Journal of Multivariate Analysis 115 (2013) 374–395

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Robust monitoring of CAPM portfolio betas

Ondřej Chochola [a], Marie Hušková [a], Zuzana Prášková [a], Josef G. Steinebach [b,*]

[a] Charles University in Prague, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Sokolovská 83, CZ – 18675 Praha 8, Czech Republic

[b] Universität zu Köln, Mathematisches Institut, Weyertal 86 – 90, D – 50931 Köln, Germany

ARTICLE INFO

Article history:
Received 2 May 2012
Available online 19 November 2012

AMS 2000 subject classifications:
60F17
60G10
60J65
62F35
62L10
62P05

Keywords:
Robust monitoring
Capital asset pricing model
Portfolio beta
M-estimate
Change-point detection

ABSTRACT

Some robust sequential procedures for the detection of structural breaks in the Capital Asset Pricing Model (CAPM) are proposed and studied. Most of the existing procedures for this model are based on ordinary least squares (OLS) estimates. Here we propose a class of cumulative sum (CUSUM)-type procedures based on M-estimates and partial weighted sums of M-residuals. The theoretical results are accompanied by a simulation study that compares the proposed procedures with those based on OLS estimates. An application to a real data set is also presented.

1. Introduction and statistical framework

The Capital Asset Pricing Model (CAPM), introduced by Sharpe [25] and subsequently modified by many authors (see, e.g. Lintner [22], Merton [24] and others), is an important and widely used model for evaluating the risk of a portfolio of assets with respect to the market risk. Despite of some shortcomings pointed out by theoreticians and practitioners as well, the wide-spread use of the CAPM is also well-documented (cf., e.g., the report of Martin and Simin [23]). A main advantage of the model is its simplicity in describing the sensitivity of an asset's risk against the market risk, which is essentially expressed through one parameter, the (so-called) portfolio beta. On the other hand, it is also well-known that the corresponding pricing of a portfolio asset heavily relies on the constancy of the betas over time. Confer, for example, the discussion in Ghysels [13] and recently Caporale [7]. So, it may be of great interest to find out whether portfolio betas change significantly over time or not. The latter was a main motivation in Aue et al. [1] for constructing a sequential monitoring procedure for the testing of the stability of portfolio betas, taking also high-frequency data into account. Along the lines of Chu et al. [9], the corresponding stopping rules of Aue et al. [1] are based on comparing the (ordinary) least squares estimate (OLS) of the beta from a historical data set (training period) to that from sequentially incoming new observations. A structural break (change) in the model is then confirmed when the beta significantly changes, that is, when the newly estimated beta exceeds a critical distance from the historical one.

* Corresponding author.
E-mail addresses: chochola@karlin.mff.cuni.cz (O. Chochola), huskova@karlin.mff.cuni.cz (M. Hušková), praskova@karlin.mff.cuni.cz (Z. Prášková), jost@math.uni-koeln.de (J.G. Steinebach).

0047-259X/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2012.10.019

However, it is well-known that OLS estimators are sensitive with respect to outliers and deviations from normality assumptions. Concerning the possible application of the CAPM this has led to an extensive discussion and numerous suggestions for "robustifying" the use of beta estimates in the prediction of portfolio risks (confer, e.g., Genton and Ronchetti [12] and Martin and Simin [23] together with the works mentioned therein). Indeed, this has also motivated our present paper in which we propose a robust monitoring procedure for testing the stability of CAPM portfolio betas. In doing so, we try to take into account various aspects of the model which should allow for a broader applicability in practice. First of all we suggest a multivariate approach allowing for dependencies within the portfolio. Second we work with a time series model describing possible dependencies over time, and last but not least our approach is based on (multivariate) M-estimators in order to reduce the sensitivity of the statistical decisions against outliers and non-normality assumptions. For some related work on the M-estimation in linear models with dependent errors confer also Wu [26].

In view of the latter aspects the monitoring procedure proposed below extends other sequential testing procedures for detecting an instability of parameters in regression models when a training sample is available (e.g. Chu et al. [9] or Aue et al. [2]), which are typically based on OLS estimators and related L_2-residuals. Here we shall make use of general M-residuals which to the best of our knowledge have not been applied in this context. Only in case of a univariate linear regression model with independent observations, Koubková [21] already studied some similar robust sequential procedures based on cumulative sum (CUSUM)-type test statistics. We would also like to mention that our procedure can be extended to general multivariate linear regression models or even to functional data setups, but this is beyond the scope of the present work and will be studied elsewhere.

In the sequel our statistical framework will be as follows. We consider the model

$$\mathbf{r}_i = \boldsymbol{\alpha}_i + \boldsymbol{\beta}_i r_{iM} + \boldsymbol{\varepsilon}_i, \quad i \in \mathbb{Z}, \tag{1.1}$$

where $\mathbf{r}_i = (r_{i,1}, \dots, r_{i,d})^T$ is a d-dimensional vector of daily log-returns at time i, r_{iM} is the log-return of the market portfolio at time i, and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i,1}, \dots, \varepsilon_{i,d})^T$ are d-dimensional error terms. The $\boldsymbol{\alpha}_i$'s and $\boldsymbol{\beta}_i$'s are d-dimensional unknown parameters, and the $\boldsymbol{\beta}_i$'s are the parameters of interest, usually called the "portfolio betas". Note that the sequence $\{(\mathbf{r}_i, r_{iM})\}$ is a $(d+1)$-dimensional time series satisfying certain conditions to be specified below.

We assume that a training sample of size m with no instabilities is available, i.e.,

$$\boldsymbol{\alpha}_1 = \cdots = \boldsymbol{\alpha}_m =: \boldsymbol{\alpha}_0, \qquad \boldsymbol{\beta}_1 = \cdots = \boldsymbol{\beta}_m =: \boldsymbol{\beta}_0, \tag{1.2}$$

where $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}_0$ are unknown parameters. The problem of the instability of the portfolio betas is formulated as a testing problem, that is, we want to test the null hypothesis

$$H_0 : \boldsymbol{\beta}_1 = \cdots = \boldsymbol{\beta}_m = \boldsymbol{\beta}_{m+1} = \cdots$$

of no change versus the alternative

$$H_A : \boldsymbol{\beta}_1 = \cdots = \boldsymbol{\beta}_{m+k^*} \neq \boldsymbol{\beta}_{m+k^*+1} = \cdots$$

of a structural break at an unknown change-point $k^* = k^*_m$.

For later convenience we reformulate our model as follows:

$$r_{i,j} = \alpha_j^0 + \beta_j^0 \widetilde{r}_{iM} + (\alpha_j^1 + \beta_j^1 \widetilde{r}_{iM})\delta_m I\{i > m + k^*\} + \varepsilon_{i,j}, \quad j = 1, \dots, d,\ i = 1, 2, \dots, \tag{1.3}$$

where $k^* = k^*_m$ is the change-point, $\alpha_j^0, \beta_j^0, \alpha_j^1, \beta_j^1, \delta_m$ are unknown parameters, and

$$\widetilde{r}_{iM} = r_{iM} - \overline{r}_{mM}, \quad \text{with } \overline{r}_{mM} = \frac{1}{m}\sum_{i=1}^{m} r_{iM}. \tag{1.4}$$

Our test procedures will be generated by convex loss functions $\varrho_1, \dots, \varrho_d$ with a.s. derivatives $\varrho_j' = \psi_j$ called score functions having further properties to be specified later. The estimators $\widehat{\alpha}_{jm} = \widehat{\alpha}_{jm}(\psi_j)$, $\widehat{\beta}_{jm} = \widehat{\beta}_{jm}(\psi_j)$ of $\alpha_j^0,\ \beta_j^0$ based on the training sample are defined as minimizers of

$$\sum_{i=1}^{m} \varrho_j(r_{i,j} - a_j - b_j \widetilde{r}_{iM}) \tag{1.5}$$

w.r.t. a_j, b_j for $j = 1, \dots, d$.

Generally, having $m + k$ observations (the training sample of size m plus k new observations) it would be natural to construct the test procedure via comparing estimators of $\beta_1^0, \dots, \beta_d^0$ based on $\mathbf{r}_1, \dots, \mathbf{r}_m$ and on $\mathbf{r}_{m+1}, \dots, \mathbf{r}_{m+k}$, respectively. This, however, would be computationally quite demanding. Therefore we propose a test procedure based on functionals of partial sums of weighted M-residuals which is asymptotically equivalent.

The M-residuals to be used are defined as follows:

$$\boldsymbol{\psi}(\widehat{\boldsymbol{\varepsilon}}_i) = (\psi_1(\widehat{\varepsilon}_{i,1}), \dots, \psi_d(\widehat{\varepsilon}_{i,d}))^T \tag{1.6}$$

with

$$\widehat{\boldsymbol{\varepsilon}}_i = (\widehat{\varepsilon}_{i,1}, \dots, \widehat{\varepsilon}_{i,d})^T,$$
$$\widehat{\varepsilon}_{i,j} = r_{i,j} - \widehat{\alpha}_{jm} - \widetilde{r}_{iM}\widehat{\beta}_{jm}. \tag{1.7}$$

A suitable test statistic based on the first $m+k$ observations is

$$\widehat{Q}(k, m) = \left(\frac{1}{\sqrt{m}} \sum_{i=m+1}^{m+k} \widetilde{r}_{iM} \boldsymbol{\psi}(\widehat{\boldsymbol{\varepsilon}}_i)\right)^T \widehat{\boldsymbol{\Sigma}}_m^{-1} \left(\frac{1}{\sqrt{m}} \sum_{i=m+1}^{m+k} \widetilde{r}_{iM} \boldsymbol{\psi}(\widehat{\boldsymbol{\varepsilon}}_i)\right) \tag{1.8}$$

where the matrix $\widehat{\boldsymbol{\Sigma}}_m$ is an estimator of the asymptotic variance (matrix)

$$\boldsymbol{\Sigma} = \lim_{m\to\infty} var \left\{ \frac{1}{\sqrt{m}} \sum_{i=1}^{m} (r_{iM} - Er_{iM}) \boldsymbol{\psi}(\boldsymbol{\varepsilon}_i) \right\} \tag{1.9}$$

based on the first m observations. Details will be discussed later.

We reject the null hypothesis as soon as the test statistic exceeds a critical level for the first time, i.e., when

$$\widehat{Q}(k, m)/q_\gamma(k/m) \geq c$$

for an appropriately chosen $c = c_\gamma(\alpha)$, where $q_\gamma(t),\ t \in (0, \infty)$ is a suitable boundary (weight) function. In this case we stop the procedure and confirm a structural break, otherwise we continue monitoring. The associated stopping rule is given by

$$\tau_m = \tau_m(\gamma) = \inf\{1 \leq k \leq \lfloor mT \rfloor : \widehat{Q}(k, m)/q_\gamma(k/m) \geq c\}, \tag{1.10}$$

with $\inf \emptyset := \infty$. Here T is a fixed positive number, that is, we have a so-called *closed-end procedure*. This is very practical since in applications the upper bound for the maximum number of possible observations is usually specified a priori. The following class of the weight functions q_γ can be used, e.g.,

$$q_\gamma(t) = (1+t)^2 \left(\frac{t}{t+1}\right)^{2\gamma}, \quad t \in (0, \infty), \tag{1.11}$$

where γ is a tuning constant taking values in $[0, 1/2)$. The critical value c will be chosen such that, under H_0, for $\alpha \in (0, 1)$ (fixed),

$$\lim_{m\to\infty} P\big(\tau_m < \infty\big) = \alpha, \tag{1.12}$$

i.e., the overall asymptotic level (false alarm rate) is α and, under H_A,

$$\lim_{m\to\infty} P\big(\tau_m < \infty\big) = 1, \tag{1.13}$$

i.e., the test is consistent (has asymptotic power 1).

The rest of the paper is organized as follows. The main results including the assumptions and limit properties of the test procedures are presented and discussed in Section 2. Section 3 reports on the results of a small simulation study and an application to a real data set. The proofs of our main results are given in Section 4, whereas Section 5 contains some auxiliary lemmas to be used in the proofs.

2. Assumptions and main results

We start with formulating the assumptions on the sequence $\{(\boldsymbol{\varepsilon}_{i,1}, \dots, \boldsymbol{\varepsilon}_{i,d},\ r_{iM})\}_{i\in\mathbb{Z}}$ and on the loss functions $\varrho_1, \dots, \varrho_d$ (or equivalently on the score functions $\psi_1, \dots, \psi_d$).

We assume on ψ_j and the distribution function F_j of $\varepsilon_{i,j},\ j = 1, \dots, d$:

(A.1) ψ_j are nondecreasing functions, $\lambda_j(t) = -\int \psi_j(x-t)dF_j(x),\ t \in \mathbb{R}, \lambda_j(0) = 0, \lambda_j'(0) > 0, \lambda_j'(t)$ exists in a neighborhood of 0 and is Lipschitz in a neighborhood of 0, e.g., for $|t| \leq c_0$ with some $c_0 > 0$.

(A.2) $\int |\psi_j(t)|^2 dF_j(t) < \infty$ and

$$\int |\psi_j(x+t_2) - \psi_j(x+t_1)|^2 dF_j(x) \leq C_1|t_2 - t_1|^\kappa, \quad |t_1|, |t_2| \leq c_0$$

for some $1 \leq \kappa \leq 2, c_0, C_1 > 0$.

