Detecting At-Most-m-Changes in Linear Regression Models

Horvath, Lajos; Pouliot, William; Wang, Shixuan

DOI:
10.1111/jtsa.12228

.

DETECTING AT-MOST-$\mathfrak{m}$ CHANGES IN LINEAR REGRESSION MODELS

LAJOS HORVÁTH, WILLIAM POULIOT, AND SHIXUAN WANG

Abstract. In this paper we provide a new procedure to test for at-most-$\mathfrak{m}$ changes in the time–dependent regression model $y_t = \mathbf{x}_t^\top \boldsymbol{\beta}_t + e_t, 1 \le t \le T$, i.e. $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \ldots = \boldsymbol{\beta}_T$ under the no change null hypothesis against the alternative $y_t = \boldsymbol{x}_t^\top \boldsymbol{\beta}^{(i)} + e_t$, if $k_{i-1}^* < t \le k_i^*, 1 \le i \le \mathfrak{m}+1$ and $\boldsymbol{\beta}^{(j)} \ne \boldsymbol{\beta}^{(\ell)}$ for some $1 \le j, \ell \le \mathfrak{m}+1$ with $k_0^* = 0, 1 < k_1^* < k_2^* < \ldots < k_{\mathfrak{m}}^* < T, k_{\mathfrak{m}+1}^* = T$. Our procedure is based on weighted sums of the residuals, incorporating the possibility of $\mathfrak{m}$ changes. The weak limit of the proposed test statistic is the sum of two double exponential random variables. A small Monte Carlo simulation illustrates the applicability of the limit results in case of small and moderate sample sizes. We compare the new method to the CUSUM and standardized (weighted) CUSUM procedures and obtain the power curves of the test statistics under the alternative. We apply our method to find changes in the unconditional four factor CAPM.

1. Introduction

In the paper we are interested in the time–dependent regression model

$$y_t = \mathbf{x}_t^\top \boldsymbol{\beta}_t + e_t, \quad 1 \le t \le T. \tag{1.1}$$

We wish to test the null hypothesis of constant $\boldsymbol{\beta}_t$'s

$$H_0: \quad \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \ldots = \boldsymbol{\beta}_T$$

against the at-most-$\mathfrak{m}$ change points alternative. With the notations $k_0^* = 0$ and $k_{\mathfrak{m}+1}^* = T$, the case of at-most-$\mathfrak{m}$ changes alternative can be formulated as

$$\begin{aligned} H_A: \quad y_t =& \boldsymbol{x}_t^\top \boldsymbol{\beta}^{(i)} + e_t, \quad \text{if } k_{i-1}^* < t \le k_i^*, \ 1 \le i \le \mathfrak{m}+1 \text{ and } \boldsymbol{\beta}^{(j)} \ne \boldsymbol{\beta}^{(\ell)} \\ & \text{for some } 1 \le j, \ell \le \mathfrak{m}+1. \end{aligned}$$

Testing for possible changes was initiated by Quandt (1958, 1960) who suggested maximally selected statistics and provided practical advise how to get critical values. Gombay and Horváth (1994), Horváth (1995) and Horváth and Shao (1995) obtained the limit distributions of some of the test statistics proposed by Quandt (1958, 1960) including maximally selected F–statistics and the likelihood ratio. McCabe and Harrison (1980) also contribute to this literature and advise the use of ordinary least squares residuals rather than recursive in CUSUM-type tests. Later McCabe (1988), using a multiple decision theory approach, shows that the CUSUM test is Bayes for structural stability in scale and variance models, and also that the CUSUM-of-squares test is a localised Bayes rule for structural stability in variance of linear regression models. Turning to estimation of the time change of change,

1991 *Mathematics Subject Classification.* 62E29, 60F05, 62J05.

Key words and phrases. change point, Bernoulli shifts, weak approximation, weighted CUSUM, residuals.

The research of Lajos Horváth was supported by NSF grant DMS 1305858 and a Visiting Fellowship at the Department of Economics, University of Birmingham, UK. The research of Shixuan Wang was supported by the Economic and Social Research Council, UK, grant ES/J50001X/1 and a Royal Economic Society Junior Fellowship.

Hušková (1996) gave large sample approximation for the estimator of the time of change assuming that we have exactly one change in the regressor during the observation period. The independence of the error terms are assumed in these early papers. Andrews (1993) provides a general methodology to test for the stability of random systems from an economic view point. Ghysels et al. (1997), Bai (1999), Bai and Perron (1998), Hall et al. (2012) followed the suggestions of Andrews (1993) and they also used the maximally selected statistics but the maxima were not computed for all observations points, a fraction of early and late observations were excluded. Aue et al. (2008, 2012a) used the maximally selected likelihood ratio method to test for stability of the parameter against exactly one change. However, they also showed that the derived tests are consistent against several changes under the alternative.

Our test for H_0 against H_A uses the residuals

$$\hat{e}_t = y_t - \mathbf{x}_t^\top \hat{\boldsymbol{\beta}}_T, \quad 1 \le t \le T, \tag{1.2}$$

where

$$\hat{\boldsymbol{\beta}}_T = (\mathbf{X}_T \mathbf{X}_T^\top)^{-1} \mathbf{X}_T^\top \mathbf{Y}_T,$$

with

$$\mathbf{Y}_T = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} \quad \text{and} \quad \mathbf{X}_T = \begin{bmatrix} x_{1,1} & x_{2,1} & \cdots & x_{T,1} \\ x_{1,2} & x_{2,2} & \cdots & x_{T,2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1,d} & x_{2,d} & \cdots & x_{T,d} \end{bmatrix}.$$

In this paper we suggest three test statistics based on the sums of the residuals. The classical CUSUM statistic

$$D_T = T^{-1/2} \max_{1 \le \ell \le T} \left| \sum_{t=1}^{\ell} \hat{e}_t - \frac{\ell}{T} \sum_{t=1}^{T} \hat{e}_t \right|$$

which together with the standardized CUSUM

$$H_T = T^{1/2} \max_{1 \le \ell < T} \left(\ell (T - \ell) \right)^{-1/2} \left| \sum_{t=1}^{\ell} \hat{e}_t - \frac{\ell}{T} \sum_{t=1}^{T} \hat{e}_t \right|$$

are one of the most often used statistics in change point analysis. Aue and Horváth (2013) contains a review of change point detection in time series and it also provides a historical account of CUSUM procedures. Horváth and Rice (2014) explain how mathematical and probabilistic tools can be used to extend classical change point methods to time series models. The motivation for H_T and D_T is based on the likelihood ratio method when there is exactly one change in the parameters under the alternative. Since we allow up to $\mathfrak{m}$ changes under the alternative, we propose a modification of the CUSUM statistics. Let

$$M(k_1, \ldots, k_{\mathfrak{m}}) = |M_1(k_1)| + |M_2(k_1, k_2)| + \ldots + |M_{\mathfrak{m}}(k_{\mathfrak{m}-1}, k_{\mathfrak{m}})| + |M_{\mathfrak{m}+1}(k_{\mathfrak{m}})|,$$

$1 \le k_1 \le k_2 \le \ldots \le k_{\mathfrak{m}} < T$, where

$$M_1(k_1) = \frac{1}{\sqrt{k_1}} \left(\sum_{t=1}^{k_1} \hat{e}_t - \frac{k_1}{T} \sum_{t=1}^{T} \hat{e}_t \right),$$

$$M_i(k_{i-1}, k_i) = \frac{1}{\sqrt{T}} \left(\sum_{t=k_{i-1}+1}^{k_i} \hat{e}_t - \frac{k_i - k_{i-1}}{T} \sum_{t=1}^{T} \hat{e}_t \right), \quad 2 \le i \le \mathfrak{m},$$

and

$$M_{\mathfrak{m}+1}(k_{\mathfrak{m}}) = \frac{1}{\sqrt{T-k_{\mathfrak{m}}}}\left(\sum_{t=k_{\mathfrak{m}}+1}^{T}\hat{e}_t - \frac{T-k_{\mathfrak{m}}}{T}\sum_{t=1}^{T}\hat{e}_t\right)$$

and define

$$M_T = \max_{\mathfrak{M}} M(k_1, k_2, \ldots, k_{\mathfrak{m}}), \quad \text{where } \mathfrak{M} = \{1 \le k_1 \le k_2 \le \ldots \le k_{\mathfrak{m}} < T\}$$

We would like to note that $M(k_1, k_2 \ldots, k_{\mathfrak{m}}), 1 \le k_1 \le k_2 \le \ldots \le k_{\mathfrak{m}} < T$ is also a generalization of the classical CUSUM process (cf. Csörgő and Horváth (1997)), since $M_1(k_1)$ and $M_{\mathfrak{m}+1}(k_{\mathfrak{m}})$, $1 \le k_1, k_{\mathfrak{m}} < T$ are standardized CUSUM processes starting from the first and the last residual, respectively. The components $M_1(k_1)$ and $M_{\mathfrak{m}+1}(k_{\mathfrak{m}})$ of $M(k_1, k_2, \ldots, k_{\mathfrak{m}})$ are self–normalized and therefore they could be derived from a likelihood argument. First we obtain the joint asymptotic distribution of D_T, H_T and M_T in Theorem 2.1. In Theorem 2.2 we derive the joint limit distribution of $M_T^*, M_T^{(1)}$ and $M_T^{(2)}$, where

$$M_T^* = \max_{\mathfrak{M}^*}\sum_{j=2}^{\mathfrak{m}} M_j(k_{j-1}, k_j) \quad \text{with} \quad \mathfrak{M}^* = \{1 < k_2 \le \ldots \le k_{\mathfrak{m}} < T\},$$

$$M_T^{(1)} = \max_{1\le k\le T} M_1(k) \quad \text{and} \quad M_T^{(2)} = \max_{1\le k<T} M_{\mathfrak{m}+1}(k).$$

Due to the standardization, the limit distributions of $M_1(k_1)$ and $M_{\mathfrak{m}+1}(k_{\mathfrak{m}})$ are non–standard, they do not follow from weak convergence type results. For the application of the Lagrange multiplier type statistics using the whole sample we refer to Hidalgo and Seo (2013). Jeng (2015) surveys CUSUM and related procedures in financial applications. We also discuss the behavior of M_T under the alternative H_A.

In this paper the test statistics are based on the residuals $\hat{e}_t, 1 \le t \le T$ but in a similar matter we can use the weighted residuals $\tilde{\mathbf{e}}_t = \mathbf{x}_t\hat{e}_t, 1 \le t \le T$ (cf. Hušková (1996)). Analogously to D_T and H_T one can define

$$\tilde{D}_T = \frac{1}{\sqrt{T}}\max_{1\le\ell\le T}\left(\left(\sum_{t=1}^{\ell}\tilde{\mathbf{e}}_t - \frac{\ell}{T}\sum_{t=1}^{T}\tilde{\mathbf{e}}_t\right)^{\top}\mathbf{\Sigma}^{-1}\left(\sum_{t=1}^{\ell}\tilde{\mathbf{e}}_t - \frac{\ell}{T}\sum_{t=1}^{T}\tilde{\mathbf{e}}_t\right)\right)^{1/2}$$

and

$$\tilde{H}_T^2 = \max_{1\le\ell<T}\left(\frac{T}{T(T-\ell)}\left(\sum_{t=1}^{\ell}\tilde{\mathbf{e}}_t - \frac{\ell}{T}\sum_{t=1}^{T}\tilde{\mathbf{e}}_t\right)^{\top}\mathbf{\Sigma}^{-1}\left(\sum_{t=1}^{\ell}\tilde{\mathbf{e}}_t - \frac{\ell}{T}\sum_{t=1}^{T}\tilde{\mathbf{e}}_t\right)\right)^{1/2},$$

where

$$\mathbf{\Sigma} = \sum_{t=-\infty}^{\infty} E\left[\mathbf{x}_0 e_0 (\mathbf{x}_t e_t)^{\top}\right],$$

the long run covariance matrix of the sum of the weighted innovations $\mathbf{x}_t e_t$. Now we define

$$\tilde{M}(k_1, \ldots, k_{\mathfrak{m}}) = \tilde{M}_1^{1/2}(k_1) + \tilde{M}_2^{1/2}(k_1, k_2) + \ldots + \tilde{M}_{\mathfrak{m}}^{1/2}(k_{\mathfrak{m}-1}, k_{\mathfrak{m}}) + \tilde{M}_{\mathfrak{m}+1}^{1/2}(k_{\mathfrak{m}}),$$

$1 \le k_1 \le k_2 \le \ldots \le k_{\mathfrak{m}} < T$, where

$$\tilde{M}_1(k_1) = \frac{1}{k_1}\left(\sum_{t=1}^{k_1}\tilde{\mathbf{e}}_t - \frac{k_1}{T}\sum_{t=1}^{T}\tilde{\mathbf{e}}_t\right)^\top \mathbf{\Sigma}^{-1}\left(\sum_{t=1}^{k_1}\tilde{\mathbf{e}}_t - \frac{k_1}{T}\sum_{t=1}^{T}\tilde{\mathbf{e}}_t\right),$$

$$\tilde{M}_i(k_{i-1},k_i) = \frac{1}{T}\left(\sum_{t=k_{i-1}+1}^{k_i}\tilde{\mathbf{e}}_t - \frac{k_i-k_{i-1}}{T}\sum_{t=1}^{T}\tilde{\mathbf{e}}_t\right)^\top \mathbf{\Sigma}^{-1}\left(\sum_{t=k_{i-1}+1}^{k_i}\tilde{\mathbf{e}}_t - \frac{k_i-k_{i-1}}{T}\sum_{t=1}^{T}\tilde{\mathbf{e}}_t\right), \quad 2 \le i \le \mathfrak{m},$$

and

$$\tilde{M}_{\mathfrak{m}+1}(k_{\mathfrak{m}}) = \frac{1}{T-k_{\mathfrak{m}}}\left(\sum_{t=k_{\mathfrak{m}}+1}^{T}\tilde{\mathbf{e}}_t - \frac{T-k_{\mathfrak{m}}}{T}\sum_{t=1}^{T}\tilde{\mathbf{e}}_t\right)^\top \mathbf{\Sigma}^{-1}\left(\sum_{t=k_{\mathfrak{m}}+1}^{T}\tilde{\mathbf{e}}_t - \frac{T-k_{\mathfrak{m}}}{T}\sum_{t=1}^{T}\tilde{\mathbf{e}}_t\right).$$

The statistics $\tilde{D}_T$, $\tilde{H}_T$ and $\max_{1\le k_1\le k_2\le\ldots\le k_{\mathfrak{m}}<T}\tilde{M}(k_1,k_2,\ldots,k_{\mathfrak{m}})$ can also be applied to to test H_0 against H_A. The derivation of their asymptotic properties can be the subject of future research.

2. Assumptions and Main Results

The vectors $\{\mathbf{x}_t, e_t, -\infty < t < \infty\}$ form a stationary time series. Our main assumption is that the sequence is a Bernoulli shift which can be approximated with finitely dependent time series. Let $\|\cdot\|$ denote the Euclidean norm of vectors and matrices.

Assumption 2.1. *The sequence $\{\mathbf{x}_t, e_t - \infty < t < \infty\}$ is a Bernoulli shift, i.e. there are measurable functionals $\mathbf{g}$ and f such that $\mathbf{x}_t = \mathbf{g}(\varepsilon_t, \varepsilon_{t-1}, \ldots)$ and $e_t = f(\varepsilon_t, \varepsilon_{t-1}, \ldots)$, where $\{\varepsilon_t, -\infty < t < \infty\}$ are independent and identically distributed random variables in some space. Also,*

$$\mathbb{E}e_0 = 0,\ \mathbb{E}|e_0|^\nu < \infty,\ \ \mathbb{E}e_0x_{0,i} = 0 \ \text{ and } \ \mathbb{E}\|x_{0,i}\|^\nu < \infty \ \text{ with some } \ \nu > 4, \tag{2.1}$$

and

$$\left(\mathbb{E}\left|e_{t,\ell} - e_t\right|^\nu\right)^{1/\nu} = O(\ell^{-\alpha}) \text{ and } \left(\mathbb{E}\left\|\mathbf{x}_{t,\ell} - \mathbf{x}_t\right\|^\nu\right)^{1/\nu} = O(\ell^{-\alpha}) \text{ with some } \alpha > 2, \tag{2.2}$$

where $e_{t,\ell} = f(\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-\ell}, \varepsilon_{t,\ell,t-\ell-1}, \varepsilon_{t,\ell,t-\ell-2}, \ldots)$,
$\boldsymbol{x}_{t,\ell} = \mathbf{g}(\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-\ell}, \varepsilon_{t,\ell,t-\ell-1}, \varepsilon_{t,\ell,t-\ell-2}, \ldots)$ and the $\varepsilon_{i,j,k}$'s are independent and identically distributed copies of ε_0.

The Bernoulli shifts $\mathbf{x}_{t,m}$ and $e_{t,m}$ are random variables that closely approximate $\mathbf{x}_t$ and e_t in the sense specified in Assumption 2.1. They used to establish some of the theorems that follow. Assumption 2.1 implies immediately that $e_t, \mathbf{x}_t, -\infty < t < \infty$ is a stationary sequence. For results on change point detection in linear models with nonstationary errors we refer to Hansen (1992), Busetti and Taylor (2004), Harvey et al. (2006), Cavaliere and Taylor (2008) and Kejriwal and Perron (2008).
We prove in Lemma A.1 that

$$\frac{1}{T}\mathbf{X}_T\mathbf{X}_T^\top \to \mathbf{A} \ \text{ a.s.}$$

The next assumption postulates that $\mathbf{A}^{-1}$ exists.

