*, *23*, 1–16.

Bai, J. (1999). Likelihood ratio tests for multiple structural changes. *Journal of Econometrics*, *91*, 299–323.

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

Bardsley, P., Horváth, L., Kokoszka, P., & Young, G. (2017). Change point tests in functional factor models with application to yield curves. *Econometrics Journal*, *20*, 373–403.

Beckmann, J., Czudaj, R., & Vipin, A. (2017). *The Relationship Between Oil Prices and Exchange Rates: Theory and Evidence*. US Energy Information Administration.

Berkes, I., Horváth, L., Kokoszka, P. and and Shao, Q-M.: On discriminating between long–range dependence and changes in the mean. *Annals of Statistics 34*(2006) 1140–1165.

Berkes, I., Horváth, L., Kokoszka, P. and and Shao, Q–M.: Almost sure convergence of the Bartlett estimator. *Periodica Mathematica Hungarica 31*(2005) 11–25.

Billingsley, P. (1968). *Convergence of probability measures*. Wiley.

Brown, R. L., Durbin, J., & Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society, Series B*, *37*, 149–192.

Csörgő, M., & Horváth, L. (1997). *Limit theorems in change–point analysis*. Wiley.

Gombay, E., & Horváth, L. (1994). Limit theorems for change in linear regression. *Journal of Multivariate Analysis*, *48*, 43–69.

Górecki, T., Horváth, L., & Kokoszka, P. (2017). Change point detection in heteroscedastic time series. *Econometrics & Statistics*, *20*, 86–117.

Harris, D., Kew, H., & Taylor, A. M. 2017. *Level shift estimation in the presence of non–stationary volatility with an application to the unit root testing problem*. Essex Finance Centre Working Papers 20329, University of Essex, Essex Business School.

Harvey, D., Leybourne, S., Sollis, R., & Taylor, A. M. (2016). Tests for explosive financial bubbles in the presence of non–stationary volatility. *Journal of Empirical Finance*, *38*, 548–574.

Hidalgo, J., & Seo, M. H. (2013). Testing for structural stability in the whole sample. *Journal of Econometrics*, *175*, 84–93.

Horváth, L., Miller, C., & Rice, G. (2020). A new class of change point test statistics of Rényi type. *Journal of Business and Economic Statistics*, *38*, 570–579.

Ji, Q., Liu, B.–. Y., & Fan, Y. (2019). Risk dependence of CoVaR and structural change between oil prices and exchange rates: A time-varying copula model. *Energy Economics*, *77*, 80–92.

Kim, H.–. J., & Siegmund, D. (1989). The likelihood ratio test for a change–point in simple linear regression. *Biometrika*, *76*, 409–423.

Krugman, P. (1980). *Oil and the dollar*. National Bureau of Economic Research.

Krugman, P. (1983). *Oil shocks and exchange rate dynamics*. In J. A. Frenkel (Ed.), *Exchange rates and international macroeconomics* (pp. 259–284). University of Chicago Press.

Novet, J. (2019). Zoom rocketed 72% on the first day of trading. *CNBC*. https://www.cnbc.com/2019/04/18/zoom-ipo-stock-begins-trading-on-nasdaq.html.

Ploberger, W., & Krämer, W. (1992). The CUSUM test with OLS residuals. *Econometrica*, *60*, 271–285.

Quandt, R. E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American Statististical Association*, *55*, 324–330.

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

Wu, W., & Zhou, Z. (2018). Gradient-based structural change detection for nonstationary time series M-estimation. *Annals of Statistics*, *46*, 1197–1224.

How to cite this article: Horváth L, Miller C, Rice G. Detecting early or late changes in linear models with heteroscedastic errors. *Scand J Statist*. 2021;48:577–609. https://doi.org/10.1111/sjos.12507