[It can be assumed that c_0 is the same in both assumptions.] Note that $\kappa = 1$ corresponds to L_1.

These assumptions are quite standard in robust statistics. For further information on the choice of ψ_j see the classical works on robust methods, e.g., Jurečková and Sen [18] and Huber [17] for the univariate situations and the papers by Koenker and Portnoy [19] and Bai et al. [3,4] for the multivariate ones.

Let us recall some of the most often considered ψ_j-functions. The classical choice $\psi_j(x) = x,\ x \in \mathbb{R}^1$, leads to the ordinary least squares (OLS) and L_2-residuals. A choice of $\psi_j(x) = \operatorname{sign} x,\ x \in \mathbb{R}^1$, leads to L_1-estimators and L_1-residuals. Huber [17] introduced $\psi_j(x) = x I\{|x| \le K\} + K \operatorname{sign} x I\{|x| > K\},\ x \in \mathbb{R}^1$, for some $K > 0$, which is one of the most often used score functions, usually known as the Huber function.

For a vector-valued random variable $\mathbf{X}$ define

$$\|\mathbf{X}\|_p = \left(E|\mathbf{X}|^p\right)^{1/p}, \quad p \ge 1,$$

the L_p- norm of $\mathbf{X}$, where $|\mathbf{X}|$ denotes the Euclidean norm of $\mathbf{X}$.

Concerning the assumptions on $\{r_{iM}\}$ and $\{\boldsymbol{\varepsilon}_i\}$ we follow the setup in Aue et al. [1]:

(B.1) For any $i \in \mathbb{Z}$, $r_{iM} = h(\boldsymbol{\xi}_i, \boldsymbol{\xi}_{i-1}, \dots)$, where $h(\cdot)$ is a measurable function, $\{\boldsymbol{\xi}_i\}$ is a sequence of i.i.d. random vectors with dimension q_1, and $E|r_{0M}|^{2+\Delta} < \infty$ for some $\Delta > 0$.
[Note that $\{r_{iM} : i \in \mathbb{Z}\}$ is a stationary and ergodic sequence.]

(B.2) For any $i \in \mathbb{Z}$, $\boldsymbol{\varepsilon}_i = \mathbf{g}(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_{i-1}, \dots)$, where $\mathbf{g}(\cdot)$ is a measurable function, $\{\boldsymbol{\zeta}_i\}$ is a sequence of i.i.d. random vectors with dimension q_2 and some further properties to be specified later.
[Note that $\{\boldsymbol{\varepsilon}_i : i \in \mathbb{Z}\}$ is also a stationary and ergodic sequence.]

(B.3) The sequences $\{\boldsymbol{\xi}_i\}$ and $\{\boldsymbol{\zeta}_i\}$ are independent.

(B.4) For all $i \in \mathbb{Z}$,

$$\sum_{L=1}^{\infty} \|r_{iM} - r_{iM}^{(L)}\|_{2+\Delta} < \infty,$$

where

$$r_{iM}^{(L)} = h(\boldsymbol{\xi}_i, \boldsymbol{\xi}_{i-1}, \dots, \boldsymbol{\xi}_{i-L+1}, \boldsymbol{\xi}_{i-L}^{(L)}, \boldsymbol{\xi}_{i-L-1}^{(L)}, \dots),$$

with $\boldsymbol{\xi}_{i-L}^{(L)}, \boldsymbol{\xi}_{i-L-1}^{(L)}, \dots$ being i.i.d. with the same distribution as $\boldsymbol{\xi}_0$ and independent of $\{\boldsymbol{\xi}_i\}$.
[Note that $r_{iM}^{(L)} \overset{\mathcal{D}}{=} r_{iM} \overset{\mathcal{D}}{=} r_{0M}$ for all $i \in \mathbb{Z}$ and $L \ge 1$.]

(B.5) For $\boldsymbol{\psi}(\boldsymbol{\varepsilon}_i) = (\psi_1(\varepsilon_{i,1}), \dots, \psi_d(\varepsilon_{i,d}))^T,\ i \in \mathbb{Z}$, it holds that

$$\sum_{L=1}^{\infty} \sup_{|\mathbf{a}| \le a_0} \|\boldsymbol{\psi}(\boldsymbol{\varepsilon}_i - \mathbf{a}) - \boldsymbol{\psi}(\boldsymbol{\varepsilon}_i^{(L)} - \mathbf{a})\|_2 < \infty$$

for some $a_0 > 0$, where

$$\boldsymbol{\varepsilon}_i^{(L)} = \mathbf{g}(\boldsymbol{\zeta}_i, \boldsymbol{\zeta}_{i-1}, \dots, \boldsymbol{\zeta}_{i-L+1}, \boldsymbol{\zeta}_{i-L}^{(L)}, \boldsymbol{\zeta}_{i-L-1}^{(L)}, \dots)$$

with $\boldsymbol{\zeta}_{i-L}^{(L)}, \boldsymbol{\zeta}_{i-L-1}^{(L)}, \dots$ being i.i.d. with the same distribution as $\boldsymbol{\zeta}_0$ and independent of $\{\boldsymbol{\zeta}_i\}$.

The above assumptions are motivated by the work of Hörmann and Kokoszka [14] on the concept of L_p-m-approximability, in which also the relation to other types of dependencies is discussed as well as various examples are presented. In this respect, our asymptotic analysis below differs from the approach of Wu [26] who also relaxed the independence assumption in the classical M-estimation theory and allowed for a more general class of dependent errors.

Next we present our results on the limit behavior of the test procedures both under the null hypothesis H_0 as well as under the alternative H_A.

2.1. Asymptotic results

Theorem 2.1. *Let Assumptions* (A.1)–(A.2), (B.1)–(B.5) *and* (1.11) *with* $\gamma \in [0, 1/2)$ *be satisfied and*

$$\widehat{\boldsymbol{\Sigma}}_m - \boldsymbol{\Sigma} = o_P(1) \quad (m \to \infty), \tag{2.1}$$

where

$$\begin{aligned}\boldsymbol{\Sigma} &= \lim_{m\to\infty} var\left\{\frac{1}{\sqrt{m}}\sum_{i=1}^{m}(r_{iM} - Er_{iM})\boldsymbol{\psi}(\boldsymbol{\varepsilon}_i)\right\}\\ &= E[(r_{0M} - Er_{0M})^2\boldsymbol{\psi}(\boldsymbol{\varepsilon}_0)\boldsymbol{\psi}(\boldsymbol{\varepsilon}_0)^T] + \sum_{i=1}^{\infty} E[(r_{0M} - Er_{0M})(r_{iM} - Er_{iM})(\boldsymbol{\psi}(\boldsymbol{\varepsilon}_0)\boldsymbol{\psi}(\boldsymbol{\varepsilon}_i)^T + \boldsymbol{\psi}(\boldsymbol{\varepsilon}_i)\boldsymbol{\psi}(\boldsymbol{\varepsilon}_0)^T)],\end{aligned} \tag{2.2}$$

and $\boldsymbol{\Sigma}$ is a positive definite matrix. Then, under the null hypothesis H_0,

$$\max_{1\le k\le\lfloor mT\rfloor}\left(\frac{\widehat{Q}(k,m)}{q_\gamma(k/m)}\right)\xrightarrow{\mathcal{D}}\sup_{0<t<T/(T+1)}\left(\frac{\sum_{j=1}^{d}W_j^2(t)}{t^{2\gamma}}\right)\quad(m\to\infty),$$

where $\{W_j(t),\ t\in[0,1]\}$, $j=1,\dots,d$, are independent (standard) Brownian motions (Wiener processes).

The proof of Theorem 2.1 is postponed to Section 4.

It follows from Assumptions (A.1)–(A.2) and (B.1)–(B.5) that $\{r_{iM}\}$ and $\{\boldsymbol{\psi}(\boldsymbol{\varepsilon}_i)\}$ are independent sequences. Then Lemma 2.1 and Theorem 4.2 in Hörmann and Kokoszka [14] imply that the series in (2.2) converges (component-wise) absolutely.

Now we turn to the model under local alternatives, i.e.

$$r_{i,j}=\alpha_j^0+\beta_j^0\widetilde{r}_{iM}+(\alpha_j^1+\beta_j^1\widetilde{r}_{iM})\delta_m I\{i>m+k^*\}+\varepsilon_{i,j},\quad j=1,\dots,d,\ i=1,2,\dots,\tag{2.3}$$

with $\delta_m\to 0$ and $k^*<\lfloor mT\rfloor$.

Theorem 2.2. *Let Assumptions* (A.1)–(A.2), (B.1)–(B.5) *and* (1.11) *with $\gamma\in[0,1/2)$ be satisfied and*

$$\widehat{\boldsymbol{\Sigma}}_m-\boldsymbol{\Sigma}=o_P(1)\quad(m\to\infty),$$

where $\boldsymbol{\Sigma}$ is as in Theorem 2.1.

(i) *Under* (2.3) *with $\delta_m=m^{-1/2}$ and $k^*=\lfloor ms\rfloor$, $0<s<T$,*

$$\max_{1\le k\le\lfloor mT\rfloor}\left(\frac{\widehat{Q}(k,m)}{q_\gamma(k/m)}\right)\xrightarrow{\mathcal{D}}\sup_{0<t<T/(T+1)}\left(\frac{\sum_{j=1}^{d}(W_j(t)-h_j(t,s))^2}{t^{2\gamma}}\right)\quad(m\to\infty),$$

where $\{W_j(t),\ t\in(0,1)\}$, $j=1,\dots,d$, are independent Brownian motions and

$$\mathbf{h}(t,s)=(h_1(t,s),\dots,h_d(t,s))^T,\quad 0<s<T,\ 0<t<T/(T+1),$$
$$\mathbf{h}(t,s)=\max(0,t/(1-t)-s)var\{r_{0M}\}\boldsymbol{\Sigma}^{-1/2}(\lambda_1'(0)\beta_1^1,\dots,\lambda_d'(0)\beta_d^1)^T,\quad 0<s<T,\ 0<t<T/(T+1).$$

(ii) *Under* (2.3) *with $\delta_m\to 0$, $|\delta_m|m^{1/2}\to\infty$, $\liminf_{m\to\infty}(\lfloor mT\rfloor-k^*)/m>0$, and $\beta_j^1\ne 0$ for at least one j,*

$$\max_{1\le k\le\lfloor mT\rfloor}\left(\frac{\widehat{Q}(k,m)}{q_\gamma(k/m)}\right)\xrightarrow{P}\infty\quad(m\to\infty).$$

The proof of Theorem 2.2 is also postponed to Section 4.

Remark 2.1. (a) By Theorem 2.1, the assertion (1.12) holds true if $c_\gamma(\alpha)$ satisfies

$$P\left(\sup_{0<t<T/(T+1)}\left(\frac{\sum_{j=1}^{d}W_j^2(t)}{t^{2\gamma}}\right)\ge c_\gamma(\alpha)\right)=\alpha.$$

But the explicit form of the distribution of $\sup_{0<t<T/(T+1)}\sum_{j=1}^{d}W_j^2(t)/t^{2\gamma}$ is unknown, so it does not provide an immediate approximation of the critical values. However, $c_\gamma(\alpha)$ can either be obtained by simulation of the limit distribution or by an application of a suitable form of bootstrap based on the training sample.

(b) Theorem 2.2 (ii) implies the consistency of the test, i.e., the validity of (1.13) (asymptotic power 1).

(c) Theorem 2.2 (i) deals with so-called contiguous alternatives. As expected, asymptotically we have the maximum of weighted sums of squares of shifted Wiener processes, where the shifts depend on the change-point, the amount of change and also on the choice of the loss functions $\varrho_1,\dots,\varrho_d$ (through $\lambda_1'(0),\dots,\lambda_d'(0)$).

Notice also that the limit distribution in Theorem 2.2 (i) is only sensitive w.r.t. a change in the β_j's, but not w.r.t. a change in the α_j's. Moreover, on checking the proof one can conclude that, in case of a local change in the α_j's only, the limit distribution is the same as under H_0.

2.2. Estimation of the variance matrix

In this section we deal with an estimator of the asymptotic variance (matrix) $\boldsymbol{\Sigma}$ as given in (2.2). Notice that $\boldsymbol{\Sigma} = \sum_{k=-\infty}^{\infty} \boldsymbol{\Gamma}_k$, where $\boldsymbol{\Gamma}_k = E[(r_{0M} - Er_{0M})(r_{kM} - Er_{kM})\boldsymbol{\psi}(\boldsymbol{\varepsilon}_0)\boldsymbol{\psi}(\varepsilon_k)^T]$ for $k \geq 0$ and $\boldsymbol{\Gamma}_{-k} = \boldsymbol{\Gamma}_k^T$.

We consider an estimator of $\boldsymbol{\Sigma}$ based on the first m observations defined as

$$\widehat{\boldsymbol{\Sigma}}_m = \sum_{|k|<q} \omega_q(k)\widehat{\boldsymbol{\Gamma}}_k \tag{2.4}$$

where $q = q(m)$ and ω_q is a kernel function specified below, $\widehat{\boldsymbol{\Gamma}}_k$ is the k-th lag sample covariance corresponding to $\boldsymbol{\Gamma}_k$, i.e.,

$$\widehat{\boldsymbol{\Gamma}}_k = \begin{cases} \dfrac{1}{m}\displaystyle\sum_{i=1}^{m-k} \widetilde{r}_{iM}\widetilde{r}_{i+k,M}\boldsymbol{\psi}(\widehat{\boldsymbol{\varepsilon}}_i)\boldsymbol{\psi}(\widehat{\boldsymbol{\varepsilon}}_{i+k})^T, & k \geq 0, \\ \widehat{\boldsymbol{\Gamma}}_{-k}^T, & k < 0, \end{cases} \tag{2.5}$$

with the $\widetilde{r}_{iM}$ as given in (1.4) and $\boldsymbol{\psi}(\widehat{\varepsilon}_i)$ based on the M-residuals defined in (1.6) and (1.7).