Assumption 2.2. **A** *is a nonsingular matrix.*

Let

$$\sigma^2 = \mathbb{E}e_0^2 + 2\sum_{\ell=1}^{\infty} \mathbb{E}e_0 e_\ell. \tag{2.3}$$

We show in the proof of Lemma A.3 that $\sigma^2 < \infty$. To state our main result we need to introduce further notations. The random variables ξ_1 and ξ_2 are double exponential random variables, i.e.

$$\xi_1 \text{ and } \xi_2 \text{ are independent and } \mathbb{P}\{\xi_1 \le x\} = \mathbb{P}\{\xi_2 \le x\} = \exp(-e^{-x}) \text{ for all } x, \tag{2.4}$$

and define the numerical sequences

$$a_T = (2\log\log T)^{1/2} \quad \text{and} \quad b_T = 2\log\log T + \frac{1}{2}\log\log\log T - \frac{1}{2}\log\pi.$$

Theorem 2.1. *If H_0 and Assumptions 2.1 and 2.2 hold, then we have*

$$\left(\frac{D_T}{\sigma}, a_T\frac{H_T}{\sigma} - b_T, a_T\frac{M_T}{\sigma} - 2b_T\right) \xrightarrow{\mathcal{D}} \left(\sup_{0\le t\le 1}|B(t)|, \max(\xi_1,\xi_2), \xi_1+\xi_2\right), \tag{2.5}$$

where $\{B(t), 0 \le t \le 1\}$ is a Brownian bridge, independent of ξ_1 and ξ_2 defined by (2.4).

Next we provide the joint asymptotic behaviour $M_T^*, M_T^{(1)}$ and $M_T^{(2)}$.

Theorem 2.2. *If H_0 and Assumptions 2.1 and 2.2 hold, then we have*

$$\left(\frac{M_T^*}{\sigma}, a_T\frac{M_T^{(1)}}{\sigma} - b_T, a_T\frac{M_T^{(2)}}{\sigma} - b_T\right) \xrightarrow{\mathcal{D}} \left(\bar{B}, \xi_1, \xi_2\right), \tag{2.6}$$

with

$$\bar{B} = \sup_{0\le u_1\le u_2\le\ldots\le u_{\mathfrak{m}}\le 1} \sum_{j=2}^{\mathfrak{m}} |B(u_j) - B(u_{j-1})|,$$

$u_0 = 0$, *where $\{B(t), 0 \le t \le 1\}$ is a Brownian bridge, independent of ξ_1 and ξ_2 defined by* (2.4).

We demonstrate via Monte Carlo simulations in Section 3 that the properties of D_T, H_T and M_T are different under the alternative and the power of these tests depend on the location of the change point(s). Theorem 2.1 makes it possible to combine the three tests to increase the power.

The norming sequences a_T and b_T are simple from a theoretical point of view but they are not the best choice in small to moderate sample sizes. Hence we provide an alternative version of (2.5). Let

$$a_{\phi,T} = (2\log\log[T(\log T)^\phi])^{1/2} \tag{2.7}$$

and

$$b_{\phi,T} = 2\log\log[T(\log T)^\phi] + \frac{1}{2}\log\log\log(T(\log T)^\phi) - \frac{1}{2}\log\pi, \tag{2.8}$$

where $-\infty < \phi < \infty$.

Theorem 2.3. *If H_0 and Assumptions 2.1 and 2.2 hold, then we have for all $-\infty < \phi < \infty$ that*

$$a_{\phi,T}\frac{H_T}{\sigma} - b_{\phi,T} \xrightarrow{\mathcal{D}} \max(\xi_1, \xi_2), \tag{2.9}$$

and

$$a_{\phi,T}\frac{M_T}{\sigma} - 2b_{\phi,T} \xrightarrow{\mathcal{D}} \xi_1 + \xi_2, \tag{2.10}$$

where ξ_1 and ξ_2 are defined in (2.4).

We discuss the choice of ϕ in Section 3.

Next we study the consistency of testing procedures based on Theorem 2.1. Let

$$\tilde{\boldsymbol{\beta}}_T = \sum_{\ell=1}^{\mathfrak{m}+1} \frac{k_\ell^* - k_{\ell-1}^*}{T}\boldsymbol{\beta}^{(\ell)}$$

and

$$J_T = \sqrt{k_1^*}\left|\mathbf{c}^\top(\boldsymbol{\beta}^{(1)} - \tilde{\boldsymbol{\beta}}_T)\right| + \sum_{i=2}^{\mathfrak{m}} \frac{k_i^* - k_{i-1}^*}{\sqrt{T}}\left|\mathbf{c}^\top(\boldsymbol{\beta}^{(i)} - \tilde{\boldsymbol{\beta}}_T)\right| + \sqrt{T - k_{\mathfrak{m}}^*}\left|\mathbf{c}^\top(\boldsymbol{\beta}^{(\mathfrak{m}+1)} - \tilde{\boldsymbol{\beta}}_T)\right|,$$

where $\mathbb{E}\mathbf{x}_0 = \mathbf{c}$.

Theorem 2.4. *We assume that H_A and Assumptions 2.1 and 2.2 are satisfied.*
(i) If

$$J_T \to \infty, \tag{2.11}$$

then we have that

$$\frac{D_T}{\sigma} \xrightarrow{\mathbb{P}} \infty. \tag{2.12}$$

(ii) If

$$(\log\log T)^{-1/2} J_T \to \infty, \tag{2.13}$$

then we have that

$$(\log\log T)^{-1/2}\frac{H_T}{\sigma} \xrightarrow{\mathbb{P}} \infty. \tag{2.14}$$

and

$$(\log\log T)^{-1/2}\frac{M_T}{\sigma} \xrightarrow{\mathbb{P}} \infty. \tag{2.15}$$

Assumptions (2.11) and (2.13) quantify the relationship between the locations and the sizes of the changes. If the change is early, i.e. $k_1^*/T \to 0$, then the size of the change at k_1^* should be relatively larger to be detected than if the change occurs in the middle of the data. The same comment holds for a late change, i.e. when $k_{\mathfrak{m}}^*/T \to 1$.
The extra $(\log\log T)^{-1/2}$ term in (2.13) are needed since the variables H_T and M_T are increasing to infinity with rate $(\log\log T)^{1/2}$ under the null hypothesis.

Next we consider two immediate consequences of Theorem 2.4. Let $\bar{\delta}^{(i)} = \mathbf{c}^\top(\boldsymbol{\beta}^{(i+1)} - \boldsymbol{\beta}^{(i)}), 1 \le i \le \mathfrak{m}$ denote the size of the change at k_i^*.

Corollary 2.1. *We assume that H_A, Assumptions 2.1 and 2.2 are satisfied and*

$$\lim_{T\to\infty} \frac{k_i^*}{T} = \theta_i \ \text{ and } \ 0 < \theta_1 < \theta_2 < \ldots < \theta_{\mathfrak{m}} < 1 \tag{2.16}$$

hold.
(i) If

$$T^{1/2} \max_{1\le i\le \mathfrak{m}} |\bar{\delta}^{(i)}| \to \infty,$$

then we have (2.12).
(ii) If

$$T^{1/2} (\log\log T)^{-1/2} \max_{1\le i\le \mathfrak{m}} |\bar{\delta}^{(i)}| \to \infty,$$

then we have (2.14) *and* (2.15).

Relation (2.16) means that the change occurs in the "middle" of the data. To illustrate the optimality of our results we consider a special case. We assume that $\mathfrak{m} = 1$, i.e. we have exactly one change and $\bar{\delta}$ denotes the size of the change.

Corollary 2.2. *We assume that H_A holds with $\mathfrak{m} = 1$ and Assumptions 2.1 and 2.2 are satisfied.*
(i) If

$$\left(\frac{k_1^*(T-k_1^*)}{T}\right)^{1/2} |\bar{\delta}| \to \infty$$

holds, then we have (2.12).
(ii) If

$$\left(\frac{k_1^*(T-k_1^*)}{T}\right)^{1/2} \frac{|\bar{\delta}|}{(\log\log T)^{1/2}} \to \infty$$

holds, then we have (2.14) *and* (2.15).

Conditions detailed in Corollary 2.2 are exactly the necessary and sufficient conditions for the consistency of the CUSUM and of the self–normalized CUSUM in case of independent and identically distributed errors (cf. Csörgő and Horváth (1997, p. 170–178)).

According to Corollaries 2.1and 2.2, we can detect changes if at least one of the changes is larger than $T^{-1/2}$ or $((\log\log T)/T)^{1/2}$, respectively, which also appeared as conditions for the consistency of CUSUM based tests (cf. Csörgő and Horváth (1997), Aue and Horváth (2013) and Horváth and Rice (2014)).

We show in Section A that

$$\sup_{0\le u\le 1} \left\| \frac{1}{T} \sum_{t=1}^{\lfloor Tu \rfloor} \mathbf{x}_t \mathbf{x}_t^\top - u\mathbf{A} \right\| \to 0 \quad \text{a.s.} \tag{2.17}$$

and

$$\frac{1}{T} \sum_{t=1}^{T} \|\mathbf{x}_t\|^\nu < \infty \quad \text{a.s. with some } \nu > 2. \tag{2.18}$$

The results of Theorems 2.1–2.4 remain true if we condition on $\mathbf{x}_t, -\infty < t < \infty$ assuming that (2.17) and (2.18) hold and $\mathbf{x}_t, -\infty < t < \infty$ and $e_t, -\infty < t < \infty$ are independent. Assumption (2.17) immediately rules out linear, polynomial, time trend and trigonometric regression. However, in these cases the likelihood method leads to weight functions different

from the square function in the definitions of H_T, $M_1(k_1)$ and $M_{\mathfrak{m}+1}(k_{\mathfrak{m}})$. We refer to Jarušková (1999, 2003) and Albin and Jarušková (2003) for the limit of the maximally selected likelihood ratio with changing trends and to Aue et. al. (2008, 2009, 2012b) for the more general case.

3. Finite sample performance

3.1. **Estimation of σ.** The long run variance of (2.3) is unknown and must be estimated from the sample. First we consider the case when the errors are uncorrelated, i.e.

Assumption 3.1.

$$\mathbb{E}e_t e_s = \begin{cases} 0, & \text{if } t \neq s \\ \sigma^2, & \text{if } t = s. \end{cases}$$

In case of uncorrelated errors we can use the sample variance

$$S_T^2 = \frac{1}{T}\sum_{t=1}^{T} \hat{e}_t^2. \tag{3.1}$$

Theorem 3.1. *We assume that Assumptions 2.1, 2.2 and 3.1 are satisfied.*
(i) If H_0 holds, then we have that

$$|S_T^2 - \sigma^2| = O_{\mathbb{P}}(T^{-1/2}). \tag{3.2}$$

(ii) If H_A holds, then we have that

$$S_T^2 = O_{\mathbb{P}}(1). \tag{3.3}$$

It follows immediately from (3.2) and (3.3) that the conclusions of Theorems 2.1–2.4 remain true when σ is replaced with S_T under Assumption 3.1.

If the errors are correlated we need to use a long run variance kernel estimator

$$\hat{\sigma}_T^2 = \hat{\gamma}_0 + 2\sum_{\ell=1}^{T-1} K(\ell/h)\hat{\gamma}_\ell, \tag{3.4}$$

where

$$\hat{\gamma}_\ell = \frac{1}{T}\sum_{t=1}^{T-\ell} \hat{e}_t \hat{e}_{t+\ell}$$

denotes the sample correlation of lag ℓ between the residuals. The kernel K and the window h satisfy the standard conditions:

Assumption 3.2. $K \geq 0$, $K(0) = 1$, $K(u) = 0$, *if* $|u| > c$ *with some* $c > 0$, *and* $K(u)$ *is Lipshitz continuous on the real line.*

We refer to Taniguchi and Kakizawa (2000) and Politis and Romano (1995) for discussion on the choice of $K(\cdot)$.

Assumption 3.3. $h = h(T) \to \infty$ *and* $h/T \to 0$.

Parzen (1957) points out that Assumption 3.3 is the necessary condition for the asymptotic consistency of the kernel based long run variance estimator. For the optimal choice of h we refer to Andrews (1991) and Newey and West (1994). Assumption 3.3 is sufficiently general that it includes the optimal windows specified in these references. The adaptive choice of h of Politis (2003) can also be used in our set up.

Theorem 3.2. *We assume that Assumptions 2.1, 2.2, 3.2 and 3.3 are satisfied.*
(i) If H_0 holds, then we have that

$$\hat{\sigma}_T^2 - \sigma^2 = O_{\mathbb{P}}\left(\left(\frac{h}{T}\right)^{1/2} + \frac{1}{h}\right). \tag{3.5}$$

(ii) If H_A holds, then we have that

$$\hat{\sigma}_T^2 = O_{\mathbb{P}}(1) \tag{3.6}$$

Under Assumption 3.3 the convergence of the first coordinate in (2.5),(2.6) and Theorem 2.4(i) remain true when σ is replaced with $\hat{\sigma}_T$. If the smoothing parameter h satisfies $(h \log\log T)/T \to 0$ and $(\log\log T)^{1/2}/h \to 0$ we can replace the theoretical σ with the estimator $\hat{\sigma}_T$ in Theorems 2.1, 2.2 and 2.4.

3.2. **Monte Carlo simulations under the null hypothesis.** To assess how well the asymptotic distributions detailed in Theorems 2.1 and 2.3 approximates the finite sample distributions, Monte Carlo simulations are performed. There are several results on the rate of convergence in the functional central limit theorem even in case of dependent variables, we only deal with the choice of the tuning parameter ϕ in Theorem 2.3. The choice of ϕ in (2.9) has been discussed in the literature already (cf. Csörgő and Horváth (1997) and Davis et al. (1995)) so we investigate the finite sample properties of (2.10) when σ is estimated. We consider independent standard normal, GARCH (1,1) and AR(1) errors for various sample sizes. In all cases we investigate, the choice of $\phi = 1$ gives the best results and therefore only those are reported. In our experiments $\{\mathbf{x}_t, 1 \le t \le T\}$ and $\{e_t, 1 \le t \le T\}$ are independent. Also, $\boldsymbol{\beta} = (0, 2)^\top$ and $\mathbf{x}_t = (1, x_{t,2})^\top$, where $x_{t,2}, 1 \le t \le T$ are independent and identically distributed random variables with $\mathbb{E}x_{t,2} = 1$ and $\mathrm{var}(x_{t,2}) = 1$. The outcomes of the simulations are based on 5,000 repetitions.

Example 3.1. First we consider the simplest case when the errors $\{e_t, 1 \le t \le T\}$ are independent and standard normal random variables. Since Assumption (3.1) holds, we used

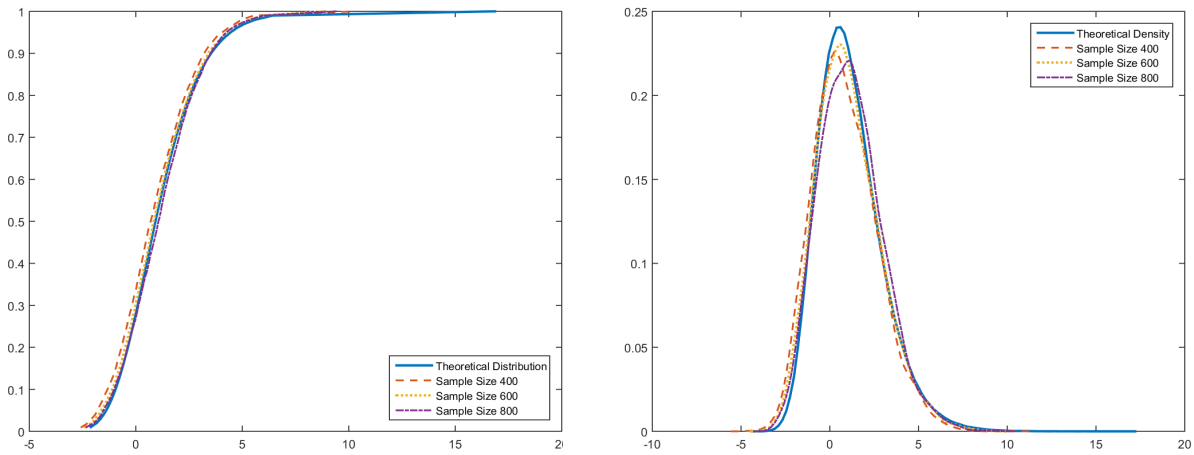
$$V_{T,1}^{(3)} = a_{1,T}\frac{M_T}{S_T} - 2b_{1,T}, \tag{3.7}$$

where $a_{1,T}$ and $b_{1,T}$ are defined in (2.7) and (2.8), and S_T^2 is the average of the squared residuals of (3.1). In Figure 3.1 we report the distribution and the density functions of $V_{T,1}^{(3)}$ for $T = 400, 600$ and 800. According to Figure 3.1, putting together Theorems 2.3 and 3.1, we obtain a good approximation for the distribution of the test statistic with the choice of $\phi = 1$.