We shall work with the Bartlett kernel, i.e.,

$$\omega_q(x) = (1 - |x|/q)\, I\{|x| \leq q\}, \quad x \in \mathbb{R}. \tag{2.6}$$

Theorem 2.3. *Let Assumptions* (A.1), (A.2), *and* (B.1)–(B.5) *be satisfied with* $\Delta \geq \kappa$, κ *from Assumption* (A.2). *Let* $\widehat{\boldsymbol{\Sigma}}_m$ *be the estimator of* $\boldsymbol{\Sigma}$ *given in* (2.4) *with the kernel* (2.6) *such that, as* $m \to \infty$, $q(m) \to \infty$ *and* $q(m)/m^{\min\{1/2,\kappa/4\}} \to 0$. *Then*

$$\widehat{\boldsymbol{\Sigma}}_m = \boldsymbol{\Sigma} + o_P(1) \quad (m \to \infty).$$

3. Simulations and applications

In this section we present some results from a small simulation study as well as an application to a real data set in order to illustrate the finite sample performance of our monitoring procedure based on the test statistic (1.8). For the computations we used the statistical software package R (R Development Core Team, 2009), version 2.13.1.

First we give a table of simulated critical values for the asymptotic distribution of the test statistic as specified in Theorem 2.1. Due to the scaling property of the Wiener process, it is enough to determine the critical values (say) $c_{\gamma,\infty}(\alpha)$ of the functional

$$\sup_{0<t<1} \left(\frac{\sum_{j=1}^{d} W_j^2(t)}{t^{2\gamma}} \right), \tag{3.1}$$

corresponding to the so-called *open-end* monitoring procedure. Then one can easily check that the critical values (say) $c_{\gamma,T}(\alpha)$ of the statistic

$$\sup_{0<t<T/(T+1)} \left(\frac{\sum_{j=1}^{d} W_j^2(t)}{t^{2\gamma}} \right)$$

for the *closed-end* procedure considered in Theorem 2.1 satisfy the relation

$$c_{\gamma,T}(\alpha) = \left(\frac{T}{T+1}\right)^{1-2\gamma} c_{\gamma,\infty}(\alpha). \tag{3.2}$$

Critical values $c_{\gamma,\infty}$ for $d = 1, 2$ have already been determined in [21]. Here we present an extended table for $d = 2, 3, 4, 5$ (cf. Table 1) for choices $\gamma = 0, 0.15, 0.25, 0.40, 0.49$ of the tuning constant and nominal levels $\alpha = 1\%, 5\%, 10\%$ of the test. Thereby the supremum of the functional given in (3.1) has been approximated by a maximum over a grid of 25,000 equidistant points, and 100,000 repetitions have been run.

In our simulation study we made the following choices:

- $\alpha = 5\%$, $d = 2$, $c_{\gamma,T}(\alpha)$ from Table 1 rescaled according to (3.2);
- ψ_j-functions corresponding to the L_2-, L_1- and Huber-estimators as described in Section 2 (in the Huber case with constant $K = 1.345\, (var\ \varepsilon_{i,j})^{1/2}$, where 1.345 is the default value in the R-package);
- $m = 100, 200, 400$; $\gamma = 0, 0.25, 0.45$;
- $\boldsymbol{\alpha}_0 = (0.5, 0.5)^T$, $\boldsymbol{\beta}_0 = (0.5, 0.5)^T$;
- Bartlett kernel variance estimator with $q = 4, 10, 20$.

Table 1
Simulated critical values for the functional given in (3.1).

d	$\alpha(\%) \setminus \gamma$	0	0.15	0.25	0.40	0.45	0.49
	10	5.83300	6.16964	6.54486	7.79693	8.90706	10.97680
2	5	7.27319	7.62029	8.01801	9.24979	10.38189	12.51981
	1	10.47212	10.81526	11.18947	12.41796	13.58373	16.08758
	10	7.55347	7.91567	8.33422	9.69223	10.89566	13.24342
3	5	9.15817	9.51428	9.92618	11.27827	12.47845	14.93875
	1	12.64423	12.97544	13.35888	14.71475	15.93770	18.61511
	10	9.15704	9.54268	9.96759	11.40482	12.68321	15.28504
4	5	10.89252	11.26607	11.67221	13.12474	14.41193	17.05890
	1	14.65064	15.00585	15.43069	16.88893	18.13029	20.88200
	10	10.63242	11.04519	11.48214	12.97519	14.35397	17.13813
5	5	12.47376	12.87663	13.31469	14.80208	16.16445	19.02006
	1	16.43966	16.84611	17.32441	18.86821	20.13233	23.11929

Table 2
Empirical sizes at nominal level $\alpha = 5\%$, $T = 10$, dependent observations.

		q	$m \setminus \gamma$	L_2			Huber			L_1		
				0	0.25	0.45	0	0.25	0.45	0	0.25	0.45
			100	8.2	9.6	8.7	7.3	8.5	7.2	7.0	8.1	5.9
$r_{iM} \sim$ AR(1),	$\boldsymbol{\varepsilon}_i \sim$ IID	4	200	6.1	7.1	6.6	5.6	6.6	5.4	5.1	6.2	4.5
			400	5.6	6.5	5.7	4.8	5.9	5.1	4.7	5.7	5.0
			100	7.5	9.6	9.8	5.9	7.2	6.0	3.8	5.0	3.4
$r_{iM} \sim$ IID,	$\boldsymbol{\varepsilon}_i \sim$ VAR(1)	4	200	5.7	6.9	6.9	5.1	6.6	5.2	4.3	5.2	3.8
			400	4.9	5.6	6.6	4.3	5.2	5.3	3.9	4.6	3.8
			100	15.6	18.3	16.5	12.7	15.0	13.0	9.4	10.7	9.1
$r_{iM} \sim$ AR(1),	$\boldsymbol{\varepsilon}_i \sim$ VAR(1)	4	200	11.1	13.3	12.5	9.7	11.1	10.0	6.6	7.8	6.7
			400	7.4	9.3	9.5	7.2	8.7	8.7	7.0	8.1	7.6
			100	13.8	16.3	14.9	11.5	13.1	11.9	8.8	10.2	7.9
$r_{iM} \sim$ AR(1),	$\boldsymbol{\varepsilon}_i \sim$ VAR(1)	10	200	9.4	11.1	10.5	7.9	9.1	8.6	5.7	6.9	5.5
			400	5.8	7.1	7.5	5.5	7.0	6.9	5.7	6.9	5.7
			100	12.1	14.0	12.4	8.9	10.7	8.7	6.4	7.7	6.1
$r_{iM} \sim$ AR(1),	$\boldsymbol{\varepsilon}_i \sim$ VAR(1)	20	200	7.2	8.6	8.5	6.6	7.6	6.2	5.1	6.7	5.6
			400	5.1	6.8	7.3	5.0	6.6	6.1	4.3	5.5	5.0

Table 2 shows the empirical sizes of the test procedure under the null hypothesis H_0 for $T = 10$. We simulated dependent data, where both the market portfolio r_{iM} as well as the errors $\boldsymbol{\varepsilon}_i$ may be dependent. In this case, the r_{iM}'s were generated as an autoregressive sequence of order 1 with coefficient 0.5 and i.i.d. $N(0, 1)$-innovations (denoted $r_{iM} \sim \mathrm{AR}(1)$), and the $\boldsymbol{\varepsilon}_i$'s were chosen as a vector autoregression with coefficient matrix $\mathbf{A} = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}$ and i.i.d. innovations $\boldsymbol{\zeta}_i \sim N_2(0, \mathbf{A})$, i.e., $\boldsymbol{\varepsilon}_i = \mathbf{A}\boldsymbol{\varepsilon}_{i-1} + \boldsymbol{\zeta}_i$ (denoted $\boldsymbol{\varepsilon}_i \sim \mathrm{VAR}(1)$). When no particular dependency is described, then the sequences are just i.i.d. innovations (IID) as specified before.

We can see that the empirical levels are getting closer to the nominal one when the size m of the historical sample increases. The L_1-procedure is the most conservative one, but all procedures perform quite well also for dependent data. Naturally, when the dependency is weaker, the performance is better. Also, in case of weak dependencies, the performance was relatively stable with respect to the choice of the bandwidth q in the Bartlett kernel variance estimator, and $q = 4$ turned out to be a reasonable choice. Only in case of heavier dependencies in both, market portfolio and errors, the performance improved with increasing q (confer the last three row blocks in Table 2). Concerning the choice of γ, there is some experience from other studies (cf., e.g., Černíková et al. [8] or Horváth et al. [15]). For example, if a change is to be expected "early" after the training period, then γ near to 0.5 is advisable. For "late change scenarios", however, small γ's are recommended. A choice of $\gamma = 0.25$ seems to strike a reasonably good balance between these two scenarios.

Next we concentrate on the robustness aspects of our method. For simplicity we used i.i.d. errors here with independent components, having either a normal N, t_4 or t_1 (Cauchy) distribution. Table 3 presents the corresponding empirical sizes (under H_0). It is obvious that, in case of t_1-distributed errors, the L_2-procedure totally fails, whereas the Huber- and L_1- type procedures still perform well.

In order to illustrate the properties of the test under the alternative hypothesis, we chose $T = 2$, $k^* = 10$ and $\delta_m \beta_j^1 = 1$, $j = 1, 2$. Table 4 gives the medians of detection delays $\tau_m - k^*$ under the same error distributions as above. Sometimes the change was not detected (ND) at all until the end of the monitoring period. Again, the Huber-type procedure shows a good performance over all error distributions. Note that, for the sake of brevity, we just reported on one scenario in Table 4, in which $d = 2$ and both parameters change, but the detector also reacts if at least one parameter changes.

Table 3
Empirical sizes at nominal level $\alpha = 5\%$, $T = 10$, $q = 4$, different error distributions.

	m\\γ	L_2			Huber			L_1		
		0	0.25	0.45	0	0.25	0.45	0	0.25	0.45
	100	7.4	8.1	6.9	5.9	6.7	5.3	3.8	4.9	3.6
N	200	4.7	5.9	5.0	4.4	5.3	4.3	4.3	5.0	3.9
	400	4.3	5.0	4.6	4.5	5.1	4.4	3.2	4.5	4.1
	100	6.2	8.5	9.3	5.7	7.6	6.5	5.2	6.7	6.5
t_4	200	6.5	8.5	9.5	5.0	6.7	4.4	4.5	5.4	6.4
	400	3.7	5.3	6.3	4.2	5.4	3.9	3.8	4.3	5.2
	100	62.6	65.6	64.2	5.5	6.5	4.2	4.2	5.1	5.6
t_1	200	65.9	68.5	67.9	5.3	6.0	4.6	3.9	4.0	5.6
	400	62.6	66.1	65.4	4.4	4.8	4.1	3.7	4.6	5.5

Table 4
Medians of detection delays, $k^* = 10$, $q = 4$.

	m\\γ	L_2			Huber			L_1		
		0	0.25	0.45	0	0.25	0.45	0	0.25	0.45
	100	30	20	15	41	28	22	62	45	39
N	200	38	22	15	50	31	22	73	47	35
	400	40	27	16	66	36	22	92	54	34
	100	38	25	20	41	27	22	63	44	38
t_4	200	48	29	20	50	31	22	73	47	34
	400	64	35	21	65	37	22	92	54	35
	100	ND	ND	186	71	50	42	90	66	60
t_1	200	ND	348	301	79	51	36	99	66	52
	400	ND	ND	766	100	59	36	120	73	48

Table 5
Empirical power of the test (in %) for t_1 errors, $q = 4$.

m\\γ	L_2			Huber			L_1		
	0	0.25	0.45	0	0.25	0.45	0	0.25	0.45
100	41.0	47.5	55.0	98	99	98	97	98	97
200	44.4	51.8	53.5	100	100	100	100	100	100
400	41.4	48.2	57.0	100	100	100	100	100	100

In Table 5, we present some empirical power values of the test procedure under t_1-distributed errors. For the other distributions the power was always 1 which is in accordance with (1.13).

Finally, as an illustration of a possible application, we investigated a data set of MSCI Global Sector Indices (net prices) that can serve as a benchmark to conduct relative valuations of sectors, industry groups and industries across countries and regions. Three sector indices – NDWUCSTA-World Consumer Staples (food, beverages, tobacco, prescription drugs and household products), NDWUFNCL-World Financials, and NDWUHC-World Health Care have been studied, and we chose NDDUWI MSCI World Index (a weighted index designed to measure the equity market performance of 24 developed country market indices) to represent the market portfolio.[1]

We have a sample of data from 29/12/2000 to 29/03/2011. The data from the period 31/12/2004 to 01/12/2006 (of length $m = 500$) was examined by an a-posteriori test for detecting a change in regression parameters based on cumulative sums (CUSUM's) of residuals (see, e.g., Csörgő and Horváth [10]). Since the (asymptotic) CUSUM test did not reject the null hypothesis of no change, we used this data set as the (stable) historical (training) period for our monitoring procedure. The length of the monitoring period is $2m = 1000$, that is, the monitoring terminates on 01/10/2010. So, critical values were chosen for $T = 2$.

Since, as described earlier, the Huber-type procedure provides a good combination of efficiency and robustness, we only present the results for this type of monitoring here. Also, as we did not know where to expect the possible change, we used $\gamma = 0.25$ as a compromise between detecting an early or late change, and, in accordance with the simulations above, we chose a bandwidth of $q = 4$ in the Bartlett kernel variance estimator. Moreover, we considered the portfolio including all three indices as well as all pairwise combinations.

1 Source: Bloomberg, 2011, http://www.msci.com/products/indices/tools/tickers/bb_eod/bloomberg_tickers_eod_sector.html

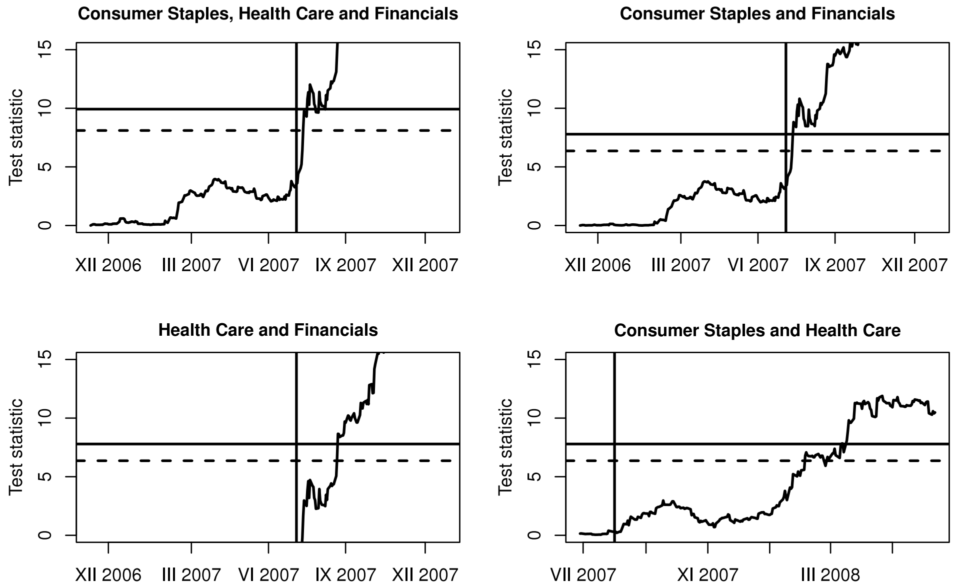

Fig. 1. Test statistics and critical values for different combinations of indices.

In Fig. 1 the values of the test statistics are shown together with the critical values; a solid line indicates the critical value for $T = \infty$ (i.e., for the open-end monitoring) and a dashed line the one for $T = 2$ (closed-end monitoring). A solid vertical line marks the date 01/08/2007, i.e., the date when the subprime mortgage crisis approximately started. The figures, particularly the first three ones, demonstrate the high sensitivity of the portfolio risk with respect to the financial sector.

4. Proofs

Proof of Theorem 2.1. The proof will be given in three steps. Let us recall that under the null hypothesis we work with the model

$$r_{i,j} = \alpha_j^0 + \beta_j^0 \widetilde{r}_{iM} + \varepsilon_{i,j}, \quad j = 1, \dots, d,\ i = 1, 2, \dots \tag{4.1}$$

as defined in (1.3).

1. In a first step we obtain asymptotic representations of the estimators $\widehat{\alpha}_{jm},\ \widehat{\beta}_{jm}$ of $\alpha_j^0,\ \beta_j^0,\ j = 1, \dots, d$. These estimators are based on the training sample only, hence we are in a non-sequential setup and can proceed in the same way as in treating the behavior of multivariate M-estimators. However, we need somewhat stronger results, since we are working under dependence.

It is convenient to introduce auxiliary estimators $\widehat{\alpha}_{jm}^*$ and $\widehat{\beta}_{jm}^*$ as minimizers of

$$\sum_{i=1}^{m} \varrho_j(\varepsilon_{i,j} - a_j^*/\sqrt{m} - b_j^* \widetilde{r}_{iM}/\sqrt{m}) \tag{4.2}$$

w.r.t. a_j^* and b_j^* for $j = 1, \dots, d$. Clearly,

$$\widehat{\alpha}_{jm}^* = \sqrt{m}(\widehat{\alpha}_{jm} - \alpha_j^0), \qquad \widehat{\beta}_{jm}^* = \sqrt{m}(\widehat{\beta}_{jm} - \beta_j^0). \tag{4.3}$$

Usually, the estimators $\widehat{\alpha}_{jm}^*$ and $\widehat{\beta}_{jm}^*$ can be obtained as solutions of the equations

$$\sum_{i=1}^{m} \psi_j(\varepsilon_{i,j} - (a_j^* + b_j^* \widetilde{r}_{iM})/\sqrt{m}) = 0, \tag{4.4}$$

$$\sum_{i=1}^{m} \psi_j(\varepsilon_{i,j} - (a_j^* + b_j^* \widetilde{r}_{iM})/\sqrt{m}) \widetilde{r}_{iM} = 0, \tag{4.5}$$

w.r.t. a_j^*, b_j^* for $j = 1, \dots, d$.

Proceeding in a standard way, Lemmas 5.2 and 5.3 below ensure that $\widehat{\alpha}^*_{jm} = O_P(1)$ and $\widehat{\beta}^*_{jm} = O_P(1)$ and, moreover, we get the asymptotic representations, as $m \to \infty$,

$$\widehat{\alpha}^*_{jm} = \sqrt{m}(\widehat{\alpha}_{jm} - \alpha_j^0) = \frac{1}{\sqrt{m}\lambda_j'(0)} \sum_{i=1}^{m} \psi_j(\varepsilon_{i,j}) + O_P(m^{-\eta}), \tag{4.6}$$

$$\widehat{\beta}^*_{jm} = \sqrt{m}(\widehat{\beta}_{jm} - \beta_j^0) = \frac{\sqrt{m}}{\lambda_j'(0)} \frac{1}{\sum_{i=1}^{m} \widetilde{r}^2_{iM}} \sum_{i=1}^{m} \psi_j(\varepsilon_{i,j})\widetilde{r}_{iM} + O_P(m^{-\eta}), \tag{4.7}$$

for some $\eta > 0$.

2. Next we show that the limit behavior of the weighted partial sums

$$\widehat{\mathbf{H}}(m,k) = (\widehat{H}_1(m,k), \dots, \widehat{H}_d(m,k))^T = \frac{1}{\sqrt{m}} \sum_{i=m+1}^{m+k} \widetilde{r}_{iM} \boldsymbol{\psi}(\widehat{\boldsymbol{\varepsilon}}_i), \quad k = 1, \dots, \lfloor mT \rfloor$$

is the same as that of

$$\mathbf{H}(m,k) = \frac{1}{\sqrt{m}} \left(\sum_{i=m+1}^{m+k} (r_{iM} - Er_{iM}) \boldsymbol{\psi}(\boldsymbol{\varepsilon}_i) - \frac{\sum_{i=m+1}^{m+k} (r_{iM} - Er_{iM})^2}{\sum_{i=1}^{m} (r_{iM} - Er_{iM})^2} \sum_{i=1}^{m} (r_{iM} - Er_{iM}) \boldsymbol{\psi}(\boldsymbol{\varepsilon}_i) \right), \quad k = 1, \dots, \lfloor mT \rfloor.$$

Again we follow the usual lines for treating the partial sums of weighted M-residuals with extension to dependent observations.

Lemmas 5.3 and 5.4 together with the asymptotic representations (4.6) and (4.7) imply that, as $m \to \infty$,

$$\max_{1 \le k < \lfloor mT \rfloor + 1} \left| \sum_{i=m+1}^{m+k} \widetilde{r}_{iM} \psi_j(\varepsilon_{i,j} - (\widehat{\alpha}^*_{jm} + \widehat{\beta}^*_{jm} \widetilde{r}_{iM})/\sqrt{m}) \right.$$
$$\left. - \left(\sum_{i=m+1}^{m+k} \widetilde{r}_{iM} \psi_j(\varepsilon_{i,j}) - \frac{\sum_{i=m+1}^{m+k} \widetilde{r}^2_{iM}}{\sum_{i=1}^{m} \widetilde{r}^2_{iM}} \sum_{i=1}^{m} \widetilde{r}_{iM} \psi_j(\varepsilon_{i,j}) \right) \right| \Big/ \left(\sqrt{m}(1 + k/m)(k/m)^{\gamma}\right) = O_P(m^{-\eta})$$

for some $\eta > 0$. Using the properties of $\{r_{iM}\}$ from Lemma 5.1 we conclude that this relation remains true even if $\widetilde{r}_{iM}$ is replaced by $r_{iM} - Er_{iM}$. This together with assumption (2.1) further implies that the limit behavior of

$$\max_{1 \le k \le \lfloor mT \rfloor} \widehat{Q}(k,m)/q_\gamma(k/m),$$

is the same as that of

$$\max_{1 \le k \le \lfloor mT \rfloor} Q(k,m)/q_\gamma(k/m),$$

where

$$Q(k,m) = \mathbf{H}(m,k)^T \boldsymbol{\Sigma}^{-1} \mathbf{H}(m,k). \tag{4.8}$$

3. Finally, we study the limit behavior of

$$\sum_{i=m+1}^{m+k} (r_{iM} - Er_{iM}) \boldsymbol{\psi}(\boldsymbol{\varepsilon}_i), \quad k = 1, \dots, \lfloor mT \rfloor,$$

and that of the related maximum of weighted quadratic forms

$$\max_{1 \le k \le \lfloor mT \rfloor} Q(k,m)/q_\gamma(k/m),$$

with $Q(k, m)$ defined by (4.8). The desired results are obtained by an application of the results in Billingsley [6]. More precisely, we introduce

$$\mathbf{Z}_i = (Z_{i,1}, \ldots, Z_{i,d})^T = (r_{iM} - Er_{iM})\boldsymbol{\psi}(\boldsymbol{\varepsilon}_i), \quad i = 1, 2, \ldots,$$
$$\mathbf{Z}_i^{(L)} = (Z_{i,1}^{(L)}, \ldots, Z_{i,d}^{(L)})^T = (r_{iM}^{(L)} - Er_{iM}^{(L)})\boldsymbol{\psi}(\boldsymbol{\varepsilon}_i^{(L)}), \quad i = 1, 2, \ldots, \quad \text{and}$$
$$\mathbf{Z}_m(t) = \frac{1}{\sqrt{m}} \sum_{i=1}^{\lfloor mt \rfloor} \mathbf{Z}_i, \quad 0 \le t \le T + 1.$$

The main step is to show that

$$\mathbf{Z}_m(\cdot) \overset{\mathcal{D}^d[0,T+1]}{\longrightarrow} \mathbf{W}_{\boldsymbol{\Sigma}}(\cdot),$$

where $\{\mathbf{W}_{\boldsymbol{\Sigma}}(t) : t \in [0, T+1]\}$ is a centered Gaussian process with covariance function $E[\mathbf{W}_{\boldsymbol{\Sigma}}(t)\mathbf{W}_{\boldsymbol{\Sigma}}^T(s)] = \min(t, s)\boldsymbol{\Sigma}$, and $\overset{\mathcal{D}^d[0,T+1]}{\longrightarrow}$ denotes weak convergence in the Skorokhod space $\mathcal{D}^d[0, T+1]$. To prove the latter relation, we make use of Billingsley [6], p. 184. Note that, in view of the triangle inequality and Assumption (B.3),

$$\begin{aligned}\|Z_{i,\ell} - Z_{i,\ell}^{(L)}\|_2 &= \|(r_{iM} - Er_{iM})\psi_\ell(\varepsilon_{i,\ell}) - (r_{iM}^{(L)} - Er_{iM}^{(L)})\psi_\ell(\varepsilon_{i,\ell}^{(L)})\|_2 \\ &\le \|[(r_{iM} - Er_{iM}) - (r_{iM}^{(L)} - Er_{iM}^{(L)})]\psi_\ell(\varepsilon_{i,\ell})\|_2 + \|(r_{iM}^{(L)} - Er_{iM}^{(L)})[\psi_\ell(\varepsilon_{i,\ell}) - \psi_\ell(\varepsilon_{i,\ell}^{(L)})]\|_2 \\ &\le 2\,\|r_{iM} - r_{iM}^{(L)}\|_2 \cdot \|\psi_\ell(\varepsilon_{0,\ell})\|_2 + \|r_{0M}^{(L)} - Er_{0M}^{(L)}\|_2 \cdot \|\psi_\ell(\varepsilon_{i,\ell}) - \psi_\ell(\varepsilon_{i,\ell}^{(L)})\|_2.\end{aligned}$$