Example 3.2. In the second study the errors e_t satisfy a GARCH (1,1) model, i.e.

$$e_t = v_t\varepsilon_t \quad \text{and} \quad v_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \alpha_2 v_{t-1}^2,$$

FIGURE 3.1. Plots of the distribution (left panel) and density functions (right panel) of $V_{T,1}^{(3)}$ for $T = 400, 600$ and 800 with the distribution and density function of $\xi_1 + \xi_2$



where the ε_t's are independent and identically distributed standard normal random variables. In our study, we used $\alpha_0 = .25, \alpha_1 = .25$ and $\alpha_2 = 0.5$. For a survey on GARCH and related processes, we refer to Francq and Zakoian (2010). It follows from Aue et al. (2014) that GARCH (1,1) with the present choice of parameters satisfies Assumption 2.1. Since GARCH (1,1) errors satisfy Assumption 3.1, one can use $V_{T,1}$ as test statistics. However, since

$$\sigma^2 = \frac{\alpha_0}{1 - \alpha_1 - \alpha_2},$$

we can use

$$\tilde{\sigma}_T^2 = \frac{\hat{\alpha}_{T,0}}{1 - \hat{\alpha}_{T,1} - \hat{\alpha}_{T,2}},$$

where $\hat{\alpha}_{T,0}, \hat{\alpha}_{T,1}$ and $\hat{\alpha}_{T,2}$ are the quasi maximum likelihood estimators for the GARCH parameters from the residuals $\hat{e}_t, 1 \le t \le T$ of (1.2). Using the basic properties of the quasi maximum likelihood estimators of the GARCH parameters discussed in Francq and Zakoian (2010), one can verify that

$$\left|\tilde{\sigma}_T^2 - \sigma^2\right| = O_{\mathbb{P}}(T^{-1/2}).$$

Hence Theorem 2.3 implies

$$V_{T,2}^{(3)} \xrightarrow{\mathcal{D}} \xi_1 + \xi_2,$$

where

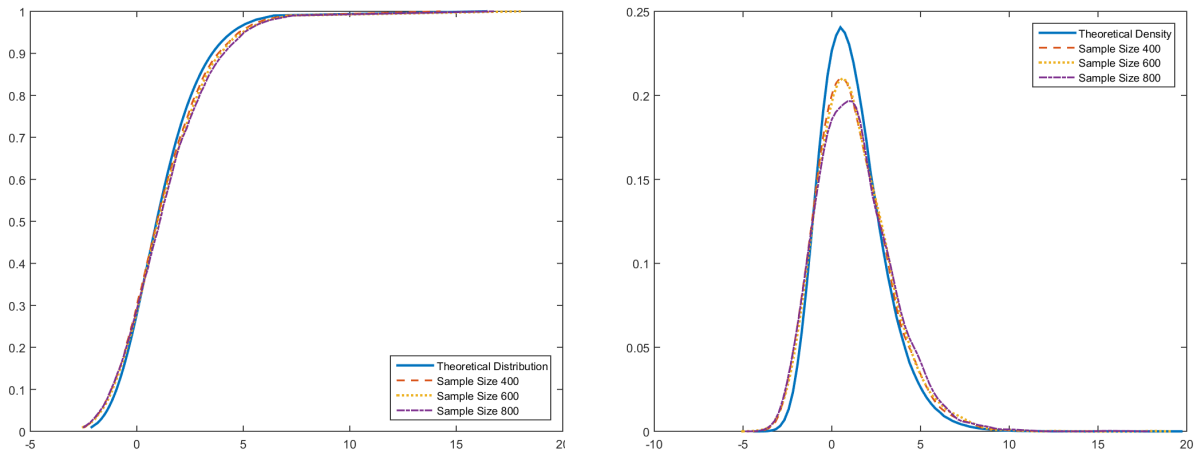
$$V_{T,2}^{(3)} = a_{1,T}\frac{M_T}{\tilde{\sigma}_T} - 2b_{1,T}.$$

The outcome of the Monte Carlo experiment is reported in Figure 3.2. Figure 3.2 shows that the dependence in the GARCH errors causes only minor difference compared to the case of independent and identically distributed e_t's.

Example 3.3. In our last experiment we simulated AR(1) errors:

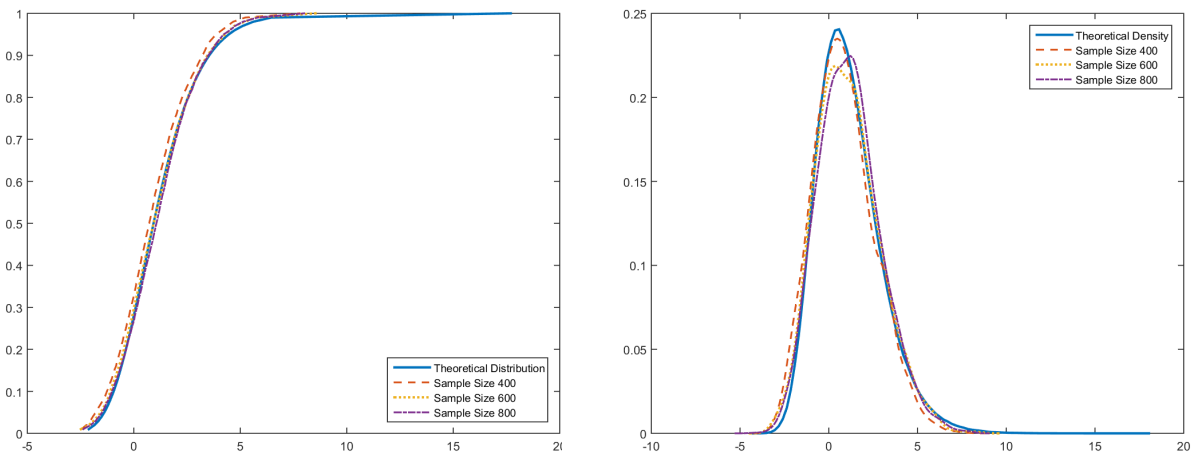
$$e_t = \frac{1}{2}e_{t-1} + \varepsilon_t,$$

FIGURE 3.2. Plots of the distribution (left panel) and density functions (right panel) of $V_{T,2}^{(3)}$ for $T = 400, 600$ and 800 with the distribution and density function of $\xi_1 + \xi_2$



where $\{\varepsilon_t, -\infty < t < \infty\}$ is a sequence of independent standard normal random variables.

FIGURE 3.3. Plots of the distribution (left panel) and density functions (right panel) of $V_{T,3}^{(3)}$ for $T = 400, 600$ and 800 with the distribution and density function of $\xi_1 + \xi_2$



We use now

$$V_{T,3}^{(3)} = a_{1,T} \frac{1}{\hat{\sigma}_T} M_T - 2b_{1,T},$$

where $\hat{\sigma}_T^2$ is the long run variance estimator of (3.4). We used the Bartlett kernel $K(x) = (1 - |x|)I\{|x| \leq 1\}$ and the window $h(T) = \lfloor 4(T/100)^{2/9} \rfloor + 1$ following the advise of Andrews and Monahan (1992). Due to the kernel estimation of the long run variance, we need somewhat larger sample sizes to achieve the same empirical accuracy as in the previous experiments and the critical values are slightly underestimated by the limit distribution.

3.3. **Monte Carlo simulations under the alternative hypothesis.** As in the numerical experiments under H_0, we assume that $\{\mathbf{x}_t, 1 \le t \le T\}$ and $\{e_t, 1 \le t \le T\}$ are independent, $\mathbf{x}_t = (1, x_{t,2})^\top$, where $x_{t,2}, 1 \le t \le T$ are independent and identically distributed normal random variables with $\mathbb{E}x_t = 1$, $\mathrm{var}(x_t) = 1$. and $e_1, e_2, \ldots$ are independent standard normal random variables. As previously, we used 5,000 repetitions.
We compare our method to the widely used maximally selected CUSUM statistic D_T/σ where σ is defined by (2.3). It is known that under mild conditions (cf. Aue and Horváth (2013)) that

$$\frac{D_T}{\sigma} \xrightarrow{\mathcal{D}} \sup_{0 \le t \le 1} |B(t)|, \tag{3.8}$$

where $B(t), 0 \le t \le 1$ denotes a Brownian bridge. Since σ is unknown, we estimate σ with S_T in case of independent and identically distributed errors resulting in

$$V_{T,1}^{(1)} = \frac{D_T}{S_T},$$

where S_T is the average of the squared residuals as in Example 3.1. Similarly, in the GARCH (1,1) model of Example 3.2 we use

$$V_{T,2}^{(1)} = \frac{D_T}{\tilde{\sigma}_T},$$

and

$$V_{T,3}^{(1)} = \frac{D_T}{\hat{\sigma}_T}$$

in case of a general stationary model (cf. Example 3.3). It follows from (3.8) and the discussions in Examples 3.1–3.3 that

$$V_{T,i}^{(1)} \xrightarrow{\mathcal{D}} \sup_{0 \le t \le 1} |B(t)|, \quad i = 1, 2, 3.$$

Similarly we introduce the statistics for AR(1) sequences

$$V_{T,1}^{(2)} = a_{1,T}\frac{H_T}{S_T} - b_{1,T}, \quad V_{T,2}^{(2)} = a_{1,T}\frac{H_T}{\tilde{\sigma}_T} - b_{1,T} \quad \text{and} \quad V_{T,3}^{(2)} = a_{1,T}\frac{H_T}{\hat{\sigma}_T} - b_{1,T}.$$

By Theorem 2.3 and Section 3.2 we have under H_0 that

$$V_{T,i}^{(2)} \xrightarrow{\mathcal{D}} \max(\xi_1, \xi_2) \quad i = 1, 2, 3.$$

First we consider the case when there is exactly one change in the parameter $\boldsymbol{\beta}_t$ at k_1^*.

Model I. We assume that $\mathfrak{m} = 1$ and

$$y_t = \begin{cases} \mathbf{x}_t^\top \boldsymbol{\beta}^{(1)} + e_t, & \text{if } 1 \le t \le k_1^*, \\ \mathbf{x}_t^\top \boldsymbol{\beta}^{(2)} + e_t, & \text{if } k_1^* + 1 \le t \le T. \end{cases}$$

Following Gombay (2010), we selected $\boldsymbol{\beta}^{(1)} = (0, 1)^\top$ and $\boldsymbol{\beta}^{(2)} = (0, 1 + \delta)^\top$, where $\delta = -2, -1.8, \ldots, 1.8, 2$. We considered three cases for the time of change $k_1^* = \lfloor T\theta_1 \rfloor$ where $\theta_1 = .2$ (early change), $\theta_2 = .5$ (change in the middle) and $\theta_2 = .9$ (late change).

Figures 3.4–3.6 exhibit the power functions of the statistics $V_{T,i}^{(j)}, 1 \le i, j \le 3$ in Model I. The power of $V_{T,i}^{(2)}$ and $V_{T,i}^{(3)}, i = 1, 2, 3$ are essentially the same but the CUSUM statistics

FIGURE 3.4. The power functions of the tests based on $V_{100,1}^{(1)}$ (upper panel), $V_{100,1}^{(2)}$ (lower left panel) and $V_{100,1}^{(3)}$ (lower right panel) in Model I under the conditions of Example 3.1

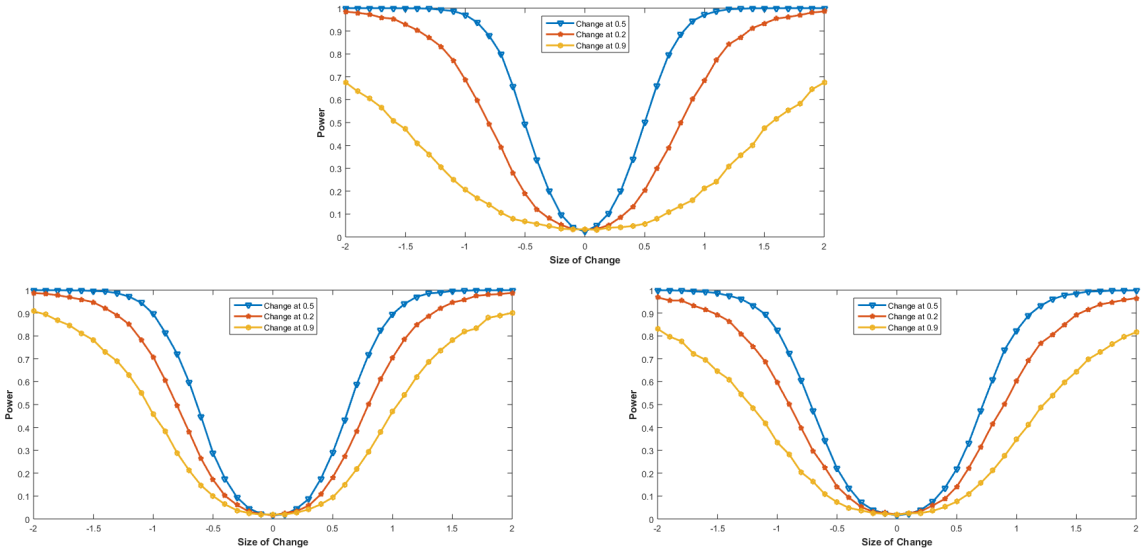
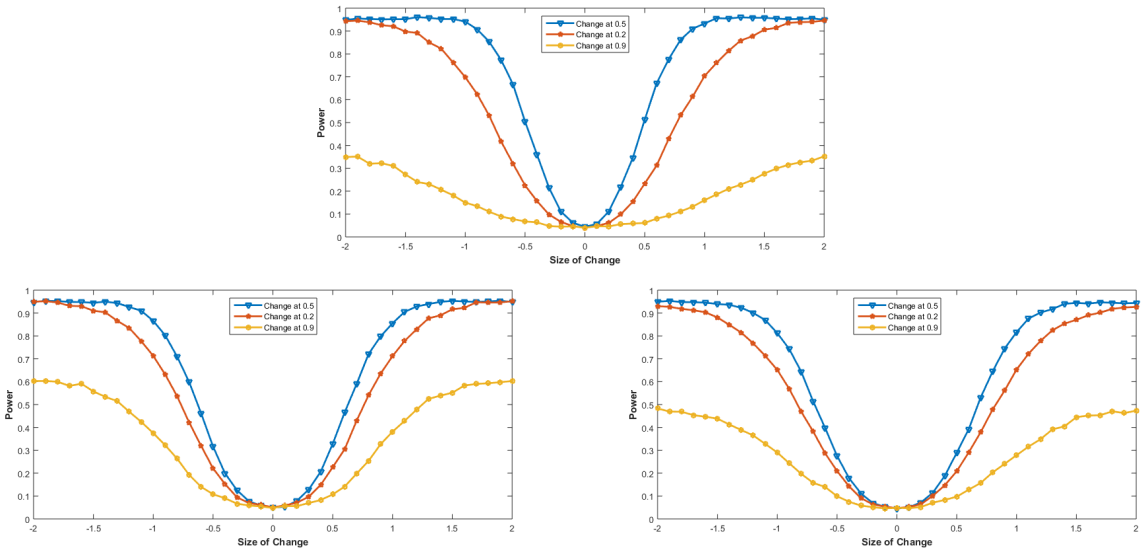


FIGURE 3.5. The power functions of the tests based on $V_{100,2}^{(1)}$ (upper panel), $V_{100,2}^{(2)}$ (lower left panel) and $V_{100,2}^{(3)}$ (lower right panel) in Model I under the conditions of Example 3.2



have higher power when the change occurs in the middle of the data.

FIGURE 3.6. The power functions of the tests based on $V_{400,3}^{(1)}$ (upper panel), $V_{400,3}^{(2)}$ (lower left panel) and $V_{400,3}^{(3)}$ (lower right panel) in Model I under the conditions of Example 3.3

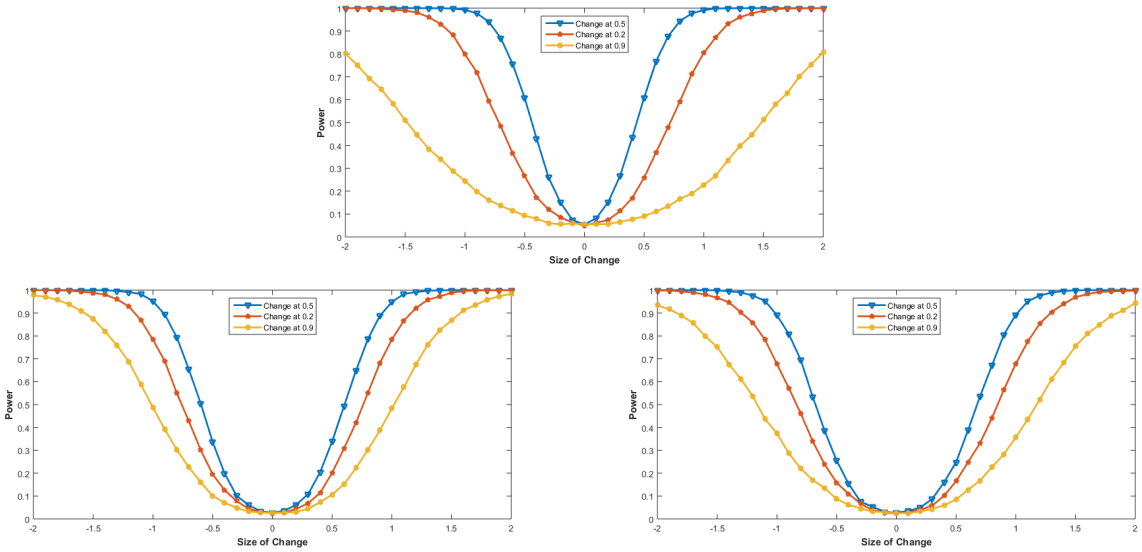
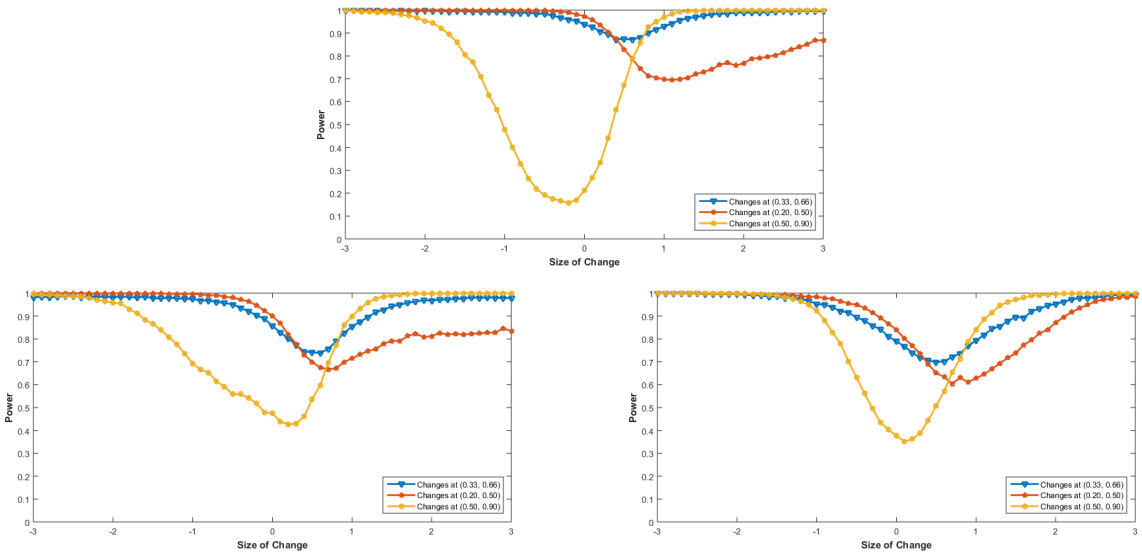