So, by Assumptions (B.4) and (B.5), for $\ell = 1, \ldots, d$,

$$\sum_{L=1}^{\infty} \|Z_{i,\ell} - Z_{i,\ell}^{(L)}\|_2 < \infty.$$

Now, according to Davidson [11], Theorem 29.16, it suffices to show that, for any set of constants $c = (c_1, \ldots, c_d)^T$, we have

$$c^T\mathbf{Z}_m(\cdot) \overset{\mathcal{D}[0,T+1]}{\longrightarrow} \mathbf{c}^T W_{\boldsymbol{\Sigma}}(\cdot)$$

as $m \to \infty$. For the latter relation, however, note that

$$\sum_{L=1}^{\infty} \Big\| \sum_{\ell=1}^{d} Z_{i,\ell} - \sum_{\ell=1}^{d} Z_{i,\ell}^{(L)} \Big\|_2 < \infty,$$

so that the desired conclusion follows from Billingsley [6], Theorem 21.1. Here we make also use of the fact that

$$\lim_{m\to\infty} var \left\{ \frac{1}{\sqrt{m}} \sum_{i=1}^{m} (r_{iM} - Er_{iM})\boldsymbol{\psi}(\boldsymbol{\varepsilon}_i) \right\} = \boldsymbol{\Sigma}.$$

In the next step, we study the process

$$\widetilde{H}(m, \lfloor mt \rfloor) = \mathbf{Z}_m(t+1) - \mathbf{Z}_m(1) - t\mathbf{Z}_m(1) = \mathbf{Z}_m(t+1) - (t+1)\mathbf{Z}_m(1), \quad 0 \le t \le T.$$

Let $\boldsymbol{\Sigma} = \mathbf{C}\mathbf{C}^T$, where $\mathbf{C}$ is a regular matrix. Via the continuous mapping theorem,

$$\mathbf{C}^{-1}\widetilde{H}(m, \lfloor m\,\cdot \rfloor) \overset{\mathcal{D}^d[0,T]}{\longrightarrow} \widetilde{\mathbf{W}}(\cdot),$$

where $\{\widetilde{\mathbf{W}}(t) : 0 \le t \le T\}$ is a centered Gaussian process with covariance function $E[\widetilde{\mathbf{W}}(t)\widetilde{\mathbf{W}}^T(s)] = (t+1)s \cdot \mathbf{I}_d$, for $0 \le s \le t \le T$, with $\mathbf{I}_d$ denoting the d-dimensional unity matrix. Thus, via another application of the continuous mapping theorem,

$$\mathbf{C}^{-1}\widetilde{H}(m, \lfloor m\,\cdot \rfloor)/(\cdot + 1) \overset{\mathcal{D}^d[0,T]}{\longrightarrow} \widetilde{\mathbf{W}}(\cdot)/(\cdot + 1) = \mathbf{W}^*(\cdot),$$

for which it is easily checked that

$$\{\mathbf{W}^*(t) : 0 \le t \le T\} \overset{\mathcal{D}}{=} \left\{ \mathbf{W}\left(\frac{t}{t+1}\right) : 0 \le t \le T \right\},$$

with $\{\mathbf{W}(t) : t \ge 0\}$ denoting a standard Brownian motion.

To complete the proof, note that, in view of Lemma 5.1, as $m \to \infty$,

$$\sup_{0\le t\le T}\left|\frac{\sum_{i=m+1}^{m+\lfloor mt\rfloor}(r_{iM}-Er_{iM})^2}{\sum_{i=1}^{m}(r_{iM}-Er_{iM})^2}-t\right| \xrightarrow{P} 0,$$

which implies that also

$$\mathbf{C}^{-1}H(m,\lfloor m\cdot\rfloor)/(\cdot+1) \xrightarrow{\mathcal{D}^d[0,T]} \mathbf{W}\left(\frac{\cdot}{\cdot+1}\right).$$

Finally, in view of the law of iterated logarithm for a Brownian motion,

$$\mathbf{W}\left(\frac{t}{t+1}\right)\Big/\left(\frac{t}{t+1}\right)^{\delta} \to \mathbf{0} \quad P\text{-a.s.} \quad \text{as} \quad t\searrow 0,$$

for every $0 \le \delta < 1/2$, and

$$\sup_{1/m\le t\le T}\left|\frac{q_\gamma(t)}{q_\gamma(\lfloor mt\rfloor/m)}-1\right| = O\left(\frac{1}{m}\Big/\frac{1}{m^{2\gamma}}\right) \to 0 \qquad \text{as } m\to\infty,$$

since $0 \le \gamma < 1/2$.

Now, an application of the continuous mapping theorem, restricted to the set of functions

$$\left\{\mathbf{x}\in\mathcal{D}^d[0,T] : \lim_{t\searrow 0}|\mathbf{x}(t)|/t^{\delta} = 0\,\forall\, 0\le\delta<1/2\right\},$$

completes this step and thus the proof of Theorem 2.1. □

Proof of Theorem 2.2. Recall that the model under the considered alternative has the form

$$r_{i,j} = \alpha_j^0 + \beta_j^0\widetilde{r}_{iM} + (\alpha_j^1 + \beta_j^1\widetilde{r}_{iM})\delta_m I\{i > m+k^*\} + \varepsilon_{i,j}, \quad j=1,\dots,d,\ i=1,2,\dots,$$

where $\delta_m \to 0$. Notice that in this situation we have, for $k^* < k \le \lfloor mT\rfloor$,

$$\sum_{i=m+1}^{m+k}\widetilde{r}_{iM}\psi_j(\widehat{\varepsilon}_{i,j}) = \sum_{i=m+1}^{m+k}\widetilde{r}_{iM}\psi_j(\varepsilon_{i,j} - (\widehat{\alpha}_{jm}^* + \widehat{\beta}_{jm}^*\widetilde{r}_{iM})/\sqrt{m} + (\alpha_j^1 + \beta_j^1\widetilde{r}_{iM})\delta_m I\{i > m+k^*\}) \tag{4.9}$$

with

$$\widehat{\alpha}_{jm}^* = O_P(1), \qquad \widehat{\beta}_{jm}^* = O_P(1) \quad (m\to\infty),$$

based on the training sample only. Moreover, in view of (5.4) below, it is enough to work on the set

$$\left\{\sup_{|a|+|b|\le C}\max_{i=1,\dots,m}|a+b\widetilde{r}_{iM}|/\sqrt{m}\le a_0\right\},$$

with an arbitrary $C > 0$ and a_0 from Assumption (B.5).

(i) Similar to the proof of Theorem 2.1 we need to study

$$L_j(a,b,m,k) = \frac{1}{\sqrt{m}}\sum_{i=m+1}^{m+k}\widetilde{r}_{iM}\psi_j(\varepsilon_{i,j} - (a+b\widetilde{r}_{iM})/\sqrt{m} + (\alpha_j^1+\beta_j^1\widetilde{r}_{iM})\delta_m I\{i>m+k^*\}).$$

Along the lines of the proof of Lemma 5.4 we get that

$$\max_{1\le k\le\lfloor mT\rfloor}\left(\frac{|\{L_j(a,b,m,k) - E^*L_j(a,b,m,k)\}_{a=\widehat{\alpha}_{jm}^*,b=\widehat{\beta}_{jm}^*}|}{(k/m)^{\gamma}}\right) = O_P(m^{-\eta}) \quad (m\to\infty),$$

uniformly in $|a|+|b|\le C$ for some $\eta > 0$, where E^* denotes the conditional expectation given $(r_{1M},\dots,r_{mM})$.

The conditional expectation of $L_j(a,b,m,k)$ has to be calculated more carefully. For doing so, notice that, using the notation $d_i = a + b\widetilde{r}_{iM}$ (see also (5.3) below),

$$E^*L_j(a,b,m,k) = \frac{1}{\sqrt{m}}\sum_{i=m+1}^{m+k}\widetilde{r}_{iM}E^*\psi_j(\varepsilon_{i,j} - d_i/\sqrt{m} + (\alpha_j^1+\beta_j^1\widetilde{r}_{iM})\delta_m I\{i>m+k^*\})$$

$$= -\frac{1}{\sqrt{m}} \sum_{i=m+1}^{m+k} \widetilde{r}_{iM} \Big(\lambda_j'(0)(d_i/\sqrt{m} - (\alpha_j^1 + \beta_j^1 \widetilde{r}_{iM})\delta_m I\{i > m + k^*\})$$
$$+ O_P\big(((a + b\widetilde{r}_{iM})/\sqrt{m} + (\alpha_j^1 + \beta_j^1 \widetilde{r}_{iM})\delta_m I\{i > m + k^*\})^2\big)\Big) \quad (m \to \infty),$$

uniformly in $|a| + |b| \le C$ and in $1 \le k \le \lfloor mT \rfloor$.

Then, in case of $\delta_m = m^{-1/2}$, an application of Lemma 5.1 results in

$$\begin{aligned} E^* L_j(a, b, m, k) &= -b\lambda_j'(0)\frac{1}{m}\sum_{i=m+1}^{m+k} \widetilde{r}_{iM}^2 + \beta_j^1 \lambda_j'(0)\frac{1}{m}\sum_{i=m+k^*+1}^{m+k} \widetilde{r}_{iM}^2\, I\{k > k^*\} \\ &\quad - a\lambda_j'(0)\frac{1}{m}\sum_{i=m+1}^{m+k} \widetilde{r}_{iM} + \alpha_j^1 \lambda_j'(0)\frac{1}{m}\sum_{i=m+k^*+1}^{m+k} \widetilde{r}_{iM}\, I\{k > k^*\} \\ &\quad + O_P\big(\{a^2 + b^2 + (\alpha_j^1)^2 + (\beta_j^1)^2\} m^{-\xi}\big) \quad (m \to \infty), \end{aligned}$$

uniformly in $|a| + |b| \le C$ and in $1 \le k \le \lfloor mT \rfloor$, for some $\xi > 0$.

Now, since $\widehat{\alpha}_{jm}^* = O_P(1)$ and $\widehat{\beta}_{jm}^* = O_P(1)$, we can plug-in the estimates $\widehat{\alpha}_{jm}^*$ and $\widehat{\beta}_{jm}^*$ for a and b, respectively, and after a few standard steps we get that the limit behavior of

$$\left\{ \frac{L_j(\widehat{\alpha}_{jm}^*, \widehat{\beta}_{jm}^*, m, \lfloor mt \rfloor)}{\big(q_\gamma(\lfloor mt \rfloor/m)\big)^{1/2}} : t \in \left[\frac{1}{m}, T\right] \right\}$$

is the same as that of

$$\left\{ \left(\frac{1}{\sqrt{m}} \sum_{i=m+1}^{\lfloor mt \rfloor} (r_{iM} - Er_{iM})\psi_j(\varepsilon_{i,j}) - \frac{\sum_{i=m+1}^{\lfloor mt \rfloor} (r_{iM} - Er_{iM})^2}{\sum_{i=1}^{m} (r_{iM} - Er_{iM})^2} \sum_{i=1}^{m} (r_{iM} - Er_{iM})\psi_j(\varepsilon_{i,j}) \right.\right.$$
$$\left.\left. + \lambda_j'(0)\beta_j^1 \frac{1}{m} \sum_{i=m+k^*+1}^{m+\lfloor mt \rfloor} (r_{iM} - Er_{iM})^2 \right) \Big/ \big(q_\gamma(\lfloor mt \rfloor/m)\big)^{1/2} : t \in \left[\frac{1}{m}, T\right] \right\}$$

for $0 \le \gamma < 1/2$ and $\lim_{m\to\infty}(\lfloor mT \rfloor - k^*)/m = T - s > 0$. Then the rest of the proof of (i) follows along the lines of that of Theorem 2.1.

(ii) In case of $\sqrt{m}\,|\delta_m| \to \infty$ and $\liminf_{m\to\infty}(\lfloor mT \rfloor - k^*)/m > 0$, the term with β_j^1 in $E^* L_j(a, b, m, mT)$ dominates. More precisely,

$$|E^* L_j(a, b, m, mT)| = \sqrt{m}\,|\delta_m|\,|\lambda_j'(0)|\,|\beta_j^1|\,\frac{1}{m} \sum_{i=m+k^*+1}^{m+\lfloor mT \rfloor} \widetilde{r}_{iM}^2 \big(1 + o_P(1)\big) \xrightarrow{P} \infty,$$

uniformly in $|a| + |b| \le C$ and in $1 \le k \le \lfloor mT \rfloor$, if $\beta_j^1 \ne 0$. Therefore also the test statistic converges to ∞ in probability, which proves (ii) and completes the proof of Theorem 2.2. □

Proof of Theorem 2.3. First define, for $k \ge 0$,

$$\widetilde{\boldsymbol{\Gamma}}_k = \frac{1}{m} \sum_{i=1}^{m-k} \widetilde{r}_{iM} \widetilde{r}_{i+k,M} \boldsymbol{\psi}(\boldsymbol{\varepsilon}_i) \boldsymbol{\psi}(\boldsymbol{\varepsilon}_{i+k})^T,$$

$$\overline{\boldsymbol{\Gamma}}_k = \frac{1}{m} \sum_{i=1}^{m-k} (r_{iM} - Er_{iM})(r_{i+k,M} - Er_{i+k,M}) \boldsymbol{\psi}(\boldsymbol{\varepsilon}_i) \boldsymbol{\psi}(\boldsymbol{\varepsilon}_{i+k})^T,$$

and, for $k < 0$, put $\widetilde{\boldsymbol{\Gamma}}_k = \widetilde{\boldsymbol{\Gamma}}_{-k}^T$ and $\overline{\boldsymbol{\Gamma}}_k = \overline{\boldsymbol{\Gamma}}_{-k}^T$, respectively.

Next, let

$$\widetilde{\boldsymbol{\Sigma}}_m = \sum_{|k|<q} \omega_q(k) \widetilde{\boldsymbol{\Gamma}}_k$$

and

$$\overline{\boldsymbol{\Sigma}}_m = \sum_{|k|<q} \omega_q(k) \overline{\boldsymbol{\Gamma}}_k.$$

Then we have

$$\widehat{\boldsymbol{\Sigma}}_m = (\widehat{\boldsymbol{\Sigma}}_m - \widetilde{\boldsymbol{\Sigma}}_m) + (\widetilde{\boldsymbol{\Sigma}}_m - \overline{\boldsymbol{\Sigma}}_m) + \overline{\boldsymbol{\Sigma}}_m.$$

First, let us consider

$$\widetilde{\boldsymbol{\Sigma}}_m - \overline{\boldsymbol{\Sigma}}_m = \sum_{|k|<q} \omega_q(k)(\widetilde{\boldsymbol{\Gamma}}_k - \overline{\boldsymbol{\Gamma}}_k).$$

Recall that $\{r_{iM}\}$ is stationary and put $Er_{0M} = \mu$. For simplicity and due to symmetry, we will only treat the terms with $k \geq 0$. Obviously,

$$\begin{aligned}\widetilde{\boldsymbol{\Gamma}}_k - \overline{\boldsymbol{\Gamma}}_k = {} & (\overline{r}_{mM} - \mu)^2 \frac{1}{m} \sum_{i=1}^{m-k} \boldsymbol{\psi}(\boldsymbol{\varepsilon}_i)\boldsymbol{\psi}(\boldsymbol{\varepsilon}_{i+k})^T - (\overline{r}_{mM} - \mu)\frac{1}{m}\sum_{i=1}^{m-k}(r_{iM} - \mu)\boldsymbol{\psi}(\boldsymbol{\varepsilon}_i)\boldsymbol{\psi}(\boldsymbol{\varepsilon}_{i+k})^T \\ & - (\overline{r}_{mM} - \mu)\frac{1}{m}\sum_{i=1}^{m-k}(r_{i+k,M} - \mu)\boldsymbol{\psi}(\boldsymbol{\varepsilon}_i)\boldsymbol{\psi}(\boldsymbol{\varepsilon}_{i+k})^T.\end{aligned}$$

According to Lemma 5.1, $\overline{r}_{mM} - \mu = O_P(m^{-1/2})$ as $m \to \infty$. Further, for any $j, \ell = 1, \dots, d$ and $\epsilon > 0$,

$$\begin{aligned} P\left(\left|\sum_{0\leq k<q} \omega_q(k)\frac{1}{m}\sum_{i=1}^{m-k}\psi_j(\varepsilon_{i,j})\psi_\ell(\varepsilon_{i+k,\ell})\right| > \epsilon\right) & \leq \frac{1}{\epsilon}\sum_{0\leq k<q}\omega_q(k)\,\frac{1}{m}\,E\left|\sum_{i=1}^{m-k}\psi_j(\varepsilon_{i,j})\psi_\ell(\varepsilon_{i+k,\ell})\right| \\ & \leq \frac{1}{\epsilon}\sum_{0\leq k<q}\omega_q(k)\,\frac{1}{m}\,\sum_{i=1}^{m-k}(E|\psi_j(\varepsilon_{i,j})|^2)^{1/2}(E|\psi_\ell(\varepsilon_{i+k,\ell})|^2)^{1/2} \leq \frac{D}{\epsilon}\,q,\end{aligned}$$

thus

$$\sum_{0\leq k<q}\omega_q(k)\frac{1}{m}\sum_{i=1}^{m-k}\boldsymbol{\psi}(\boldsymbol{\varepsilon}_i)\boldsymbol{\psi}(\boldsymbol{\varepsilon}_{i+k})^T = O_P\big(q(m)\big) \quad (m\to\infty).$$

Similarly, by iterated expectations and according to the independence of $\{r_{iM}\}$ and $\{\boldsymbol{\psi}(\varepsilon_i)\}$,

$$\begin{aligned} & P\left(\left|\sum_{0\leq k<q}\omega_q(k)\frac{1}{m}\sum_{i=1}^{m-k}(r_{iM}-\mu)\psi_j(\varepsilon_{i,j})\psi_\ell(\varepsilon_{i+k,\ell})\right| > \epsilon\right) \\ & \quad \leq \frac{1}{\epsilon}\sum_{0\leq k<q}\omega_q(k)\,\frac{1}{m}\,\sum_{i=1}^{m-k}E|r_{iM}-\mu|(E|\psi_j(\varepsilon_{i,j})|^2)^{1/2}(E|\psi_\ell(\varepsilon_{i+k,\ell})|^2)^{1/2} \leq \frac{D}{\epsilon}\,q,\end{aligned}$$

which implies

$$\sum_{0\leq k<q}\omega_q(k)\frac{1}{m}\sum_{i=1}^{m-k}(r_{iM}-\mu)\boldsymbol{\psi}(\boldsymbol{\varepsilon}_i)\boldsymbol{\psi}(\boldsymbol{\varepsilon}_{i+k})^T = O_P\big(q(m)\big) \quad (m\to\infty).$$

Analogously

$$\sum_{0\leq k<q}\omega_q(k)\frac{1}{m}\sum_{i=1}^{m-k}(r_{i+k,M}-\mu)\boldsymbol{\psi}(\boldsymbol{\varepsilon}_i)\boldsymbol{\psi}(\boldsymbol{\varepsilon}_{i+k})^T = O_P\big(q(m)\big) \quad (m\to\infty),$$

from which, together with the corresponding estimates for $\sum_{q<k<0}$, we conclude that

$$\widetilde{\boldsymbol{\Sigma}}_m - \overline{\boldsymbol{\Sigma}}_m = O_P\big(q(m)m^{-1/2}\big) \quad (m\to\infty).$$

Next, let us consider

$$\widehat{\boldsymbol{\Sigma}}_m - \widetilde{\boldsymbol{\Sigma}}_m = \sum_{|k|<q}\omega_q(k)(\widehat{\boldsymbol{\Gamma}}_k - \widetilde{\boldsymbol{\Gamma}}_k).$$

We have

$$\begin{aligned}\sum_{0\le k<q}\omega_q(k)(\widehat{\boldsymbol{\Gamma}}_k-\widetilde{\boldsymbol{\Gamma}}_k) &= \sum_{0\le k<q}\omega_q(k)\frac{1}{m}\sum_{i=1}^{m-k}\widetilde{r}_{iM}\widetilde{r}_{i+k,M}[\boldsymbol{\psi}(\widehat{\boldsymbol{\varepsilon}}_i)\boldsymbol{\psi}(\widehat{\boldsymbol{\varepsilon}}_{i+k})^T-\boldsymbol{\psi}(\boldsymbol{\varepsilon}_i)\boldsymbol{\psi}(\boldsymbol{\varepsilon}_{i+k})^T] \\ &= \sum_{0\le k<q}\omega_q(k)\frac{1}{m}\sum_{i=1}^{m-k}\widetilde{r}_{iM}\widetilde{r}_{i+k,M}[\boldsymbol{\psi}(\widehat{\boldsymbol{\varepsilon}}_i)-\boldsymbol{\psi}(\boldsymbol{\varepsilon}_i)][\boldsymbol{\psi}(\widehat{\boldsymbol{\varepsilon}}_{i+k})-\boldsymbol{\psi}(\boldsymbol{\varepsilon}_{i+k})]^T \\ &\quad+\sum_{0\le k<q}\omega_q(k)\frac{1}{m}\sum_{i=1}^{m-k}\widetilde{r}_{iM}\widetilde{r}_{i+k,M}[\boldsymbol{\psi}(\widehat{\boldsymbol{\varepsilon}}_i)-\boldsymbol{\psi}(\boldsymbol{\varepsilon}_i)]\boldsymbol{\psi}(\boldsymbol{\varepsilon}_{i+k})^T \\ &\quad+\sum_{0\le k<q}\omega_q(k)\frac{1}{m}\sum_{i=1}^{m-k}\widetilde{r}_{iM}\widetilde{r}_{i+k,M}\boldsymbol{\psi}(\boldsymbol{\varepsilon}_i)[\boldsymbol{\psi}(\widehat{\boldsymbol{\varepsilon}}_{i+k})-\boldsymbol{\psi}(\boldsymbol{\varepsilon}_{i+k})]^T \\ &= \mathbf{S}_1+\mathbf{S}_2+\mathbf{S}_3 \quad \text{(say).}\end{aligned}$$

Set $\mathbf{d}_i=(\mathbf{a}+\widetilde{r}_{iM}\mathbf{b})/\sqrt{m}$, where $\mathbf{a}=(a_1,\dots,a_d)^T,\ \mathbf{b}=(b_1,\dots,b_d)^T,\ \mathbf{d}_i=(d_{i,1},\dots,d_{i,d})^T$, and define

$$\mathbf{S}_1^0=\sum_{0\le k<q}\omega_q(k)\frac{1}{m}\sum_{i=1}^{m-k}\widetilde{r}_{iM}\widetilde{r}_{i+k,M}[\boldsymbol{\psi}(\boldsymbol{\varepsilon}_i-\mathbf{d}_i/\sqrt{m})-\boldsymbol{\psi}(\boldsymbol{\varepsilon}_i)][\boldsymbol{\psi}(\boldsymbol{\varepsilon}_{i+k}-\mathbf{d}_{i+k}/\sqrt{m})-\boldsymbol{\psi}(\boldsymbol{\varepsilon}_{i+k})]^T,$$

$$\mathbf{S}_2^0=\sum_{0\le k<q}\omega_q(k)\frac{1}{m}\sum_{i=1}^{m-k}\widetilde{r}_{iM}\widetilde{r}_{i+k,M}[\boldsymbol{\psi}(\boldsymbol{\varepsilon}_i-\mathbf{d}_i/\sqrt{m})-\boldsymbol{\psi}(\boldsymbol{\varepsilon}_i)]\boldsymbol{\psi}(\boldsymbol{\varepsilon}_{i+k})^T,$$

$$\mathbf{S}_3^0=\sum_{0\le k<q}\omega_q(k)\frac{1}{m}\sum_{i=1}^{m-k}\widetilde{r}_{iM}\widetilde{r}_{i+k,M}\boldsymbol{\psi}(\boldsymbol{\varepsilon}_i)[\boldsymbol{\psi}(\boldsymbol{\varepsilon}_{i+k}-\mathbf{d}_{i+k}/\sqrt{m})-\boldsymbol{\psi}(\boldsymbol{\varepsilon}_{i+k})]^T.$$

Now, for the (j,ℓ)-th component $S_1^0(j,\ell)$ of $\mathbf{S}_1^0$, we get

$$\begin{aligned}&P\big(|S_1^0(j,\ell)|>\epsilon\big) \\ &\quad\le\sum_{0\le k<q}\omega_q(k)\frac{1}{\epsilon\, m}\sum_{i=1}^{m-k}E(E^*|\widetilde{r}_{iM}\widetilde{r}_{i+k,M}[\psi_j(\varepsilon_{i,j}-d_{i,j}/\sqrt{m})-\psi_j(\varepsilon_{i,j})][\psi_\ell(\varepsilon_{i+k,\ell}-d_{i+k,\ell}/\sqrt{m})-\psi_\ell(\varepsilon_{i+k,\ell})]|) \\ &\quad\le\sum_{0\le k<q}\omega_q(k)\frac{1}{\epsilon\, m}\sum_{i=1}^{m-k}E(|\widetilde{r}_{iM}\widetilde{r}_{i+k,M}|E^*|\psi_j(\varepsilon_{i,j}-d_{i,j}/\sqrt{m})-\psi_j(\varepsilon_{i,j})||\psi_\ell(\varepsilon_{i+k,\ell}-d_{i+k,\ell}/\sqrt{m})-\psi_\ell(\varepsilon_{i+k,\ell})|),\end{aligned}$$

with E^* the conditional expectation given r_{iM}, $i=1,\dots,m$.

From the Cauchy–Schwarz inequality and according to Assumption (A.2),

$$\begin{aligned}&E^*|\psi_j(\varepsilon_{i,j}-d_{i,j}/\sqrt{m})-\psi_j(\varepsilon_{i,j})||\psi_\ell(\varepsilon_{i+k,\ell}-d_{i+k,\ell}/\sqrt{m})-\psi_\ell(\varepsilon_{i+k,\ell})|\le D|d_{i,j}/\sqrt{m}|^{\kappa/2}|d_{i+k,\ell}/\sqrt{m}|^{\kappa/2} \\ &\quad=Dm^{-\kappa/2}|a_j+b_j\widetilde{r}_{iM}|^{\kappa/2}|a_\ell+b_\ell\widetilde{r}_{i+k,M}|^{\kappa/2},\end{aligned}$$

which holds uniformly in $k=0,\dots,q$ and for all $\mathbf{a}$, $\mathbf{b}$ such that $\max_{1\le j\le d}(|a_j|+|b_j|)<C$ with some constant $C>0$.