FIGURE 3.7. The power functions of the tests based on $V_{100,1}^{(1)}$ (upper panel), $V_{100,1}^{(2)}$ (lower left panel) and $V_{100,1}^{(3)}$ (lower right panel) in Model II under the conditions of Example 3.1

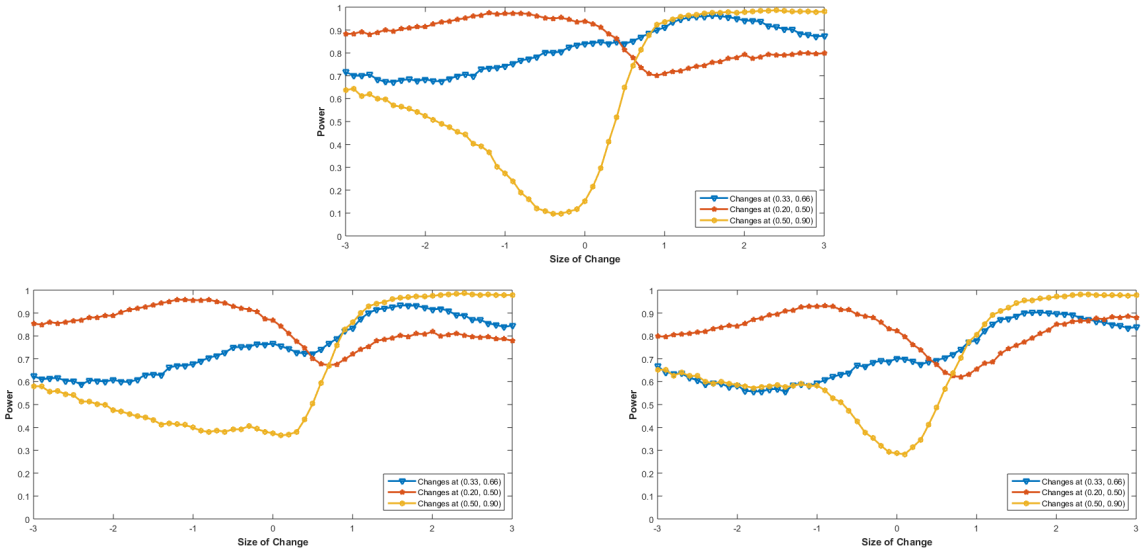


Model II. In this case $\mathfrak{m} = 2$ and

$$y_t = \begin{cases} \boldsymbol{x}_t^\top \boldsymbol{\beta}^{(1)} + e_t, & \text{if } 1 \le t \le k_1^* \\ \boldsymbol{x}_t^\top \boldsymbol{\beta}^{(2)} + e_t, & \text{if } k_1^* + 1 \le t \le k_2^* \\ \boldsymbol{x}_t^\top \boldsymbol{\beta}^{(3)} + e_t, & \text{if } k_2^* + 1 \le t \le T \end{cases}$$

with $\boldsymbol{\beta}^{(1)} = (0,1)^\top, \boldsymbol{\beta}^{(2)} = (0,1+\delta)^\top$, where $\delta = -3, -2.9, \ldots, 2.9, 3$ and $\boldsymbol{\beta}^{(3)} = (0,2)^\top$. The times of the changes in Figures 3.7–3.9 are $k_1^* = \lfloor T\theta_1 \rfloor$ and $k_2^* = \lfloor T\theta_2 \rfloor$, when $(\theta_1, \theta_2) = (.33, .66)$ (blue curves), $(.2, .5)$ (red curves) and $(.5, .9)$ (yellow curves). Due to the selection of the parameters, there is at least one change in Model II. As it is expected, the CUSUM statistics $V_{T,i}^{(1)}, i = 1,2,3$ have the lowest power nearly in all cases when the size of the change is small. Both $V_{T,i}^{(2)}$ and $V_{T,i}^{(3)}, i = 1,2,3$ have high power and $V_{T,i}^{(3)}, i = 1,2,3$ are better when the second change is late.

FIGURE 3.8. The power functions of the tests based on $V_{100,2}^{(1)}$ (upper panel), $V_{100,2}^{(2)}$ (lower left panel) and $V_{100,2}^{(3)}$ (lower right panel) in Model II under the conditions of Example 3.2



4. Change detection in the CAPM parameters

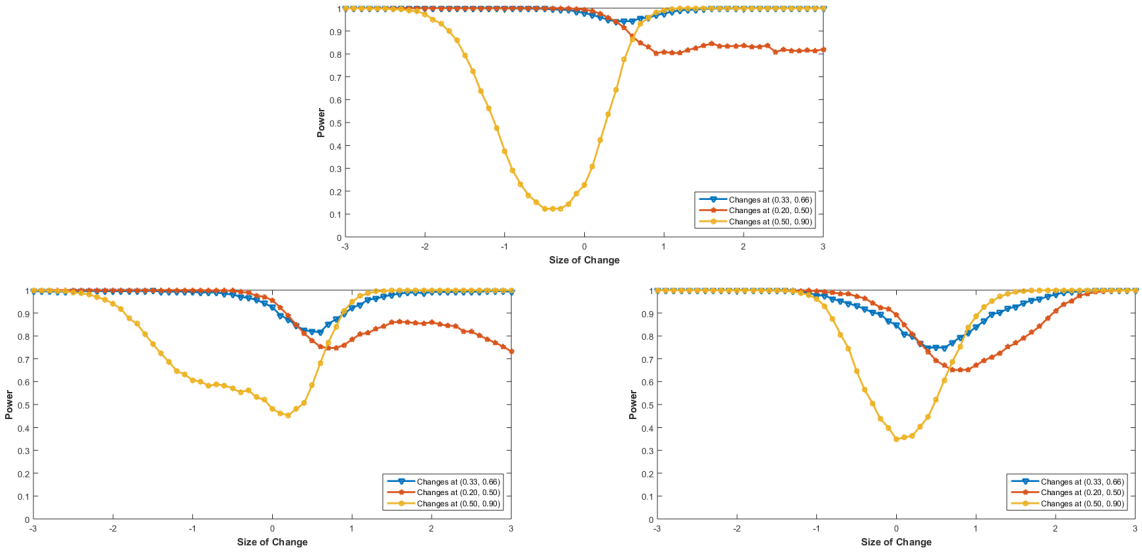
The capital asset pricing model (CAPM) of Sharpe (1964), Lintner (1965) and Merton (1973) and its extensions and modifications have been in the focus of research in applied as well as theoretical finance. In this application we follow Barras et al. (2010) and Fama and French (2010), and use the unconditional four factor CAPM of Carhart (1997) defined as

$$R_t - R_t^{\mathrm{f}} = \alpha_t + (R_t^{\mathrm{M}} - R_t^{\mathrm{f}})\beta_t^{\mathrm{M}} + R_t^{\mathrm{HML}}\beta_t^{\mathrm{HML}} + R_t^{\mathrm{SMB}}\beta_t^{\mathrm{SMB}} + R_t^{\mathrm{MOM}}\beta_t^{\mathrm{MOM}} + e_t, \tag{4.1}$$

$1 \le t \le T$, where $R_t - R_t^{\mathrm{f}}$ denotes the excess return on the mutual fund, $R_t^{\mathrm{M}} - R_t^{\mathrm{f}}$ is the access return on the market portfolio, R_t^{HML} refers to the average return on three small portfolios minus the average return on three big portfolios. The value of R_t^{SMB} is constructed as the average return on two value portfolios minus the average return on the two growth portfolios and R_t^{MOM} gives the returns on a portfolio consisting of stocks with high returns. The monthly return history of US mutual funds is available for the period January 1986 to November 2014 at the web site `http://finance.yahoo.com` and the factors are available at `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`. In the notation of model (1.1) we have that

$$\mathbf{x}_t = (1, R_t^{\mathrm{M}} - R_t^{\mathrm{f}}, R_t^{\mathrm{HML}}, R_t^{\mathrm{SMB}}, R_t^{\mathrm{MOM}})^\top \tag{4.2}$$

FIGURE 3.9. The power functions of the tests based on $V_{400,3}^{(1)}$ (upper panel), $V_{400,3}^{(2)}$ (lower left panel) and $V_{100,3}^{(3)}$ (lower right panel) in Model II under the conditions of Example 3.3



and

$$\boldsymbol{\beta}_t = (\alpha_t, \beta_t^{\mathrm{M}}, \beta_t^{\mathrm{HML}}, \beta_t^{\mathrm{SMB}}, \beta_t^{\mathrm{MOM}})^\top. \tag{4.3}$$

Barras et al. (2010) as well as Fama and French (2010) use (4.1) to evaluate the performance of the managers of 2,076 actively managed US mutual funds that existed between 1975 and 2006 assuming that $\boldsymbol{\beta}_t$ of (4.3) is not time dependent. Fama and French (2010) collect data from January 1984 to December 2006. Baras et al. (2010) provide data from 1975 to 2006. They call a fund manager skilled if the non time dependent α is positive and unskilled otherwise.

The original CAPMs assume that $\boldsymbol{\beta}_t$ is constant (not time dependent) which has been criticized by Harvey (1989) and Ferson and Harvey (1993). Jagannathan and Wang (1996), Lettau and Ludvigson (2001) and Beach (2011) advocate time–dependent betas and provide examples where time–varying beta outperform the unconditional CAPM with constant coefficients. On the other hand, Ghysels (1998) argues that a structural break model might be more suitable in applications. Caporale (2012) provides an example of structural breaks in the CAPM betas in the banking sector, finding only three breaks during the period February 1941 and January 2008. For sequential testing of the stability of high–frequency portfolio betas, we refer to Aue et al. (2012a).

If the null hypothesis is rejected, under the assumption that we have two changes, we estimate the times of change by

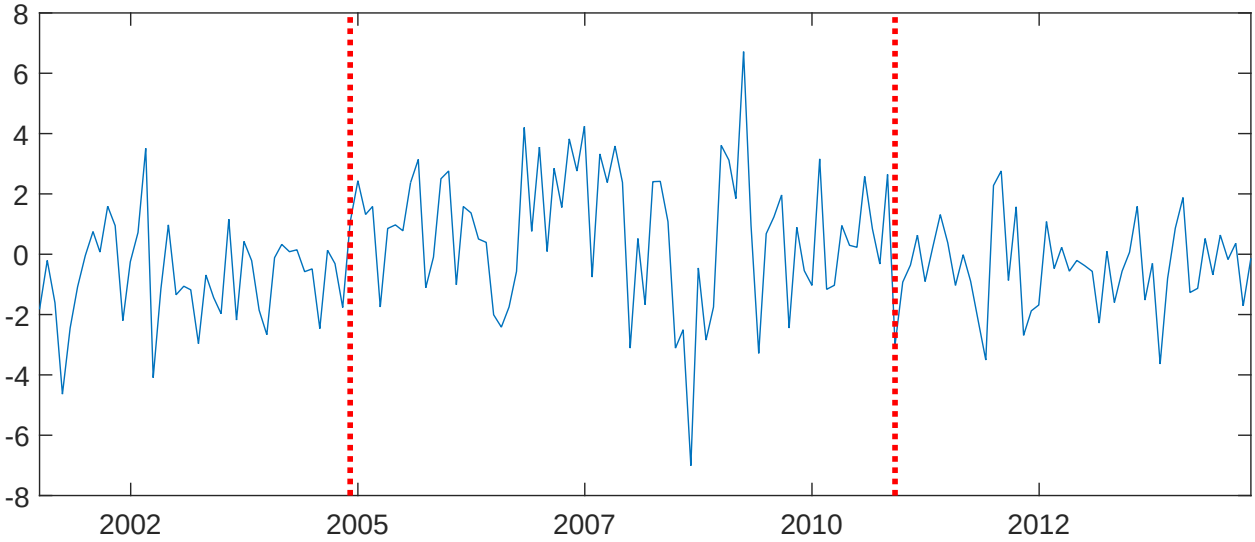
$$(\hat{k}_{1,T}, \hat{k}_{2,T}) = \operatorname{argmax}\{M(k_1, k_2), 1 \le k_1 < k_2 \le T\}.$$

If (2.16) holds and $T^{1/2}\bar{\delta}^{(1)} \to \infty$ and $T^{1/2}\bar{\delta}^{(2)} \to \infty$, then

$$\frac{1}{T}\hat{k}_{1,T} \xrightarrow{\mathbb{P}} \theta_1 \quad \text{and} \quad \frac{1}{T}\hat{k}_{2,T} \xrightarrow{\mathbb{P}} \theta_2.$$

Then the standardized CUSUM procedure is calculated in sub-samples to confirm there are only two changes. For our study we selected two mutual funds, *American Century Heritage C* (code AHGCX) and *Voya Growth and Income Port I* (code IIVGX). AHGCX seeks long-term capital growth. The fund normally invests in stocks of medium-sized and smaller companies that the adviser believes will increase in value over time. Our procedure found two changes, and the estimated time of changes are December 2004 and December 2010. The standardized CUSUM statistics for five sub-samples are reported in Table C.1 and confirms only two changes. The model estimation result of the segmentation is in Table C.2. In the terminology of Barras et al. (2010), the manager of this fund is "unskilled" in two periods, July 2001 to November 2004 and December 2010 to November 2014, since the portfolio α are negative for the two periods.

FIGURE 4.1. Plot of the residuals of the AHGCX time series

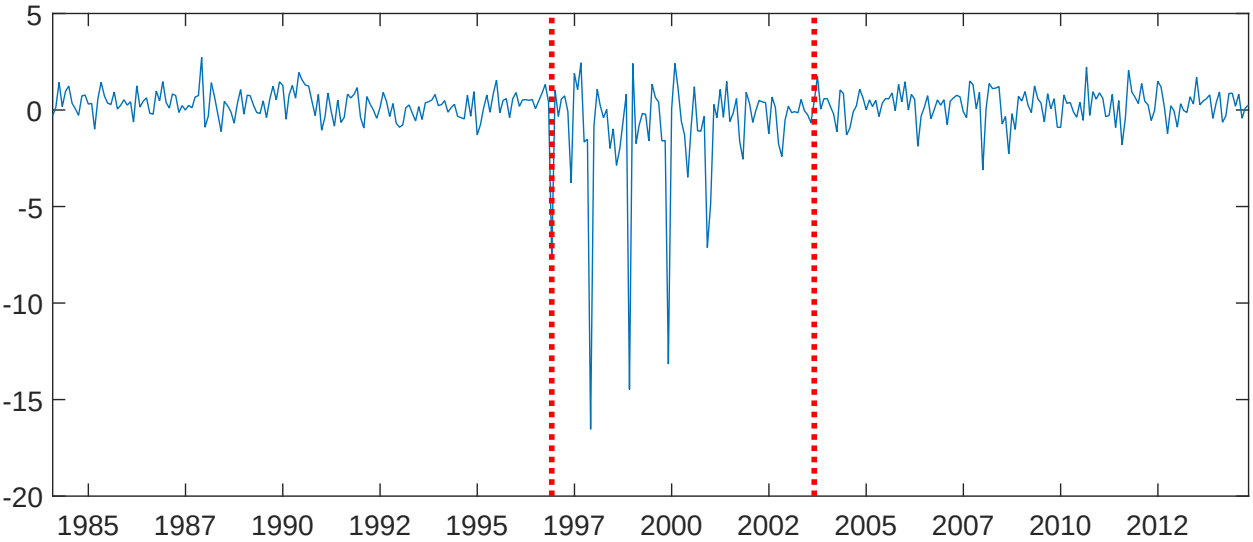


[Tables 4.1 -4.2 Go After Plot AHGCX]

IIVGX's investment aim is to maximize total return through investments in a diversified portfolio of common stock and securities convertible into common stocks. We plot the residuals for the IIVGX data in Figure 4.2. The estimated times for the changes are December 1996 and September 2003. Table C.3 presents the standardized CUSUM statistics for five sub-samples and confirms that there are only two changes. Table C.4 summarizes the outcome of the segmentation procedure. The portfolio α is very low in the period between December 1996 to August 2003.

[Tables 4.2- 4.3 Go After Plot of IIVGX]

FIGURE 4.2. Plot of the residuals of the IIVGX time series



5. SUMMARY

A new procedure has been developed that can test for an arbitrary but fixed number of changes in parameters of time–dependent regression models. This is achieved by modifying the CUSUM statistic so that it can test for at–most-$\mathfrak{m}$ changes in this model. The asymptotic properties of our modified statistic are explored under the null hypothesis of no change as well as under the alternative hypothesis. It is documented there that it converges in distribution to a sum of two independent double exponential random variables and that out test is asymptotically consistent under the alternative. Simulations show that our test statistic can detect one change when there is only one change in the parameters and when there are two changes. These simulations also allow comparison of our statistic with the standard CUSUM statistic. Our statistic is further illustrated through application to detecting time–varying risk factors in the capital asset pricing model.