From here, using the Cauchy–Schwarz inequality once again, and in view of Assumption (B.1) with $\Delta\ge\kappa$,

$$\begin{aligned}&E(|\widetilde{r}_{iM}\|\widetilde{r}_{i+k,M}\|a_j+b_j\widetilde{r}_{iM}|^{\kappa/2}|a_\ell+b_\ell\widetilde{r}_{i+k,M}|^{\kappa/2}m^{-\kappa/2}) \\ &\quad\le Dm^{-\kappa/2}[E|\widetilde{r}_{iM}|^{1+\kappa/2}|\widetilde{r}_{i+k,M}|^{1+\kappa/2}+E|\widetilde{r}_{iM}\widetilde{r}_{i+k,M}|+E|\widetilde{r}_{iM}|^{1+\kappa/2}|\widetilde{r}_{i+k,M}|+E|\widetilde{r}_{iM}\widetilde{r}_{i+k,M}|^{1+\kappa/2}] \\ &\quad\le Dm^{-\kappa/2},\end{aligned}$$

from which we conclude that $S_1^0(j,\ell)=O_P\big(q(m)m^{-\kappa/2}\big)$ and thus $\mathbf{S}_1^0=O_P\big(q(m)m^{-\kappa/2}\big)$ as $m\to\infty$, uniformly for all $\mathbf{a}$, $\mathbf{b}$ such that $\max_{1\le j\le d}(|a_j|+|b_j|)<C$ with some constant $C>0$.

Proceeding in the same way we get $\mathbf{S}_2^0=O_P\big(q(m)m^{-\kappa/4}\big)$ and $\mathbf{S}_3^0=O_P\big(q(m)m^{-\kappa/4}\big)$ as $m\to\infty$, uniformly for all $\mathbf{a}$, $\mathbf{b}$ such that $\max_{1\le j\le d}(|a_j|+|b_j|)<C$ with some $C>0$, and we can conclude that

$$\mathbf{S}_1^0+\mathbf{S}_2^0+\mathbf{S}_3^0=O_P\big(q(m)m^{-\kappa/4}\big)\quad(m\to\infty),$$

uniformly for all $\mathbf{a}$, $\mathbf{b}$ such that $\max_{1\le j\le d}(|a_j|+|b_j|)<C$ with some constant $C>0$.

Since $\widehat{\varepsilon}_{i,j}=\varepsilon_{i,j}-\widehat{\alpha}^*_{jm}/\sqrt{m}-\widehat{\beta}^*_{jm}\widetilde{r}_{iM}/\sqrt{m}$ and $\widehat{\alpha}^*_{jm}=O_P(1)$, $\widehat{\beta}^*_{jm}=O_P(1)$ for all $j=1,\dots,d$ (see (4.6) and (4.7), respectively), we obtain, due to the monotonicity of ψ_j, $j=1,\dots,d$, and with the corresponding estimates for $-q<k<0$, that

$$\widehat{\boldsymbol{\Sigma}}_m-\widetilde{\boldsymbol{\Sigma}}_m=O_P\big(q(m)m^{-\kappa/4}\big)\quad(m\to\infty).$$

It remains to show that $\overline{\boldsymbol{\Sigma}}_m \xrightarrow{P} \boldsymbol{\Sigma}$ as $m \to \infty$. But this follows easily from Theorem 16.6 in Horváth and Kokoszka [16]. Indeed, if we denote $\mathbf{Z}_i = (r_{iM} - Er_{iM})\boldsymbol{\psi}(\boldsymbol{\varepsilon}_i),\ i \in \mathbb{Z}$, it is enough to verify that $\{\mathbf{Z}_i\}$ satisfies Assumptions 16.2–16.6 in Horváth and Kokoszka [16].

The latter assumptions, however, follow easily from our Assumptions (B.1)–(B.5) together with the independence of $\{r_{iM}\}$ and $\{\boldsymbol{\psi}(\boldsymbol{\varepsilon}_i)\}$ and Lemma 2.1 in Hörmann and Kokoszka [14]. Assumption 16.5 is satisfied with the kernel (2.6) (see Example 16.4 in Horváth and Kokoszka [16]), and Assumption 16.6 is fulfilled if $m \to \infty$, $q(m) \to \infty$ and $q(m)m^{-\min\{1/2,\kappa/4\}} \to 0$. Then with

$$\overline{\boldsymbol{\Gamma}}_k = \frac{1}{m}\sum_{i=1}^{m-k} \mathbf{Z}_i \mathbf{Z}_{i+k}^T, \quad k \geq 0,$$

and $\overline{\boldsymbol{\Gamma}}_k = \overline{\boldsymbol{\Gamma}}_{-k}^T,\ k < 0$,

$$\overline{\boldsymbol{\Sigma}}_m = \sum_{|k|<q} \omega_q(k)\overline{\boldsymbol{\Gamma}}_k \xrightarrow{P} \boldsymbol{\Sigma} \quad (m \to \infty),$$

which completes the proof. □

Remark 4.1. The assertion of Theorem 2.3 can be proved with any kernel ω_q satisfying Assumption 16.5 in Horváth and Kokoszka [16].

5. Some auxiliary results

In the sequel, $D > 0$ is a generic constant, which may vary from case to case.

At first we gather some properties of the sequence $\{r_{iM}\}$.

Lemma 5.1. *Let Assumptions* (B.1) *and* (B.4) *be satisfied. Then,*

(i) *the sequence* $\{r_{iM}^2\}$ *satisfies Assumption* (B.1) *with* $2 + \Delta$ *replaced by* $1 + \Delta/2$ *and Assumption* (B.4) *with* $\|\cdot\|_{2+\Delta}$ *replaced by* $\|\cdot\|_{1+\Delta/2}$;

(ii) *there is a constant* $C > 0$ *such that, for every* $\ell \in \mathbb{Z}$ *and* $n \in \mathbb{N}$,

$$E\left|\sum_{i=\ell+1}^{\ell+n}(r_{iM} - Er_{iM})\right|^p \leq Cn^{p/2}, \quad 0 < p \leq 2 + \Delta,$$

and, for $b_1 \geq b_2 \geq \cdots \geq b_n > 0$,

$$E\max_{1\leq k\leq n}\left|b_k\sum_{i=\ell+1}^{\ell+n}(r_{iM} - Er_{iM})\right|^p \leq Cn^{p/2-1}\sum_{k=1}^{n} b_k^p, \quad 2 < p \leq 2 + \Delta; \tag{5.1}$$

(iii) *as* $m \to \infty$,

$$\sum_{i=1}^{m}(r_{iM} - Er_{iM}) = O_P\big(m^{1/2}\big),$$

$$\max_{1\leq i\leq m}|r_{iM} - Er_{iM}| = O_P\big(m^{1/(2+\Delta)}\big),$$

$$\sum_{i=1}^{m}|r_{iM} - Er_{iM}|^a = O_P\big(m^{\max(1,a/(2+\Delta))}\big)$$

for $a > 0$;

(iv) *for some* $D > 0$,

$$E\left(\max_{1\leq k\leq \lfloor mT\rfloor}\frac{\left|\sum_{i=m+1}^{m+k}(r_{iM} - Er_{iM})\right|}{\sqrt{m}\,(k/m)^{\gamma}}\right)^p \leq D, \quad 2 < p \leq 2 + \Delta;$$

(v) *as* $m \to \infty$,

$$\sup_{0\leq t\leq T}\left|\frac{\sum_{i=m+1}^{m+\lfloor mt\rfloor}(r_{iM} - Er_{iM})^2}{\sum_{i=1}^{m}(r_{iM} - Er_{iM})^2} - t\right| \xrightarrow{P} 0.$$

Proof. (i) Ergodicity and stationarity, and the assertion for (B.1) is obvious. The rest follows from the following inequality:

$$\begin{aligned}\left(E|{r_{0M}}^2-(r_{0M}^2)^{(L)}|^{p/2}\right)^{2/p} &= \left(E(|\,r_{0M}-{r_{0M}}^{(L)}\,|^{p/2}|\,r_{0M}+{r_{0M}}^{(L)})|^{p/2}\right)^{2/p}\\ &\le \left(E|\,r_{0M}-{r_{0M}}^{(L)}\,|^p\cdot E|\,r_{0M}+{r_{0M}}^{(L)}\,|^p\right)^{1/p}\le D\left(E|\,r_{0M}-{r_{0M}}^{(L)}\,|^p\right)^{1/p}.\end{aligned}$$

(ii) Making use of the $L_{2+\Delta}$-approximability from Assumption (B.4), the first bound has been obtained in Berkes et al. [5], Proposition 4. For the second one confer, e.g., Kirch [20], Theorem B.1.

(iii) By Chebyshev's inequality and assertion (ii) above, with $p=2$,

$$P\left(\left|\sum_{i=1}^m (r_{iM}-Er_{iM})\right|\ge\lambda\right)\le\frac{D}{\lambda^2}m.$$

Next, note that

$$\max_{1\le i\le m}|r_{iM}-Er_{iM}|\le D\left(\frac{1}{m}\sum_{i=1}^m|r_{iM}-Er_{iM}|^{2+\delta}\right)^{1/(2+\delta)} m^{1/(2+\delta)}$$

for any $\delta\ge 0$. Since the sequence $\{r_{iM}\}$ is stationary and ergodic, also $\{g(r_{iM})\}$ is stationary and ergodic, where g is a measurable function, and, if $E|g(r_{iM})|<\infty$, the ergodic theorem implies

$$\frac{1}{m}\sum_{i=1}^m g(r_{iM})\to Eg(r_{iM})\quad\text{a.s.}\quad(m\to\infty). \tag{5.2}$$

Hence, under Assumption (B.1),

$$\frac{1}{m}\sum_{i=1}^m|r_{iM}-Er_{iM}|^{2+\Delta}\to E|r_{0M}-r_{0M}|^{2+\Delta}\quad\text{a.s.}\quad(m\to\infty),$$

and therefore

$$\max_{1\le i\le m}|r_{iM}-Er_{iM}|=O_P\left(m^{1/(2+\Delta)}\right)\quad(m\to\infty),$$

which easily implies

$$\begin{aligned}\sum_{i=1}^m|r_{iM}-Er_{iM}|^a &= O_P\left(\sum_{i=1}^m|r_{iM}-Er_{iM}|^{\min(a,2+\Delta)}\max_{1\le i\le m}|r_{iM}-Er_{iM}|^{\max(0,a-(2+\Delta))}\right)\\ &= O_P\left(m^{\max(1,a/(2+\Delta))}\right)\quad(m\to\infty).\end{aligned}$$

(iv) It follows immediately from Assumptions (B.1) and (B.4) together with (5.1).

(v) Note that, by (5.2),

$$\frac{1}{m}\sum_{i=1}^m(r_{iM}-Er_{iM})^2\ \to\ var(r_{0M})\quad\text{a.s.}\quad(m\to\infty),$$

hence, due to the strict stationarity, also

$$\sup_{0\le t\le\lfloor mT\rfloor}\left|\frac{1}{m}\left\{\sum_{i=m+1}^{m+\lfloor mt\rfloor}(r_{iM}-Er_{iM})^2-\lfloor mt\rfloor var(r_{0M})\right\}\right|\ \xrightarrow{P}\ 0\quad(m\to\infty).$$

On combining the above two assertions, the proof of (v) can be completed. □

The following two lemmas are crucial assertions for the proof of the limit behavior of the estimators $\widehat{\alpha}^*_{jm}$, $\widehat{\beta}^*_{jm}$ from (4.2) and (4.3), respectively. The arguments in the proofs of both lemmas are standard but, since our assumptions are different from those usually considered, also some parts of the proofs differ and we focus on these differences.

In the following P^*, E^* and var^* again denote the conditional probability, conditional expectation and conditional variance given r_{iM}, $i=1,\dots,m$. In Lemmas 5.2 and 5.3 we omit the index j, i.e., we write ε_i, ψ, ... instead of ε_{ij}, ψ_j, ...

Lemma 5.2. *Let the assumptions of Theorem 2.1 be satisfied. Then, for arbitrary $C>0$ and with a_0 from Assumption* (B.5)*, we have on the set $\{\sup_{|a|+|b|\le C}\max_{i=1,\dots,m}|a+b\widetilde{r}_{iM}|/\sqrt{m}\le a_0\}$, as $m\to\infty$,*

$$\sup_{|a|+|b|\le C}|Z_m(a,b)-E^*Z_m(a,b)|=O_P\left(m^{-\eta}\right),$$

$$E^*Z_m(a,b)=\frac{\lambda'(0)}{2}\left(a^2+b^2\frac{1}{m}\sum_{i=1}^m\widetilde{r}^2_{iM}\right)+O_P\left(m^{-1/2}|a|^3+|b|^3m^{-\eta}\right),$$

and

$$\sup_{|a|+|b|\le C}\left|Z_m(a,b)-\frac{\lambda'(0)}{2}\left(a^2+b^2\frac{1}{m}\sum_{i=1}^m \widetilde{r}_{iM}^2\right)\right| = O_P(m^{-\eta}),$$

for some $\eta > 0$, *where*

$$Z_m(a,b)=\sum_{i=1}^m\big(\rho(\varepsilon_i-a/\sqrt{m}-b\widetilde{r}_{iM}/\sqrt{m})-\rho(\varepsilon_i)+(a/\sqrt{m}+b\widetilde{r}_{iM}/\sqrt{m})\psi(\varepsilon_i)\big).$$

Proof. The lines of the proof are quite standard. We just need to derive a proper approximation for the conditional expectation and variance of $Z_m(a,b)$.