A. APPENDIX: PROOFS OF THEOREMS 2.1 AND 2.2

Lemma A.1. *If Assumption 2.1 holds, then we have that*

$$\frac{1}{T}\boldsymbol{X}_T\boldsymbol{X}_T^\top \to \boldsymbol{A} \quad a.s.$$

Proof. According to the Cauchy–Schwarz inequality and (2.1), $\mathbb{E}|x_{0,i}x_{0,j}| \le (\mathbb{E}x_{0,i}^2x_{0,j}^2)^{1/2} < \infty$, $1 \le i, j \le d$. Bernoulli shifts are stationary and ergodic sequences (cf. Stout (1974)) and therefore the ergodic theorem (cf. Breiman (1968)) yields

$$\frac{1}{T}\sum_{t=1}^{T} x_{t,i}x_{t,j} \longrightarrow \mathbb{E}x_{0,i}x_{0,j}. \tag{A.1}$$

□

Under the null hypothesis (1.1) reduces to

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

where $\boldsymbol{\beta}$ denotes the common regressor.

Lemma A.2. *If H_0, Assumptions 2.1 and 2.2 hold, then we have that*

$$\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta} = O_{\mathbb{P}}(T^{-1/2})$$

Proof. Since under H_0

$$\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta} = (\boldsymbol{X}_T^\top \boldsymbol{X}_T)^{-1} \boldsymbol{X}_T \boldsymbol{E}_T,$$

where $\mathbf{E}_T = [e_1, e_2, \ldots, e_T]^\top$. Thus we get via Lemma A.1 that

$$\|\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}\| = O_{\mathbb{P}}(1/T)\|\mathbf{X}_T \mathbf{E}_T\|.$$

Since $\mathbb{E}x_{t,i}e_t = 0$ is assumed in (2.1), it is enough to show that

$$E\|\mathbf{X}_T \mathbf{E}_T\|^2 = O(T). \tag{A.2}$$

It follows from stationarity that

$$\begin{aligned}
\mathbb{E}\left(\sum_{t=1}^{T} x_{t,i}e_t\right)^2 &\le \sum_{t=1}^{T}\sum_{s=1}^{T} |\mathbb{E}x_{t,i}e_t x_{s,i}e_s| \\
&= T\mathbb{E}(x_{t,0}e_0)^2 + 2\sum_{1\le t<s\le T} |\mathbb{E}x_{t,i}e_t x_{s,i}e_s| \\
&= T\mathbb{E}(x_{t,0}e_0)^2 + 2\sum_{u=1}^{T-1}(T-u)|\mathbb{E}x_{0,i}e_0 x_{u,i}e_u| \\
&\le T\sum_{u=0}^{\infty} |\mathbb{E}x_{0,i}e_0 x_{u,i}e_u|.
\end{aligned}$$

Using the notation $\mathbf{x}_{t,m} = [x_{t,m,1}, x_{t,m,2}, \ldots, x_{t,m,d}]$, we write

$$\mathbb{E}x_{0,i}e_0 x_{u,i}e_u = \mathbb{E}x_{0,i}e_0(x_{u,i}e_u - x_{u,u,i}e_{u,u})$$

since by independence $\mathbb{E}x_{0,i}e_0x_{u,u,i}e_{u,u} = 0$. Hence by the Cauchy–Schwarz inequality we conclude

$$\begin{aligned}
\sum_{u=1}^{\infty} |\mathbb{E}x_{0,i}e_0x_{u,i}e_u| &= \sum_{u=1}^{\infty} |\mathbb{E}x_{0,i}e_0(x_{u,i}e_u - x_{u,u,i}e_{u,u})| \\
&\le (\mathbb{E}x_{0,i}^4)^{1/4}(\mathbb{E}e_0^4)^{1/4} \sum_{u=1}^{\infty} (\mathbb{E}(x_{u,i}e_u - x_{u,u,i}e_{u,u})^2)^{1/2} \\
&= (\mathbb{E}x_{0,i}^4)^{1/4}(\mathbb{E}e_0^4)^{1/4} \sum_{u=1}^{\infty} (\mathbb{E}(x_{0,i}e_0 - x_{0,u,i}e_{0,u})^2)^{1/2} \\
&\le 2(\mathbb{E}x_{0,i}^4)^{1/4}(\mathbb{E}e_i^4)^{1/4} \sum_{u=1}^{\infty} \left(\mathbb{E}(x_{0,i} - x_{0,u,i})e_0)^2 + \mathbb{E}(x_{0,u,i}(e_0 - e_{0,u})^2\right)^{1/2} \\
&\le 2(\mathbb{E}x_{0,i}^4)^{1/4}(\mathbb{E}e_0^4)^{1/2} \sum_{u=1}^{\infty} (\mathbb{E}(x_{0,i} - x_{0,u,i})^4)^{1/4} \\
&\qquad + 2(\mathbb{E}x_{0,i}^4)^{1/2}(\mathbb{E}e_0^4)^{1/4} \sum_{u=1}^{\infty} (\mathbb{E}(e_0 - e_{0,u})^4)^{1/4},
\end{aligned}$$

and therefore (A.2) follows from (2.2). □

Since under H_0 we have

$$\hat{e}_t = e_t - \boldsymbol{x}_t^\top \left(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}\right), \tag{A.3}$$

we can decompose $M_1(k)$ as

$$M_1(k) = \frac{1}{\sqrt{k}} \left(\sum_{t=1}^{k} e_t + R_k\right), \quad 1 \le k \le T, \tag{A.4}$$

where

$$R_k = R_{k,1} + R_{k,2} + R_{k,3}, \quad 1 \le k \le T,$$

with

$$R_{k,1} = -\left(\sum_{t=1}^{k} \boldsymbol{x}_t\right)^\top \left(\hat{\boldsymbol{\beta}}_{\boldsymbol{T}} - \boldsymbol{\beta}\right), \quad R_{k,2} = -\frac{k}{T}\sum_{t=1}^{T} e_t \text{ and } R_{k,3} = \frac{k}{T}\left(\sum_{t=1}^{T} \boldsymbol{x}_t\right)^\top \left(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}\right).$$

Lemma A.3. *If H_0, Assumptions 2.1 and 2.2 hold, then we have that*

$$\max_{1\le k\le T} k^{-1/2}|R_k| = O_{\mathbb{P}}(1).$$

Proof. By Assumption 2.1, $\boldsymbol{x}_t$ is stationary and ergodic, and therefore by the ergodic theorem (cf. Breiman (1968)) we have

$$\max_{1\le k\le T} \frac{1}{k}\left|\sum_{t=1}^{k} \boldsymbol{x}_t\right| = O_{\mathbb{P}}(1).$$

Hence Lemma A.2 yields

$$\max_{1\le k\le T} k^{-1/2}|R_{k,1}| = O_{\mathbb{P}}(1).$$

Following the arguments in the proof of Lemma A.2 we get that

$$\mathbb{E}\left(\sum_{t=1}^{T} e_t\right) \leq 2T \sum_{s=0}^{T} |\mathbb{E}e_0 e_s| < \infty$$

and since $\mathbb{E}e_t = 0$ we conclude

$$\left|\sum_{t=1}^{T} e_t\right| = O_{\mathbb{P}}(T^{1/2}). \tag{A.5}$$

Hence

$$\max_{1\leq k\leq T} k^{-1/2}|R_{k,2}| = O_{\mathbb{P}}(1).$$

By the ergodic theorem we have that

$$\frac{1}{T}\sum_{t=1}^{T} |x_{t,i}| = O_{\mathbb{P}}(1). \tag{A.6}$$

Applying Lemma A.2 we get that

$$\max_{1\leq k\leq T} k^{-1/2}|R_{k,3}| = \max_{1\leq k\leq T}\left|\frac{\sqrt{k}}{T}\left(\sum_{t=1}^{T}\boldsymbol{x}_t\right)^{\top}\left(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}\right)\right| = O_{\mathbb{P}}(1),$$

completing the proof of lemma A.3. □

Lemma A.4. *If Assumptions 2.1 and 2.2 hold, then we have that*

$$(2\log\log T)^{-1/2} \max_{1\leq k\leq T} \frac{1}{\sqrt{k}}\left|\sum_{t=1}^{k} e_t\right| \xrightarrow{\mathbb{P}} \sigma,$$

where σ is defined in (2.3).

Proof. It follows from Lemma 5.4 of Aue et al. (2014), that we can define Wiener processes W_T such that

$$\max_{1\leq k\leq T} k^{-1/2+\delta}\left|\sum_{t=1}^{k} e_t - \sigma W_T(k)\right| = O_{\mathbb{P}}(1) \text{ with some } \delta > 0. \tag{A.7}$$

Hence for any $1 \leq c_1 = c_1(T) \leq c_2 = c_2(T) \leq T$ we have

$$\max_{c_1\leq k\leq c_2} \frac{1}{\sqrt{k}}\left|\sum_{t=1}^{k} e_t\right| - \max_{c_1\leq k\leq c_2} \frac{\sigma}{\sqrt{k}}\left|W_T(k)\right| = O_{\mathbb{P}}(c_1^{-\delta}). \tag{A.8}$$

Since the distribution of $W_T(\cdot)$ does not depend on T, by the law of the iterated logarithm for Wiener processes we get

$$(2\log\log T)^{-1/2} \max_{1\leq k\leq T} \frac{1}{\sqrt{k}}\left|W_T(k)\right| \xrightarrow{\mathbb{P}} 1,$$

which completes the proof. □

Lemma A.5. *If H_0, Assumptions 2.1 and 2.2 hold, then we have that*

$$\mathbb{P}\left(\max_{1\leq k\leq T}|M_1(k)| = \max_{n(T)\leq k\leq m(T)}|M_1(k)|\right) \xrightarrow{\mathbb{P}} 1,$$

where $n(T) = (\log T)^{\kappa}$, and $m(T) = T/(\log T)^{\kappa}$ with any $\kappa > 0$.

Proof. It follows from Lemmas A.3 and A.4 that

$$(2\log\log T)^{-1/2}\max_{1\le k\le T}|M_1(k)|\xrightarrow{\mathbb{P}}\sigma, \tag{A.9}$$

and therefore we need to show only that

$$\max_{1\le k\le n(T)}|M_1(k)|=o_{\mathbb{P}}((\log\log T)^{1/2})$$

and

$$\max_{m(T)\le k\le T}|M_1(k)|=o_{\mathbb{P}}((\log\log T)^{1/2}).$$

On account of Lemma A.3 we need to prove only that

$$\max_{1\le k\le n(T)}\frac{1}{\sqrt{k}}\left|\sum_{t=1}^{k}e_t\right|=o_{\mathbb{P}}((\log\log T)^{1/2}) \tag{A.10}$$

and

$$\max_{m(T)\le k\le T}\frac{1}{\sqrt{k}}\left|\sum_{t=1}^{k}e_t\right|=o_{\mathbb{P}}((\log\log T)^{1/2}). \tag{A.11}$$

For any Wiener process $W(\cdot)$ we have that

$$\max_{1\le k\le n(T)}\frac{1}{\sqrt{k}}|W(k)|=O_{\mathbb{P}}((\log\log\log T)^{1/2}) \tag{A.12}$$

and

$$\max_{m(T)\le k\le T}\frac{1}{\sqrt{k}}|W(k)|=O_{\mathbb{P}}((\log\log\log T)^{1/2}) \tag{A.13}$$

(cf. Csörgő and Horváth (1997)). The claims in (A.10) and (A.11) follow immediately from (A.8) and (A.12), (A.13). □

According to Lemma A.5, $|M_1(k)|$ reaches it largest value on the interval $n(T), m(T)$ with probability converging to 1.

Lemma A.6. *If H_0, Assumptions 2.1 and 2.2 hold, then we have that*

$$\max_{n(T)\le k\le m(T)}|M_1(k)|-\max_{n(T)\le k\le m(T)}\frac{1}{\sqrt{k}}\left|\sum_{t=1}^{k}e_t\right|=o_{\mathbb{P}}((\log\log T)^{-1/2}),$$

where $n(T)$ and $m(T)$ are defined in Lemma A.5.

Proof. We use again (A.4). Combining Lemma A.2 and (A.1) we get that

$$\max_{n(T)\le k\le m(T)}k^{-1/2}|R_{k,1}|=O_{\mathbb{P}}(1)\max_{n(T)\le k\le m(T)}k^{1/2}T^{-1/2}=O_{\mathbb{P}}(1)\,(\log T)^{-\kappa/2}$$

and similarly

$$\max_{n(T)\le k\le m(T)}k^{-1/2}|R_{k,3}|=O_{\mathbb{P}}(1)\,(\log T)^{-\kappa/2}$$

It follows from (A.5) that

$$\max_{n(T)\le k\le m(T)}k^{-1/2}|R_{k,2}|=O_{\mathbb{P}}(1)T^{-1/2}m^{1/2}(T)=O_{\mathbb{P}}(1)\,(\log T)^{-\kappa/2},$$

which concludes the proof. □

Next we consider $M_{\mathfrak{m}+1}(\ell)$, since its definition is similar to that of $M_1(k)$. However, due to time reversal, i.e. the CUSUM starts with residual $\hat{e}_T$, we need to modify Lemmas A.3–A.6. As in the decomposition of $M_1(k)$, we have

$$M_{\mathfrak{m}+1}(\ell) = \frac{1}{\sqrt{T-\ell}}\left(\sum_{t=\ell}^{T} e_t + R_\ell^*\right), \quad 1\ell < T,$$

where

$$R_\ell^* = R_{\ell,1}^* + R_{\ell,2}^* + R_{\ell,3}^*, \quad 1\ell < T,$$

with

$$R_{\ell,1}^* = -\left(\sum_{t=\ell}^{T} \mathbf{x}_t\right)^\top \left(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}\right), \;\; R_{\ell,2}^* = -\frac{T-\ell}{T}\sum_{t=1}^{T} e_t$$

and

$$R_{\ell,3}^* = \frac{T-\ell}{T}\left(\sum_{t=1}^{T} \mathbf{x}_t\right)^\top \left(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}\right).$$

Lemma A.7. *If H_0, Assumptions 2.1 and 2.2 hold, then we have that*

$$\max_{1\le\ell<T} \frac{|R_\ell^*|}{\sqrt{T-\ell}} = O_{\mathbb{P}}(1).$$

Proof. First we show that

$$\max_{1\le\ell<T} \frac{1}{T-\ell}\left|\sum_{t=\ell}^{T} x_{t,i}\right| = O_{\mathbb{P}}(1). \tag{A.14}$$

Using Aue et al. (2014), we can find Wiener processes $W_{T,i}(\cdot)$ such that

$$\max_{1\le\ell<T} \frac{1}{\sqrt{T-\ell}}\left|\sum_{t=\ell}^{T}(x_{t,i} - \mathbb{E}x_{t,i})\right| - \max_{1\le\ell<T}\left(\frac{\mathrm{var}(x_{0,i})}{T-\ell}\right)^{1/2}\left|W_{T,i}(T-\ell)\right| = O_{\mathbb{P}}(1),$$

and therefore (A.14) follows from the law of the iterated logarithm for Wiener processes. Hence Lemma A.2 yields

$$\max_{1\le\ell<T}(T-\ell)^{-1/2}|R_{\ell,1}^*| = O_{\mathbb{P}}(1) \quad \text{and} \quad \max_{1\le\ell<T}(T-\ell)^{-1/2}|R_{\ell,3}^*| = O_{\mathbb{P}}(1).$$

Using (A.5) we get immediately that

$$\max_{1\le\ell<T}(T-\ell)^{-1/2}|R_{\ell,2}^*| = O_{\mathbb{P}}(1).$$

□

We continue with the analogue of Lemma A.4.

Lemma A.8. *If Assumptions 2.1 and 2.2 hold, then we have that*

$$(2\log\log T)^{-1/2}\max_{1\le\ell<T}\frac{1}{\sqrt{T-\ell}}\left|\sum_{t=\ell}^{T} e_t\right| \xrightarrow{\mathbb{P}} \sigma,$$

where σ is defined in (2.3).

Proof. Using again Lemma 5.4 of Aue et al. (2014), we can define Wiener processes W_T^* such that

$$\max_{1\le \ell<T}(T-\ell)^{-1/2+\delta}\left|\sum_{t=\ell}^{T} e_t-\sigma W_T^*(T-\ell)\right| = O_{\mathbb{P}}(1) \text{ with some } \delta>0. \tag{A.15}$$

Change of variable and the Darling–Erdős law (cf. Csörgő and Horváth (1997)) gives

$$(2\log\log T)^{-1/2}\max_{1\le \ell<T}\frac{1}{\sqrt{T-\ell}}\left|W_T^*(T-\ell)\right| \xrightarrow{\mathbb{P}} 1,$$

and therefore Lemma A.8 follows from (A.15) along the lines of (A.7) and (A.8). □

Lemma A.9. *If H_0, Assumptions 2.1 and 2.2 hold, then we have that*

$$\mathbb{P}\left(\max_{1\le \ell<T}|M_{\mathfrak{m}+1}(\ell)| = \max_{T-m(T)\le k\le T-n(T)}|M_{\mathfrak{m}+1}(\ell)|\right) \longrightarrow 1,$$

where $n(T)$ and $m(T)$ are defined in Lemma A.5.

Proof. It follows from Lemmas A.7 and A.8 that

$$(2\log\log T)^{-1/2}\max_{1\le \ell<T}(T-\ell)^{-1/2}|M_{\mathfrak{m}+1}(\ell)| \xrightarrow{\mathbb{P}} \sigma.$$

Thus, in light of Lemmas A.7 and A.8, it is enough to establish that

$$\max_{1\le \ell\le T-m(T)}(T-\ell)^{-1/2}\left|\sum_{t=\ell}^{T} e_t\right| = o_{\mathbb{P}}((\log\log T)^{1//2}) \tag{A.16}$$

and

$$\max_{T-n(T)\le \ell<T}(T-\ell)^{-1/2}\left|\sum_{t=\ell}^{T} e_t\right| = o_{\mathbb{P}}((\log\log T)^{1//2}). \tag{A.17}$$

Using (A.15) we get for any $1\le c_1=c_1(T)\le c_2=c_2(T)<T$

$$\max_{T-c_2\le \ell\le T-c_1}\frac{1}{\sqrt{T-\ell}}\left|\sum_{t=\ell}^{T} e_t\right| - \max_{T-c_2\le \ell\le T-c_1}\frac{\sigma}{\sqrt{T-\ell}}\left|W_T^*(T-\ell)\right| = O_{\mathbb{P}}(c_1^{-\delta}),$$

and therefore (A.12) and (A.13) imply (A.16) and (A.17). □

Lemma A.10. *If H_0, Assumptions 2.1 and 2.2 hold, then we have that*

$$\max_{T-m(T)\le \ell\le T-n(T)}|M_{\mathfrak{m}+1}(\ell)| - \max_{T-m(T)\le \ell\le T-n(T)}\frac{1}{\sqrt{T-\ell}}\left|\sum_{t=\ell}^{T} e_t\right| = o_{\mathbb{P}}((\log\log T)^{-1/2}),$$

where $n(T)$ and $m(T)$ are defined in Lemma A.5.

Proof. Putting together Lemma A.2 and (A.14) we conclude

$$\max_{T-m(T)\le \ell\le T-n(T)}(T-\ell)^{-1/2}|R_{\ell,1}^*| = O_{\mathbb{P}}(1)(\log T)^{-\kappa/2}.$$

Similar arguments give

$$\max_{T-m(T)\le \ell\le T-n(T)}(T-\ell)^{-1/2}|R_{\ell,2}^*| = O_{\mathbb{P}}(1)(\log T)^{-\kappa/2}$$

and

$$\max_{T-m(T)\le \ell\le T-n(T)}(T-\ell)^{-1/2}|R_{\ell,3}^*| = O_{\mathbb{P}}(1)(\log T)^{-\kappa/2}.$$

□

Lemma A.11. *If H_0, Assumptions 2.1 and 2.2 hold, then we have that*

$$\max_{1\le k\le \ell\le T} |M_j(k,\ell)| = O_{\mathbb{P}}(1), \quad 2\le j\le \mathfrak{m}.$$

Proof. Using (A.3) we write

$$M_j(k,\ell) = R_{k,\ell,1} + \ldots + R_{k,\ell,4},$$

where

$$R_{k,\ell,1} = T^{-1/2}\sum_{t=k}^{\ell} e_t, \quad R_{k,\ell,2} = -T^{-1/2}\left(\sum_{t=k}^{\ell}\mathbf{x}_t\right)^{\top}\left(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}\right), \quad R_{k,\ell,3} = -\frac{\ell-k}{T^{3/2}}\sum_{t=1}^{T} e_t$$

and

$$R_{k,\ell,4} = \frac{\ell-k}{T^{3/2}}\left(\sum_{t=1}^{T}\mathbf{x}_t\right)^{\top}\left(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}\right).$$

It follows from Lemma 5.4 of Aue et al. (2014) that

$$T^{-1/2}\sum_{t=1}^{Tu} e_t \xrightarrow{\mathcal{D}[0,1]} \sigma W(t),$$

where $W(\cdot)$ is a Wiener process. Hence

$$\sup_{0\le u,v\le 1} T^{-1/2}\left|\sum_{t=Tu}^{Tv} e_t\right| \xrightarrow{\mathcal{D}} \sigma \sup_{0\le u,v\le 1}|W(u)-W(v)|,$$

which implies

$$\max_{1\le k\le \ell\le T}|R_{k,\ell,1}| = O_{\mathbb{P}}(1).$$

Applying Lemma A.2 and the ergodic theorem we conclude

$$\max_{1\le k\le \ell\le T}|R_{k,\ell,2}| = O_{\mathbb{P}}(1)\frac{1}{T}\max_{1\le k\le \ell\le T}\left\|\sum_{t=k}^{\ell}\mathbf{x}_t\right\| = O_{\mathbb{P}}(1)\frac{1}{T}\max_{1\le k\le \ell\le T}\sum_{t=k}^{\ell}\|\mathbf{x}_t\| = O_{\mathbb{P}}(1).$$

Similar arguments yield

$$\max_{1\le k\le \ell\le T}|R_{k,\ell,3}| = O_{\mathbb{P}}(1) \quad \text{and} \quad \max_{1\le k\le \ell\le T}|R_{k,\ell,4}| = O_{\mathbb{P}}(1).$$

□

Let us define $A_T = [n(T), m(T)] \times [T-m(T), T-n(T)]$, where $n(T)$ and $m(T)$ are defined ine Lemma A.5.

Lemma A.12. *If H_0, Assumptions 2.1 and 2.2 hold, then we have that*

$$\max_{(k,\ell)\in A_T}|M_j(k,\ell)| = o_{\mathbb{P}}\left((\log\log T)^{-1/2}\right), \quad 2\le j\le \mathfrak{m}.$$

Proof. On account of (A.3) we can decompose M_j as

$$M_j(k,\ell) = M_{2,1}(k) + M_{2,2}(k) + M_{2,3}(\ell) + M_{2,4}(\ell),$$

where

$$M_{2,1}(k) = -T^{-1/2}\sum_{t=1}^{k-1}\hat{e}_t, \quad M_{2,2}(k) = \frac{k}{T^{3/2}}\sum_{t=1}^{T}\hat{e}_t, \quad M_{2,3}(\ell) = -T^{-1/2}\sum_{t=\ell+1}^{T}\hat{e}_t,$$

and

$$M_{2,4}(\ell) = \frac{T-\ell}{T^{3/2}}\sum_{t=1}^{T}\hat{e}_t.$$

We note that by (A.3) that

$$\sum_{t=u}^{v}\hat{e}_t = \sum_{t=u}^{v} e_t - \left(\sum_{t=u}^{v}\mathbf{x}_t\right)^{\top}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) = O_{\mathbb{P}}(T^{1/2}).$$

Thus we obtain that from Lemma A.2, (A.5) and (A.6) that

$$\max_{n(T)\le k\le m(T)} |M_{2,2}(k)| = O_{\mathbb{P}}((\log T)^{-\kappa/2}) \quad \text{and} \quad \max_{T-m(T)\le \ell\le T-n(T)} |M_{2,4}(\ell)| = O_{\mathbb{P}}((\log T)^{-\kappa/2}).$$

Applying (A.7) we get

$$\begin{aligned} T^{-1/2}&\max_{n(T)\le k\le m(T)}\left|\sum_{t=1}^{k} e_t\right| \qquad (A.18)\\ &\le T^{-1/2}\max_{n(T)\le k\le m(T)}\sigma\,|W_T(k)| + T^{-1/2}\max_{n(T)\le k\le m(T)}\left|\sum_{t=1}^{k} e_t - \sigma W_T(k)\right|\\ &= O_{\mathbb{P}}(1)T^{-1/2}\max_{n(T)\le k\le m(T)} k^{1/2-\delta} + T^{-1/2}\max_{n(T)\le k\le m(T)}\sigma\,|W_T(k)|\\ &= O_{\mathbb{P}}(1)\left(T^{-1/2}m^{1/2-\delta}(T) + (\log T)^{-\kappa/2}\right)\\ &= o_{\mathbb{P}}\left((\log\log T)^{-1/2}\right), \end{aligned}$$

since by the scale transformation of the Wiener process $W(\cdot)$ we have

$$\sup_{0\le u\le m(T)}|W(t)| \overset{\mathcal{D}}{=} m^{1/2}(T)\sup_{0\le v\le 1}|W(v)|.$$

By the ergodic theorem and Lemma A.2 we obtain that

$$T^{-1/2}\max_{n(T)\le k\le m(T)}\left|\left(\sum_{t=1}^{k-1}\mathbf{x}_t\right)^{\top}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})\right| = O_{\mathbb{P}}(1),$$

and therefore (A.18) implies

$$\max_{n(T)\le k\le m(T)}|M_{2,1}(k)| = o_{\mathbb{P}}((\log\log T)^{-1/2}).$$

Next we write

$$|M_{2,3}(\ell)| \le \left|\sum_{t=\ell+1}^{T} e_t\right| + \left|\left(\sum_{t=\ell+1}^{T}\mathbf{x}_t\right)^{\top}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})\right|.$$

Replacing (A.7) with (A.15), one can establish along the lines of the proof of (A.18) that

$$T^{-1/2} \max_{T-m(T)\le k\le T-n(T)} \left| \sum_{t=\ell+1}^{T} e_t \right| = o_{\mathbb{P}}((\log\log T)^{-1/2}).$$

Lemma A.2 and (A.14) give

$$\begin{aligned} T^{-1/2} \max_{T-m(T)\le \ell\le T-n(T)} \left| \left(\sum_{t=\ell+1}^{T} \mathbf{x}_t \right)^{\top} (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) \right| &= O_{\mathbb{P}}(1/T) \max_{T-m(T)\le \ell\le T-n(T)} (T-\ell) \\ &= O_{\mathbb{P}}(m(T)/T), \end{aligned}$$

which completes the proof of

$$\max_{T-m(T)\le \ell\le T-n(T)} M_{2,3}(\ell) = o_{\mathbb{P}}((\log\log T)^{-1/2}).$$

□

Proof of Theorem 2.1. It follows from Lemmas A.5–A.12 that

$$H_T = \max\left(\max_{n(T)\le k\le m(T)} \frac{1}{\sqrt{k}} \left| \sum_{t=1}^{k} e_t \right|, \max_{T-m(T)\le \ell\le T-n(T)} \frac{1}{\sqrt{T-\ell}} \left| \sum_{t=\ell}^{T} e_t \right| \right) + o_{\mathbb{P}}((\log\log T)^{-1/2})$$

and

$$M_T = \max_{n(T)\le k\le m(T)} \frac{1}{\sqrt{k}} \left| \sum_{t=1}^{k} e_t \right| + \max_{T-m(T)\le \ell\le T-n(T)} \frac{1}{\sqrt{T-\ell}} \left| \sum_{t=\ell}^{T} e_t \right| + o_{\mathbb{P}}((\log\log T)^{-1/2}).$$

We get from (A.3) that

$$D_T = T^{-1/2} \max_{1\le \ell\le T} \left| \sum_{t=1}^{\ell} e_t - \frac{\ell}{T} \sum_{t=1}^{T} e_t - \left(\sum_{t=1}^{\ell} \mathbf{x}_t - \frac{\ell}{T} \sum_{t=1}^{T} \mathbf{x}_t \right)^{\top} (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) \right|.$$

By Assumption 2.1 we can use the approximation in Aue et al. (2014) and get

$$T^{-1/2} \max_{1\le \ell\le T} \left\| \sum_{t=1}^{\ell} \mathbf{x}_t - \frac{\ell}{T} \sum_{t=1}^{T} \mathbf{x}_t \right\| = O_{\mathbb{P}}(1)$$

and therefore by Lemma A.2 we have

$$D_T = T^{-1/2} \max_{1\le \ell\le T} \left| \sum_{t=1}^{\ell} e_t - \frac{\ell}{T} \sum_{t=1}^{T} e_t \right| + o_{\mathbb{P}}(1).$$

Also, by Aue et al. (2014) we also conclude that

$$\begin{aligned} &T^{-1/2} \max_{1\le \ell\le T} \left| \sum_{t=1}^{\ell} e_t - \frac{\ell}{T} \sum_{t=1}^{T} e_t \right| \\ &\qquad = T^{-1/2} \max_{m(T)+1\le \ell\le T-m(T)-1} \left| \sum_{t=m(T)+1}^{\ell} e_t - \frac{\ell}{T} \sum_{t=m(T)+1}^{T-m(T)-1} e_t \right| + o_{\mathbb{P}}(1). \end{aligned} \tag{A.19}$$

Thus we need to show only

$$\left(a_T \max_{n(T)\le k\le m(T)} \frac{1}{\sigma\sqrt{k}}\left|\sum_{t=1}^{k} e_t\right| - b_T,\ a_T \max_{T-m(T)\le \ell\le T-n(T)} \frac{1}{\sigma\sqrt{T-\ell}}\left|\sum_{t=\ell}^{T} e_t\right| - b_T,\right. \tag{A.20}$$

$$\left.\frac{1}{\sigma\sqrt{T}} \max_{m(T)+1\le \ell\le T-m(T)-1}\left|\sum_{t=m(T)+1}^{\ell} e_t - \frac{\ell}{T}\sum_{t=m(T)+1}^{T-m(T)-1} e_t\right|\right)$$

$$\xrightarrow{\mathcal{D}} \left(\xi_1, \xi_2, \sup_{0\le t\le 1}|B(t)|\right),$$

where ξ_1 and ξ_2 are random variables with distribution defined in (2.4) and $\{B(t), 0\le t\le 1\}$ is a Brownian bridge. Also, ξ_1, ξ_2 and $\{B(t), 0\le t\le 1\}$ are independent. In (A.7) and (A.15) we obtained weighted approximations for $\sum_{t=1}^{k} e_t$ and $\sum_{t=\ell}^{T} e_t$ with some Wiener processes. However, the proof of (A.20) requires a joint approximation of these partial sums on $(k,\ell)\in A_T$ and $\sum_{t=u}^{v} e_t, m(T)+1\le u<v\le T-m(T)-1$. The approximation in Aue et al. (2014) uses the blocking technique and therefore we can define three independent Wiener processes $W_{T,1}(\cdot), W_{T,2}(\cdot)$ and $W_{T,3}(\cdot)$ such that

$$\max_{n(T)\le k\le m(T)} k^{1/2-\delta}\left|\sum_{t=1}^{k} e_t - \sigma W_{T,1}(k)\right| = O_{\mathbb{P}}(1), \tag{A.21}$$

$$\max_{T-m(T)\le \ell\le T-n(T)} (T-\ell)^{1/2-\delta}\left|\sum_{t=\ell}^{T} e_t - \sigma W_{T,2}(T-\ell)\right| = O_{\mathbb{P}}(1) \tag{A.22}$$

with some $\delta>0$ and

$$\max_{m(T)+1\le k\le T-m(T)-1}\left|\sum_{t=m(T)+1}^{k} e_t - \sigma W_{T,3}(k-m(T))\right| = o\mathbb{P}(1). \tag{A.23}$$

The approximation in (A.23) yields that

$$\frac{1}{\sqrt{T}} \max_{m(T)+1\le \ell\le T-m(T)-1}\left|\sum_{t=m(T)+1}^{\ell} e_t - \frac{\ell}{T}\sum_{t=m(T)+1}^{T-m(T)-1} e_t\right.$$

$$\left. - \sigma\left(W_{T,3}(\ell-m(T)) - \frac{\ell}{T}W_{T,3}(T-2m(T)-1)\right)\right| = o_{\mathbb{P}}(1)$$

and the modulus of continuity of the Wiener process (cf. Csörgő and Révész (1981)) gives

$$\frac{1}{\sqrt{T}} \max_{m(T)+1\le \ell\le T-m(T)-1}\left|W_{T,3}(\ell-m(T)) - \frac{\ell}{T}W_{T,3}(T-2m(T)-1)\right|$$

$$= \frac{1}{\sqrt{T}} \sup_{0\le x\le T}\left|W_{T,3}(x) - \frac{x}{T}W_{T,3}\right| + o_{\mathbb{P}}(1).$$

Thus we have

$$\frac{1}{\sqrt{T}} \max_{m(T)+1\le \ell\le T-m(T)-1} \left| \sum_{t=m(T)+1}^{\ell} e_t - \frac{\ell}{T} \sum_{t=m(T)+1}^{T-m(T)-1} e_t \right| \tag{A.24}$$
$$= \frac{\sigma}{\sqrt{T}} \sup_{0\le x\le T} \left| W_{T,3}(x) - \frac{x}{T} W_{T,3}(x) \right| + o_{\mathbb{P}}(1).$$

It is easy to see that

$$\frac{1}{\sqrt{T}} \sup_{0\le x\le T} \left| W_{T,3}(x) - \frac{x}{T} W_{T,3}(x) \right| \stackrel{\mathcal{D}}{=} \sup_{0\le t\le 1} |B(t)|, \tag{A.25}$$

where $B(t)$ denotes a Brownian bridge. It follows from (A.21) that

$$\max_{n(T)\le k\le m(T)} \frac{1}{\sqrt{k}} \left| \sum_{t=1}^{t} e_t \right| - \max_{n(T)\le k\le m(T)} \frac{\sigma}{\sqrt{k}} \left| W_{T,1}(k) \right| = O_{\mathbb{P}}((\log T)^{-\kappa\delta}) \tag{A.26}$$

and (A.22) yields

$$\max_{T-m(T)\le k\le T-n(T)} \frac{1}{\sqrt{T-\ell}} \left| \sum_{t=\ell}^{T} e_t \right| - \max_{T-m(T)\le \ell\le T-n(T)} \frac{\sigma}{\sqrt{T-\ell}} \left| W_{T,2}(T-\ell) \right| \tag{A.27}$$
$$= O_{\mathbb{P}}((\log T)^{-\kappa\delta}).$$

The asymptotic independence in (A.20) is an immediate consequence of (A.24), (A.26) and (A.27).

The Darling–Erdős limit result (cf. Appendix A in Csörgő and Horváth (1997)) states that

$$a_{m(T)/n(T)} \max_{n(T)\le k\le m(T)} \frac{1}{\sqrt{k}} |W(k)| - b_{m(T)/n(T)} \stackrel{\mathcal{D}}{\to} \xi, \tag{A.28}$$

for any Wiener process $W(\cdot)$, where ξ is a random variable with distribution function $\exp(-e^{-x})$. Elementary calculations give

$$|a_T - a_{m(T)/n(T)}| = O((\log\log T)^{1/2}/\log T) = o((\log\log T)^{-1/2})$$

and similarly

$$|b_T - b_{m(T)/n(T)}| = o(1).$$

Hence (A.28) can be rewritten as

$$a_T \max_{n(T)\le k\le m(T)} \frac{1}{\sqrt{k}} |W(k)| - b_T \stackrel{\mathcal{D}}{\to} \xi. \tag{A.29}$$

Now (A.20) follows from (A.24)–(A.27) and (A.29). □

Proof of Theorem 2.2. We follow the proof of Theorem 2.1. By Lemmas A.5–A.12 we have that

$$M_T^{(1)} = \max_{n(T)\le k\le m(T)} \frac{1}{\sqrt{k}} \left| \sum_{t=1}^{k} e_t \right| + o_{\mathbb{P}}((\log\log T)^{-1/2}), \tag{A.30}$$

$$M_T^{(2)} = \max_{T-m(T)\le k\le T-n(T)} \frac{1}{\sqrt{T-\ell}} \left| \sum_{t=\ell}^{T} e_t \right| + o_{\mathbb{P}}((\log\log T)^{-1/2}), \tag{A.31}$$

and

$$M_T^* = \max_{\mathfrak{M}_0^*} \sum_{i=2}^{\mathfrak{m}} \frac{1}{\sqrt{T}} \left| \sum_{t=k_{i-1}}^{k_i} e_t - \frac{k_i - k_{i-1}}{T} \sum_{t=m(T)+1}^{T-m(T)-1} e_t \right|, \tag{A.32}$$

where $\mathfrak{M}_0^* = \{n(T) \le k_1 \le k_2 \le \dots k_{\mathfrak{m}} \le T - m(T) - 1\}$ (cf. (A.19)). Repeating the arguments used in the proof of Theorem 2.1, we obtain the asymptotic independence of $a(T)M_T^{(1)}/\sigma - b(T), a(T)M_T^{(2)}/\sigma - b(T)$ amd M_T^* from (A.30)–(A.32). We also conclude from Lemmas A.5–A.12 and (A.30), (A.31) that

$$\left(a_T \frac{M_T^{(1)}}{\sigma} - b_T, a_T \frac{M_T^{(2)}}{\sigma} - b_T \right) \xrightarrow{\mathcal{D}} (\xi_1, \xi_2),$$

where ξ_1 and ξ_2 defined in (2.4). We obtain from (A.24) and the continuity of the Wiener process that

$$\begin{aligned} &\max_{\mathfrak{M}_0^*} \sum_{i=2}^{\mathfrak{m}} \frac{1}{\sqrt{T}} \left| \sum_{t=k_{i-1}}^{k_i} e_t - \frac{k_i - k_{i-1}}{T} \sum_{t=m(T)+1}^{T-m(T)-1} e_t \right| \\ &\qquad \xrightarrow{\mathcal{D}} \sigma \max_{0 \le u_1 \le u_2 \le \dots \le u_{\mathfrak{m}} \le 1} \sum_{i=2}^{\mathfrak{m}} \left| W(u_i) - W(u_{i-1}) - (u_i - u_{i-1}) W(1) \right|, \end{aligned}$$

with $u_0 = 0$, where $W(t), 0 \le t \le 1$ is a Wiener process. Since $B(t) = W(t) - tW(1)$ is a Brownian bridge, the proof of Theorem 2.2 is complete. □

B. Appendix: Proof of Theorem 2.4.

The OLS estimator $\boldsymbol{\beta}_T$ under H_A can be decomposed as

$$\begin{aligned} \hat{\boldsymbol{\beta}}_T &= (\mathbf{X}_T \mathbf{X}_T^\top)^{-1} \mathbf{X}_T \mathbf{Y}_T \\ &= (\mathbf{X}_T \mathbf{X}_T^\top)^{-1} \left(\sum_{i=1}^{\mathfrak{m}+1} \sum_{t=k_{i-1}^*+1}^{k_i^*} \mathbf{x}_t \mathbf{x}_t^\top \boldsymbol{\beta}^{(i)} + \sum_{t=1}^{T} \mathbf{x}_t e_t \right). \end{aligned}$$

We continue with the analogue of Lemma A.2 under H_A.

Lemma B.1. *If H_A, Assumptions 2.1 and 2.2 hold, then we have that*

$$\hat{\boldsymbol{\beta}}_T = \tilde{\boldsymbol{\beta}}_T + O_{\mathbb{P}}(T^{-1/2}). \tag{B.1}$$

Proof. Since $x_{t,i}x_{j,t}, -\infty < t < \infty$ is a Bernoulli shift, repeating the arguments leading to A.2 we obtain that

$$\mathbb{E} \left(\sum_{t=1}^{N} (x_{t,i}x_{j,t} - \mathbb{E}x_{t,i}x_{j,t}) \right)^2 = O(N), \quad \text{as } N \to \infty.$$

and therefore

$$(\mathbf{X}_T \mathbf{X}^\top)^{-1} = \frac{1}{T} \mathbf{A}^{-1} + O_{\mathbb{P}}(T^{-3/2})$$

and

$$\sum_{t=k_{i-1}^*+1}^{k_i^*} \mathbf{x}_t\mathbf{x}_t^\top = (k_i^* - k_{i-1}^*)\mathbf{A} + O_{\mathbb{P}}\left(\sqrt{k_i^* - k_{i-1}^*}\right),$$

and therefore by (A.2) we have (B.1). □

Proof of Theorem 2.4. Similarly to A.2 we have

$$\mathbb{E}\left\|\sum_{t=T_1}^{T_2}(\mathbf{x}_t - \mathbf{c})\right\|^2 = O(T_2 - T_1), \quad \text{as } N \to \infty, \tag{B.2}$$

and therefore

$$\sum_{t=1}^{T} \mathbf{x}_t = T\mathbf{c} + O_{\mathbb{P}}(\sqrt{T}) \text{ and } \sum_{t=k_{i-1}^*+1}^{k_i^*} \mathbf{x}_t + O_{\mathbb{P}}\left(\left(\sqrt{k_i^* - k_{i-1}^*}\right)\right), \quad 1 \le i \le \mathfrak{m}+1.$$

Since

$$\hat{e}_t = y_t - \mathbf{x}_t^\top \hat{\boldsymbol{\beta}}_T = e_t - \mathbf{x}_t^\top\left(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}^{(i)}\right), \quad \text{if } k_{i-1}^* + 1 \le t \le k_i^*, 1 \le i \le \mathfrak{m}+1, \tag{B.3}$$

we get

$$\begin{aligned} M_1(k_1^*) &= \frac{1}{\sqrt{k_1^*}}\left(\sum_{t=1}^{k_1^*}\mathbf{x}_t^\top(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}^{(1)}) - \frac{k_1^*}{T}\sum_{t=1}^{T}\mathbf{x}_t^\top(\hat{\boldsymbol{\beta}}_T - \tilde{\boldsymbol{\beta}}_T)\right) + O_{\mathbb{P}}(1) \\ &= \sqrt{k_1^*}\mathbf{c}^\top(\tilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta}^{(1)})(1 + o_{\mathbb{P}}(1)) + O_{\mathbb{P}}(1) \end{aligned} \tag{B.4}$$

and

$$M_{\mathfrak{m}+1}(k_{\mathfrak{m}}^*) = \sqrt{T - k^*\mathfrak{m}}\mathbf{c}^\top(\tilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta}^{(\mathfrak{m}+1)})(1 + o_{\mathbb{P}}(1)) + O_{\mathbb{P}}(1). \tag{B.5}$$

Using again Lemma B.1 and (B.2) we conclude for all $2 \le i \le \mathfrak{m}$ that

$$M_i(k_{i-1}^*, k_i^*) = \frac{k_i^* - k_{i-1}^*}{\sqrt{T}}\mathbf{c}^\top(\tilde{\boldsymbol{\beta}}_T - \boldsymbol{\beta}^{(i)})(1 + o_{\mathbb{P}}(1)) + O_{\mathbb{P}}(1). \tag{B.6}$$

The result now follows from (B.4)–(B.6). □

Proofs of Corollaries 2.1 and 2.2. The results can be derived from Theorem 2.4 by elementary calculations. □

C. Appendix: Proofs of Theorems 3.1 and 3.2

Proof of Theorem 3.1. Using (A.3) we get

$$S_T^2 = \sum_{t=1}^{T} e_t^2 - 2\left(\sum_{t=1}^{T} e_t\mathbf{x}_t\right)^\top(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) + \sum_{t=1}^{T}\left(\mathbf{x}_t^\top(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})\right)^2. \tag{C.1}$$

Assumption 2.1 implies along the lines of the proof of (A.5) that

$$\left|\sum_{t=1}^{T} e_t^2 - T\sigma^2\right| = O_{\mathbb{P}}(T^{1/2}). \tag{C.2}$$

Putting together Lemma A.1 and (A.2) we conclude

$$\left|\left(\sum_{t=1}^{T} e_t \mathbf{x}_t\right)^\top (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})\right| = O_{\mathbb{P}}(1),$$

and by the ergodic theorem we get

$$\sum_{t=1}^{T} \left(\mathbf{x}_t^\top (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})\right)^2 = O_{\mathbb{P}}(1).$$

This completes the proof of (3.2).
Lemma B.1 and (A.2) give

$$\left|\left(\sum_{t=1}^{T} e_t \mathbf{x}_t\right)^\top (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})\right| = O_{\mathbb{P}}(T), \quad \sum_{t=1}^{T} \left(\mathbf{x}_t^\top (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta})\right)^2 = O_{\mathbb{P}}(1),$$

and therefore (3.3) follows from (C.2).

Proof of Theorem 3.2. It follows from (A.3)

$$\begin{aligned}\hat{\gamma}_\ell = \frac{1}{T}\sum_{t=1}^{T-\ell} e_t e_{t+\ell} - \frac{1}{T}\sum_{t=1}^{T-\ell} e_t \mathbf{x}_{t+\ell}^\top (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) - \frac{1}{T}\sum_{t=1}^{T-\ell} e_{t+\ell} \mathbf{x}_t^\top (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) \\ + \frac{1}{T}\sum_{t=1}^{T-\ell} \mathbf{x}_t^\top (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}) \mathbf{x}_{t+\ell}^\top (\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}).\end{aligned} \tag{C.3}$$

It is easy to see that

$$\begin{aligned}&\mathbb{E}\left(\sum_{\ell=1}^{T-1} K(\ell/h) \frac{1}{T}\sum_{t=1}^{T-\ell} (e_t e_{t+\ell} - \mathbb{E} x_0 x_\ell)\right)^2 \\ &= \frac{1}{T^2}\sum_{\ell,\ell'=1}^{T-1}\sum_{t=1}^{T-\ell}\sum_{s=1}^{T-\ell'} K(\ell/h)K(\ell'/h)(\mathbb{E}[e_t e_{t+\ell} e_s e_{s+\ell'}] - \mathbb{E} e_0 e_\ell \mathbb{E} e_0 e_{\ell'}) \\ &= Q_T.\end{aligned}$$

Following the proof of Horváth and Rice (2015) we write by Assumption 3.2

$$\begin{aligned}Q_T &= O(1/T^2) \sum_{1\le t\le s\le T-1} \sum_{1\le \ell,\ell'\le ch} |\mathbb{E}[e_t e_{t+\ell} e_s e_{s+\ell'}] - \mathbb{E} e_0 e_\ell \mathbb{E} e_0 e_{\ell'}| \\ &= O(1/T^2) \sum_{1\le t\le s\le T-1} \sum_{1\le \ell,\ell'\le ch} |\mathbb{E}[e_t e_{t+\ell} e_s e_{s+\ell'}] - \mathbb{E} e_0 e_\ell \mathbb{E} e_0 e_{\ell'}| \\ &= O(1/T) \sum_{1\le v\le T-1} \sum_{1\le \ell,\ell'\le ch} |\mathbb{E}[e_0 e_\ell e_v e_{v+\ell'}] - \mathbb{E} e_0 e_\ell \mathbb{E} e_0 e_{\ell'}| \\ &= O(1/T)(Q_{T,1} + Q_{T,2}),\end{aligned}$$

where

$$Q_{T,1} = \sum_{(v,\ell,\ell')\in G_{T,1}} |\mathbb{E}[e_0 e_\ell e_v e_{v+\ell'}] - \mathbb{E} e_0 e_\ell \mathbb{E} e_0 e_{\ell'}|$$

and

$$Q_{T,2} = \sum_{(v,\ell,\ell')\in G_{T,2}} |\mathbb{E}[e_0 e_\ell e_v e_{v+\ell'}] - \mathbb{E}e_0 e_\ell \mathbb{E}e_0 e_{\ell'}|$$

with

$$G_{T,1} = \{(v,\ell,\ell') : h+1 \le v \le T, 1 \le \ell, \ell' \le h\},$$
$$G_{T,2} = \{(v,\ell,\ell') : 1 \le v \le h, 1 \le \ell, \ell' \le h\}.$$

Next we define

$$\bar{e}_{v,v-\ell-1} = f(\varepsilon_v, \varepsilon_{v-1}, \ldots, \varepsilon_{\ell+1}, \varepsilon'_\ell, \varepsilon'_{\ell-1}, \ldots)$$

and

$$\bar{e}_{v+\ell',v+\ell'-\ell-1} = f(\varepsilon_{v+\ell'}, \varepsilon_{v+\ell'-1}, \ldots, \varepsilon_{\ell+1}, \varepsilon'_\ell, \varepsilon'_{\ell-1}, \ldots),$$

where $\varepsilon'_\ell, -\infty < \ell < \infty$ are independent copies of ε_0, independent of $\varepsilon_j, -\infty < j < \infty$. It follows from the Bernoulli assumption that (e_0, e_ℓ) is independent of $(\bar{e}_{v,v-\ell-1}, \bar{e}_{v+\ell',v+\ell'-\ell-1})$. Also, according to the construction, $(\bar{e}_{v,v-\ell-1}, \bar{e}_{v-\ell',v+\ell'-\ell-1})$ and $(e_v, e_{v+\ell'})$ are identically distributed. Hence

$$\mathbb{E}[e_0 e_\ell e_v e_{v+\ell'}] - \mathbb{E}e_0 e_\ell \mathbb{E}e_0 e_{\ell'} = \mathbb{E}e_0 e_\ell [e_v e_{v+\ell'} - \bar{e}_{v,v-\ell-1}\bar{e}_{v+\ell',v+\ell'-\ell-1}].$$

It follows from Assumption 2.1 that

$$(\mathbb{E}(e_v - \bar{e}_{v,v-\ell-1})^4)^{1/4} \le c(v-\ell)^{-\alpha} \text{ and } (\mathbb{E}(e_{v+\ell'} - \bar{e}_{v+\ell',v+\ell'-\ell-1})^4)^{1/4} \le c(v+\ell'-\ell)^{-\alpha}$$

with some constant c for all $(v,\ell,\ell') \in G_{T,1}$. Hence

$$\begin{aligned}
&\mathbb{E}|e_0 e_\ell [e_v e_{v+\ell'} - \bar{e}_{v,v-\ell-1}\bar{e}_{v+\ell',v+\ell'-\ell-1}]| \\
&\quad \le \mathbb{E}|e_0 e_\ell e_v [e_{v+\ell'} - \bar{e}_{v+\ell',v+\ell'-\ell-1}]| + \mathbb{E}|e_0 e_\ell \bar{e}_{v+\ell',v+\ell'-\ell-1}[e_v - e_{v-\ell+1}]| \\
&\quad \le \left(\mathbb{E}e_0^4 \mathbb{E}e_\ell^4 \mathbb{E}e_v^4 \mathbb{E}|e_{v+\ell'} - \bar{e}_{v+\ell',v+\ell'-\ell-1}|^4\right)^{1/4} + \left(\mathbb{E}e_0^4 \mathbb{E}e_\ell^4 \mathbb{E}\bar{e}^4_{v+\ell',v+\ell'-\ell-1} \mathbb{E}|e_v - \bar{e}_{v,v-\ell-1}|^4\right)^{1/4} \\
&\quad \le c_*(v-\ell)^{-\alpha}
\end{aligned}$$

with some constant c_*. Hence elementary arguments give that

$$Q_{T,1} = O(1) \sum_{(v,\ell,\ell')\in G_{T,1}} (v-\ell)^{-\alpha} = O(h) \int_1^h \int_{h+1}^T (x-y)^{-\alpha} dx dy = O(h).$$

Next we note

$$Q_{T,2} \le \sum_{(v,\ell,\ell')\in G_{T,2}} |\mathbb{E}[e_0 e_\ell e_v e_{v+\ell'}]| + \sum_{(v,\ell,\ell')\in G_{T,2}} |\mathbb{E}e_0 e_\ell \mathbb{E}e_0 e_{\ell'}|$$

and

$$\sum_{(v,\ell,\ell')\in G_{T,2}} |\mathbb{E}e_0 e_\ell \mathbb{E}e_0 e_{\ell'}| \le h \left(\sum_{\ell=1}^\infty |\mathbb{E}e_0 e_\ell|\right)^2 < \infty.$$

Using the variables $e_{t,m}$ defined in Assumption 2.1 we write for all $0 \le s \le t \le v \le 2h$

$$\begin{aligned}
e_0 e_s e_t e_v =& e_0 e_{s,s}(e_t - e_{t,t-s})(e_v - e_{v,v-t}) + e_0 e_{s,s}(e_t - e_{t,t-s})e_{v,v-t} + e_0 e_{s,s} e_{t,t-s}(e_v - e_{v,v}) \\
&+ e_0 e_{s,s} e_{t,t-s} e_{v,v} + e_0(e_s - e_{s,s})(e_t - e_{t,t-s})e_v + e_0(e_s - e_{s,s})e_{t,t-s}e_{v,v-t} \\
&+ e_0(e_s - e_{s,s})e_{t,t-s}(e_v - e_{v,v-t}).
\end{aligned}$$

It follows from the definitions of $e_{t,m}$ that $e_{v,v-t}$ is independent of $e_0 e_{s,s}(e_t - e_{t,t-s})$, e_0 is independent of $e_{s,s}e_{t,t-s}e_{v,v}$, $e_{v,v-t}$ is independent of $e_0(e_s - e_{s,s})e_{t,t-s}$ and therefore

$$\mathbb{E}[e_0 e_{s,s}(e_t - e_{t,t-s})e_{v,v-t}] = 0, \ \ \mathbb{E}[e_0 e_{s,s} e_{t,t-s} e_{v,v}] = 0 \text{ and } \mathbb{E}[e_0(e_s - e_{s,s})e_{t,t-s}e_{v,v-t}] = 0.$$

Using Assumption 2.1 we obtain that

$$\begin{aligned}\sum_{0\le s\le t\le v\le 2h} &|\mathbb{E}e_0e_{s,s}(e_t-e_{t,t-s})(e_v-e_{v,v-t})| \\ &\le (\mathbb{E}e_0^4)^{1/2}\sum_{0\le s\le t\le v\le 2h}\left(\mathbb{E}(e_t-e_{t,t-s})^4\right)^{1/4}\left(\mathbb{E}(e_v-e_{v,v-t})^4\right)^{1/4} \\ &= O(h).\end{aligned}$$

Similarly

$$\sum_{0\le s\le t\le v\le 2h}|\mathbb{E}e_0(e_s-e_{s,s})(e_t-e_{t,t-s})e_v| = O(h),$$

$$\sum_{0\le s\le t\le v\le 2h}|\mathbb{E}e_0(e_s-e_{s,s})e_{t,t-s}(e_v-e_{v,v-t})| = O(h)$$

and

$$\sum_{0\le s\le t\le v\le 2h}|\mathbb{E}e_0e_{s,s}e_{t,t-s}(e_v-e_{v,v})| = O(1)\int_0^{2h}\int_s^{2h}\int_t^{2h}v^{-\alpha}dvdtds = O(h).$$

Thus we conclude that

$$\mathbb{E}\left(\sum_{\ell=1}^{T-1}\frac{1}{T}\sum_{t=1}^{T-\ell}(e_te_{t+\ell}-\mathbb{E}x_0x_\ell)\right)^2 = O\left(\frac{h}{T}\right)$$

and therefore

$$\sum_{\ell=1}^{T-1}K(\ell/T)\frac{1}{T}\sum_{t=1}^{T-\ell}e_te_{t+\ell} = \sum_{\ell=1}^{T-1}K(\ell/h)\frac{T-\ell}{T}\mathbb{E}e_0e_\ell + O_{\mathbb{P}}\left((h/T)^{1/2}\right). \tag{C.4}$$

Let

$$w_{\ell,i} = \frac{1}{T}\sum_{t=1}^{T-\ell}e_tx_{t+\ell,i}$$

and define

$$\hat{a}_{T,i} = \sum_{\ell=1}^{T-1}K(\ell/h)w_{\ell,i}.$$

Similarly to (C.4) we have for all $1\le i\le d$

$$\begin{aligned}\hat{a}_{T,i} &= \sum_{\ell=1}^{T-1}K(\ell/h)\frac{T-\ell}{T}\mathbb{E}e_0x_{\ell,i} + O_{\mathbb{P}}\left((h/T)^{1/2}\right) \\ &= O(1)\sum_{\ell=1}^{\infty}|\mathbb{E}e_0x_{\ell,i}| + O_{\mathbb{P}}\left((h/T)^{1/2}\right) \\ &= O_{\mathbb{P}}(1),\end{aligned} \tag{C.5}$$

since by Assumption 2.1 we have $\mathbb{E}e_0x_{\ell,i}| = \mathbb{E}(e_0-e_{0,\ell})x_{\ell,i}$ and therefore by the Cauchy–Schwartz inequality

$$\sum_{\ell=1}^{\infty}|\mathbb{E}e_0x_{\ell,i}| = \sum_{\ell=1}^{\infty}|\mathbb{E}(e_0-e_{0,\ell})x_{\ell,i}| \le \mathbb{E}(x_{0,i}^2)^{1/2}\sum_{\ell=1}^{\infty}(\mathbb{E}(e_0-e_{0,\ell}))^{1/2} < \infty.$$

Along the lines of the arguments leading to (C.4) and (C.5) we get

$$\hat{b}_{T,i} = \sum_{\ell=1}^{T-1} K(\ell/h) \frac{1}{T} \sum_{t=1}^{T-\ell} x_{t,i} e_{t+\ell} = O_{\mathbb{P}}(1) \text{ for all } 1 \le i \le d, \tag{C.6}$$

and

$$\hat{c}_{T,i,j} = \sum_{\ell=1}^{T-1} K(\ell/h) \frac{1}{T} \sum_{t=1}^{T-\ell} x_{t,i} x_{t+\ell,j} = O_{\mathbb{P}}(1) \text{ for all } 1 \le i, j \le d. \tag{C.7}$$

Thus we conclude by Lemma

$$\hat{\sigma}_T^2 = \sum_{\ell=1}^{T-1} K(\ell/h) \frac{T-\ell}{T} \mathbb{E} e_0 e_\ell + O_{\mathbb{P}}\left((h/T)^{1/2}\right) + O_{\mathbb{P}}(T^{-1/2}).$$

Assumption 3.2 yields

$$\sum_{\ell=1}^{T-1} K(\ell/h) \frac{T-\ell}{T} \mathbb{E} e_0 e_\ell = \sum_{\ell=1}^{\infty} \mathbb{E} e_0 e_\ell + \sum_{\ell=1}^{hc} \left(K(\ell/h) \frac{T-\ell}{T} - 1 \right) \mathbb{E} e_0 e_\ell - \sum_{\ell=hc+1}^{\infty} \mathbb{E} e_0 e_\ell$$

and

$$\begin{aligned} \left| \sum_{\ell=1}^{hc} \left(K(\ell/h) \frac{T-\ell}{T} - 1 \right) \mathbb{E} e_0 e_\ell \right| &\le \sum_{\ell=1}^{hc} |K(\ell/h) - 1| \, |\mathbb{E} e_0 e_\ell| + \sum_{\ell=1}^{hc} K(\ell/h) \left| \frac{T-\ell}{T} - 1 \right| |\mathbb{E} e_0 e_\ell| \\ &= O(1) \left(\frac{1}{h} + \frac{1}{T} \right) \sum_{\ell=1}^{\infty} \ell |\mathbb{E} e_0 e_\ell|. \end{aligned}$$

By Assumption 2.1 we get

$$\sum_{\ell=1}^{\infty} \ell |\mathbb{E} e_0 e_\ell| = \sum_{\ell=1}^{\infty} \ell |\mathbb{E} e_0 e_\ell| = \sum_{\ell=1}^{\infty} \ell |\mathbb{E} e_0 (e_\ell - e_{\ell,\ell})| \le (\mathbb{E} e_0^2)^{1/2} \sum_{\ell=1}^{\infty} \ell (\mathbb{E}(e_\ell - e_{\ell,\ell})^2)^{1/2} < \infty$$

and

$$\sum_{\ell=hc+1}^{\infty} |\mathbb{E} e_0 e_\ell| \le \frac{1}{h} \sum_{\ell=1}^{\infty} \ell |\mathbb{E} e_0 e_\ell|.$$

Since $\hat{\gamma}_0 - \mathbb{E} e_0^2 = O_P(T^{-1/2})$, (3.5) is proven.
We use (B.3) to prove (3.6). It follows from (C.5)–(C.7) and Lemma B.1 that

$$|\hat{a}_{T,i}| \|\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}\| = O_{\mathbb{P}}(1), \quad |\hat{b}_{T,i}| \|\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}\| = O_{\mathbb{P}}(1) \quad \text{and} \quad |\hat{c}_{T,i,j}| \|\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}\| = O_{\mathbb{P}}(1). \tag{C.8}$$

The result in (3.6) is an immediate consequence of (C.3), (C.4) and (C.8).

References

[1] Albin, J.M.P. and Jarušková. D.: On a test statistic for linear trend. *Extremes* **6**(2003), 247–258.

[2] Andrews, D.W.K.: Heteroskedacity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**(1991), 817–858.

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

[4] Andrews, D. W. and Monahan, J. C.: An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* **60**(1992), 953-966.

[5] Aue, A., Hörmann, S., Horváth, L., Hušková, M. and Steinebach, J.G.: Sequential testing for the stability of high–frequency portfolio betas. *Econometric Theory* **28**(2012a), 804–837.

[6] Aue, A., Hörmann, S., Horváth, L. and Hušková, M.: Dependent functional linear models with applications to monitoring structural change. *Statistica Sinica* **24**(2014), 1043–1073.
[7] Aue, A. and Horváth, L.: Structural breaks in time series. *Journal of Time Series Analysis* **34**(2013), 1–16.
[8] Aue, A., Horváth, L. and Hušková, M.: Extreme value theory for stochastic integrals of Legendre polynomials. *Journal of Multivariate Analysis* **100**(2009), 1029–1043.
[9] Aue, A., Horváth, L. and Hušková, M.: Segmenting mean–nonstationary time series via trending regressions. *Journal of Econometrics* **168**(2012b), 367–381.
[10] Aue, A., Horváth, L., Hušková, M. and Kokoszka, P.: Testing for changes in polynomial regression. *Bernoulli* **14**(2008), 637–660.
[11] Bai, J.: Likelihood ratio test for multiple structural changes. *Journal of Econometrics* **91**(1999), 299–323.
[12] Bai, J. and Perron, P.: Estimating and testing linear models with multiple structural changes. *Econometrica* **66**, 47–78.
[13] Barras, L., Scaillet, O. and Wermers, R.: False discoveries in mutual fund performance: measuring luck in estimated alphas. *Journal of Finance* **65**(2010), 179–216.
[14] Beach, S.L.: Semivariance decomposition of country–level return. *International Review of Economics and Finance* **20**(2011), 607–623.
[15] Breiman, L.: *Probability.* Addison Wesley, 1968.
[16] Busetti, F. and Taylor, A.R.M.: Test of stationarity against a change in persistence. *Journal of Econometrics* **123**(2004), 33–66.
[17] Caporale, T.: Time varying CAPM betas and banking sector risks. *Economics Letters* **115**(2012), 293–295.
[18] Carhart, M.M.: On persistence in mutual fund performance. *Journal of Finance* **52**(1997), 57–82.
[19] Cavaliere, G. and Taylor, A.M.R.: Testing for a change in persistence in the presence of non–stationary volatility. *Journal of Econometrics* **147**(2008), 84–98.
[20] Csörgő, M. and Horváth, L.: *Limit Theorems in Change–Point Analysis.* Wiley, 1997.
[21] Csörgő, M. and Révész, P.: *Strong Approximations in Probability and Statistics.* Academic Press, 1981.
[22] Davis, R.A., Huang, D. and Yao, Y.C.: Testing for a change in the parameter values and order of an autoregressive model. *Annals of Statistics* **23**(1995), 282–304.
[23] Fama, E.F. and French, K.R.: Luck versus skill in the cross–section of mutual fund returns. *Journal of Finance* **65**(2010), 1915–1947.
[24] Ferson, W. and Harvey, C.: The risk and predictabiliy of of international equity retruns. *The Review of Financial Studies* **6**(1993), 527–566.
[25] Francq, C. and Zakoian, J.-M.: *GARCH Models.* Wiley, 2010.
[26] Ghysels, E.: On stable factor structures in the pricing of risk: Do time–varying betas help or hurt? *Journal of Finance* **53**(1998), 549–573.
[27] Ghysels, E., Guay, A. and Hall, A.: Predictive tests for structural change with unknown breakpoint. *Journal of Econometrics* **82**(1997), 209–233.
[28] Gombay, E.: Change detection in linear regression with time series errors. *Canadian Journal of Statistics* **38**(2010), 65–79.
[29] Gombay, E. and Horváth, L.: Limit theorems for change in linear regression. *Journal of Multivariate Analysis* **48**(1994), 43–69.
[30] Hall, A.R., Han, S. and Boldea, O.: Inference regarding multiple structural changes in linear models with endogenous regressors. *Journal of Econometrics* **170**(2012), 281–302.
[31] Hansen, B.E.: Tests for parameter instability in regression with $I(1)$ processes. *Journal of Business and Economics Statistics* **10**(1992), 321–335.
[32] Harvey, C.: Time–varying conditional covariance in tests of asset pricing models. *Journal of Financial Economics* **24**(1989), 289–317.
[33] Harvey, D.I., Leybourne, S.J. and Taylor, A.M.R.: Modified tests for a change in persistence. *Journal of Econometrics* **134**(2006), 441–469.
[34] Hidalgo, J. and Seo, M.H.: Testing for structural stability in the whole sample. *Journal of Econometrics* **175**(2013), 84–93.

[35] Horváth, L. and Rice, G.: Extensions of some classical methods in change point analysis (with discussions) *TEST* **23**(2014), 219–290.
[36] Hušková, M.: Estimation of a change in linear models. *Statistics & Probability Letters* **26**(1996), 13–24.
[37] Jagannathan, R. and Wang, Z.: The conditional CAPM and the cross–section of expected returns. *Journal of Finance* **51**(1996), 3–54.
[38] Jarušková, D.: Testing appearance of polynomial trend. *Extremes* **1**(1999), 25–37.
[39] Jarušková, D.: Asymptotic distribution of a statistic testing a change in simple linear regression with equidistant design. *Statistics & Probability Letters* **64**(2003), 89–95.
[40] Jeng, J.–L.: *Analyzing Event Statistics in Corporate Finance.* Palgrave and MacMillen, 2015.
[41] Kejriwal, M. and Perron, P.: The limit distribution of the estimates in cointegrated regression models with multiple structural changes. *Journal of Econometrics* **146**(2008), 59–73.
[42] Lettau, M. and Ludvigson, S.: Resurracting the CAPM: A cross–sectional test when risk premia are time–varying. *Journal of Political Economy* **109**(2001), 1238–1287.
[43] Lintner, J.: The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics* **47**(1965), 13–37.
[44] McCabe, B.P.M.: A Multiple Decision Theory Analysis of Structural Stability in Regression. *Econometric Theory* **4** (1988), 499-508.
[45] McCabe, B.P.M. and Harrison, M.J.: Testing the Constancy of Regression Relationships Over Time Using Least Squares Residuals. *Journal of the Royal Statistical Society. Series C* **29**(1980), 142–148.
[46] Merton, R.: An intertemporal asset pricing model. *Econometrica* **41**(1973), 867–880.
[47] Newey, W.K. and West, K.D.: Automatic lag selection for covariance matrix estimation. *Review of Economic Studies* **61**(1994), 631–53.
[48] Parzen, E.: On choosing an estimate of the spectral density function of a stationary time series. *Annals of Mathematical Statistics* **28**(1957), 921-932.
[49] Politis, D.M.: Adaptive bandwidth choice. *Journal of Nonparametric Statistics* **25**(2003), 517–533.
[50] Politis, D.M. and Romano, J.P.: Bias–corrected nonparametric spectral estimation. *Journal of Time Series Analysis* **16**(1995), 67-103.
[51] Quandt, R.E.: Tests of the hypothesis that a linear regression system obeys two seperate regimes. *Journal of the American Statistical Association* **53**(1958), 873–880.
[52] Quandt, R.E.: The estimation of the parameters of a linear regression system obeying two seperate regimes. *Journal of the American Statistical Association* **55**(1960), 324–330.
[53] Sharpe, W.F.: Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* **19**(1964), 424–442.
[54] Stout, W. F.: *Almost Sure Convergence.* Academic Press, 1974.
[55] Taniguchi, A. and Kakizawa, Y.: *Asymptotic Theory of Statistical Inference for Time Series.* Springer, New York, 2000.

Table C.1. Standardized CUSUM for Sub-samples

	Standardized CUSUM	P-Value
Jul. 2001–Nov. 2010	3.46	3.08%
Nov. 2004–Nov. 2014	3.23	3.89%
Jul. 2001–Nov. 2004	0.87	34.32%
Dec. 2004–Nov. 2010	0.69	39.48%
Dec. 2010–Nov. 2014	0.37	49.86%

Table C.2. The segmentation for the AHGCX time series

	Jul. 2001–Nov. 2004	Dec. 2004–Nov. 2010	Dec. 2010–Nov. 2014
α	-1.09	0.65	-0.39
β^{M}	0.99	1.22	1.05
β^{HML}	0.05	-0.30	-0.41
β^{SMB}	0.48	0.35	0.40
β^{MOM}	0.25	0.18	0.09

Table C.3. Standardized CUSUM for Sub-samples

	Standardized CUSUM	P-Value
Feb. 1984–Aug. 2003	5.84	0.29%
Nov. 1996–Nov. 2014	6.02	0.24%
Feb. 1984–Nov. 1996	1.99	12.80%
Dec. 1996–Aug. 2003	0.66	40.30%
Sep. 2003–Nov. 2014	-0.62	84.54%

Table C.4. The segmentation for the IIVGX time series

	Feb. 1984 – Nov. 1996	Dec. 1996 – Aug. 2003	Sep. 2003 – Nov. 2014
α	0.11	-1.21	-0.06
β^{M}	0.83	0.80	1.01
β^{HML}	-0.06	0.10	-0.03
β^{SMB}	-0.14	-0.07	-0.10
β^{MOM}	0.03	-0.11	0.01

Lajos Horváth, Department of Mathematics, University of Utah, Salt Lake City, UT 84112–0090 USA

William Pouliot, Department of Economics, University of Birmingham, Birmingham, UK, B15 2TT

Shixuan Wang, Department of Economics, University of Birmingham, Birmingham, UK, B15 2TT