Whenever convenient we use the short-hand notations

$$d_i = a+b\widetilde{r}_{iM} \quad \text{and} \quad g(\varepsilon_i,x,d_i)=\operatorname{sign} d_i\big(-\psi(\varepsilon_i-x\operatorname{sign} d_i)+\psi(\varepsilon_i)\big), \quad i\in\mathbb{Z}. \tag{5.3}$$

Note that, for any d,

$$\rho(\varepsilon_i-d)-\rho(\varepsilon_i)+d\psi(\varepsilon_i)=\operatorname{sign} d\int_0^{|d|}\big(-\psi(\varepsilon_i-x\operatorname{sign} d)+\psi(\varepsilon_i)\big)dx\ge 0,\quad i\in\mathbb{Z}.$$

Direct calculations in combination with Lemma 5.1 result in

$$\begin{aligned}E^*Z_m(a,b) &= E^*\sum_{i=1}^m\operatorname{sign} d_i\int_0^{|d_i|/\sqrt{m}} g(\varepsilon_i,x,d_i)dx=\sum_{i=1}^m\lambda'(0)d_i^2\frac{1}{2m}+O_P\left(\sum_{i=1}^m|d_i|^3\frac{1}{m^{3/2}}\right)\\ &=\frac{1}{2}\lambda'(0)\left(a^2+b^2\frac{1}{m}\sum_{i=1}^m\widetilde{r}_{iM}^2\right)+O_P\big(m^{-1/2}|a|^3+|b|^3m^{-3/2+\max(1,3/(2+\Delta))}\big),\end{aligned}$$

uniformly in $|a|+|b|\le C$.

For the conditional variance we obtain

$$\begin{aligned}var^*\{Z_m(a,b)\} &= E^*\left(\sum_{i=1}^m\int_0^{|d_i|/\sqrt{m}}\big(g(\varepsilon_i,x,d_i)-E^*g(\varepsilon_i,x,d_i)\big)dx\right)^2\\ &=\sum_{i_1=1}^m E^*\left(\int_0^{|d_{i_1}|/\sqrt{m}}\big(g(\varepsilon_{i_1},x,d_{i_1})-E^*g(\varepsilon_{i_1},x,d_{i_1})\big)dx\right)^2\\ &\quad+2E^*\sum_{1\le i_1<i_2\le m}\left\{\left(\int_0^{|d_{i_1}|/\sqrt{m}}\big(g(\varepsilon_{i_1},x,d_{i_1})-E^*g(\varepsilon_{i_1},x,d_{i_1})\big)dx\right)\right.\\ &\quad\times\left.\left(\int_0^{|d_{i_2}|/\sqrt{m}}\big(g(\varepsilon_{i_2},y,d_{i_2})-E^*g(\varepsilon_{i_2},y,d_{i_2})\big)dy\right)\right\}=I_1+I_2\quad\text{(say)}.\end{aligned}$$

Using Assumption (A.2) together with the Cauchy–Schwarz inequality, we get

$$\begin{aligned}I_1 &= \sum_{i_1=1}^m E^*\left(\int_0^{|d_{i_1}|/\sqrt{m}}\big(g(\varepsilon_{i_1},x,d_{i_1})-E^*g(\varepsilon_{i_1},x,d_{i_1})\big)dx\right)^2\le D\sum_{i_1=1}^m\left(\int_0^{|d_{i_1}|/\sqrt{m}}y^{\kappa/2}dy\right)^2\\ &\le D\sum_{i_1=1}^m(|d_{i_1}|/\sqrt{m})^{\kappa+2}\le D\max_{i=1,\ldots,m}(|d_i|/\sqrt{m})^{\kappa}\frac{1}{m}\sum_{i_1=1}^m|d_{i_1}|^2,\end{aligned}$$

uniformly in $|a|+|b|\le C$, where κ is from Assumption (A.2).

In view of Lemma 5.1(iii),

$$\sup_{|a|+|b|\le C}\max_{i=1,\ldots,m}(|d_i|/\sqrt{m})^{\kappa}=O_P\big(m^{-\eta}\big)\quad(m\to\infty),$$

with some $\eta > 0$, hence

$$P\left(\sup_{|a|+|b|\le C}\max_{i=1,\ldots,m}|d_i|/\sqrt{m}\le a_0\right)\to 1\quad(m\to\infty), \tag{5.4}$$

which, in view of the ergodic theorem, results in

$$\sup_{|a|+|b|\le C} |I_1| = O_P\big(m^{-\eta}\big) \quad (m \to \infty),$$

for some $\eta > 0$.

Concerning I_2 we have, due to the independence of $\{r_{iM}\}$ and $\{\boldsymbol{\varepsilon}_i\}$,

$$\begin{aligned}
I_2 \;&\le\; 2\sum_{i_1=1}^{m-1}\sum_{v=1}^{m-i_1}\int_0^{|d_{i_1}|/\sqrt{m}}\int_0^{|d_{i_1+v}|/\sqrt{m}}\Big(E^*(g(\varepsilon_{i_1},x,d_{i_1}))^2\\
&\quad\times\big(E^*(-\psi(\varepsilon_{i_1+v}-y)+\psi(\varepsilon^{(v)}_{i_1+v}-y))^2+E^*(-\psi(\varepsilon_{i_1+v})+\psi(\varepsilon^{(v)}_{i_1+v}))^2\big)\Big)^{1/2}dxdy\\
&\le D\sum_{i_1=1}^{m-1}|d_{i_1}/\sqrt{m}|^{\kappa/2+1}\sum_{v=1}^{m-i_1}|d_{i_1+v}/\sqrt{m}|\sup_{|a|\le a_0}\Big(E^*\big(\psi(\varepsilon_{i_1+v}-a)-\psi(\varepsilon^{(v)}_{i_1+v}-a)\big)^2\Big)^{1/2}\\
&\le D\max_{i=1,\dots,m}(|d_i|/\sqrt{m})^{\kappa/2}\frac{1}{m}\sum_{i_1=1}^{m-1}\sum_{v=1}^{m-i_1}|d_{i_1}d_{i_1+v}|\sup_{|a|\le a_0}\Big(E\big(\psi(\varepsilon_0-a)-\psi(\varepsilon^{(v)}_0-a)\big)^2\Big)^{1/2},
\end{aligned}$$

if $\max_{1\le i\le m}|d_i|/\sqrt{m} \le a_0$, where a_0 is from Assumption (B.5). Since $E|d_{i_1}d_{i_1+v}| \le E|d_0|^2$, the latter assumption gives

$$\sup_{|a|+|b|\le C} |I_2| = O_P\big(m^{-\eta}\big) \quad (m \to \infty),$$

with some $\eta > 0$.

On combining the above estimates for $E^*Z_m(a,b), I_1, I_2$, we conclude that Lemma 5.2 holds true. □

Lemma 5.3. *Let the assumptions of Theorem 2.1 be satisfied. Then, for arbitrary $C > 0$ and with a_0 from Assumption (B.5), we have on the set $\{\sup_{|a|+|b|\le C}\max_{i=1,\dots,m}|a+b\widetilde{r}_{iM}|/\sqrt{m}\le a_0\}$, as $m\to\infty$,*

$$\sup_{|a|+|b|\le C}|\mathbf{M}_m(a,b)-E^*\mathbf{M}_m(a,b)| = O_P(m^{-\eta}),$$

$$E^*\mathbf{M}_m(a,b) = -\frac{1}{m}\lambda'(0)\left(a\,m, b\sum_{i=1}^m\widetilde{r}^2_{iM}\right)^T + O_P\big(m^{-\eta}\big),$$

and

$$\sup_{|a|+|b|\le C}\Big|\mathbf{M}_m(a,b)+\frac{1}{m}\lambda'(0)(a\,m, b\sum_{i=1}^m\widetilde{r}^2_{iM})^T\Big| = O_P\big(m^{-\eta}\big),$$

with some positive η, where

$$\mathbf{M}_m(a,b) = \frac{1}{\sqrt{m}}\sum_{i=1}^m(1,\widetilde{r}_{iM})^T\big(\psi(\varepsilon_i-(a+b\widetilde{r}_{iM})/\sqrt{m})-\psi(\varepsilon_i)\big).$$

Proof. Again one has to get suitable approximations for the conditional expectation $\mathbf{M}_m(a,b)$ and the conditional (2×2)-variance matrix

$$var^*\{\mathbf{M}_n(a,b)\} = E^*\Big(\mathbf{M}_n(a,b)-E^*\mathbf{M}_n(a,b)\Big)\Big(\mathbf{M}_n(a,b)-E^*\mathbf{M}_n(a,b)\Big)^T.$$

We start with the conditional expectation

$$\begin{aligned}
E^*\mathbf{M}^T_m(a,b) \;&=\; \frac{1}{\sqrt{m}}\sum_{i=1}^m(1,\widetilde{r}_{iM})(-\lambda(d_i/\sqrt{m}))\\
&= -\frac{1}{\sqrt{m}}\lambda'(0)\sum_{i=1}^m(1,\widetilde{r}_{iM})\big(d_i/\sqrt{m}+O_P(|d_i/\sqrt{m}|^2)\big)\\
&= -\frac{1}{m}\lambda'(0)\left(a\,m, b\sum_{i=1}^m\widetilde{r}^2_{iM}\right)+O_P\big((a^2+b^2)m^{-1/2}+b^2m^{\max(-1/2,-3\Delta/(2(2+\Delta)))}\big),
\end{aligned}$$

uniformly in $|a| + |b| \le C$. Note that

$$E^*\big(\mathbf{M}_m^T(a_1, b_1) - \mathbf{M}_m^T(a_2, b_2)\big) = -\frac{1}{m}\lambda'(0)\Big((a_1 - a_2)m, (b_1 - b_2)\sum_{i=1}^{m}\widetilde{r}_{iM}^2\Big) + O_P\big((a_1^2 + b_1^2 + a_2^2 + b_2^2)m^{-1/2} + (b_1^2 + b_2^2)m^{\max(-1/2, -3\Delta/(2(2+\Delta)))}\big).$$

We only calculate one term of the conditional variance matrix. The calculation of the others is similar and will therefore be omitted. We have

$$var^*\Big\{\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\widetilde{r}_{iM}\big(\psi(\varepsilon_i - d_i/\sqrt{m}) - \psi(\varepsilon_i)\big)\Big\} = \frac{1}{m}\sum_{i=1}^{m}\widetilde{r}_{iM}^2 E^*\big(\psi(\varepsilon_i - d_i/\sqrt{m}) - \psi(\varepsilon_i) + \lambda(d_i/\sqrt{m})\big)^2$$
$$+ 2\frac{1}{m}\sum_{i=1}^{m}\sum_{v=1}^{m-i}\widetilde{r}_{iM}\widetilde{r}_{i+v,M}E^*\big(\psi(\varepsilon_i - d_i/\sqrt{m}) - \psi(\varepsilon_i) + \lambda(d_i/\sqrt{m})\big)$$
$$\times\big(\psi(\varepsilon_{i+v} - d_{i+v}/\sqrt{m}) - \psi(\varepsilon_{i+v}) + \lambda(d_{i+v}/\sqrt{m})\big) = J_1 + 2J_2 \quad \text{(say)}.$$

In view of Assumption (A.2), a similar estimate as that for I_1 in the proof of Lemma 5.2 gives

$$J_1 = \tfrac{1}{m}\sum_{i=1}^{m}\widetilde{r}_{iM}^2 E^*\big(\psi(\varepsilon_i - d_i/\sqrt{m}) - \psi(\varepsilon_i) + \lambda(d_i/\sqrt{m})\big)^2$$
$$\le D\tfrac{1}{m^{1+\kappa/2}}\Big(|a|^\kappa\sum_{i=1}^{m}\widetilde{r}_{iM}^2 + |b|^\kappa\sum_{i=1}^{m}|\widetilde{r}_{iM}|^{2+\kappa}\Big) = O_P\big(m^{-\xi}\big),$$

with some $\xi > 0$, uniformly in $|a| + |b| \le C$.

Concerning J_2 we obtain

$$J_2 = \frac{1}{m}\sum_{i=1}^{m}\sum_{v=1}^{m-i}\widetilde{r}_{iM}\widetilde{r}_{i+v,M}E^*\big(\psi(\varepsilon_i - d_i/\sqrt{m}) - \psi(\varepsilon_i) + \lambda(d_i/\sqrt{m})\big)$$
$$\times\big(\psi(\varepsilon_{i+v} - d_{i+v}/\sqrt{m}) - \psi(\varepsilon_{i+v}) - (\psi(\varepsilon_{i+v}^{(v)} - d_{i+v}/\sqrt{m}) - \psi(\varepsilon_{i+v}^{(v)}))\big),$$

and

$$|J_2| \le D\frac{1}{m}\sum_{i=1}^{m}\sum_{v=1}^{m-i}|\widetilde{r}_{iM}\widetilde{r}_{i+v,M}|(|d_i|/\sqrt{m})^{\kappa/2}$$
$$\times\big\{E^*\big(\psi(\varepsilon_{i+v} - d_{i+v}/\sqrt{m}) - \psi(\varepsilon_{i+v}) - (\psi(\varepsilon_{i+v}^{(v)} - d_{i+v}/\sqrt{m}) - \psi(\varepsilon_{i+v}^{(v)}))\big)^2\big\}^{1/2}$$
$$\le D\max_{i=1,\dots,m}(|d_i|/\sqrt{m})^{\kappa/2}\frac{1}{m}\sum_{i=1}^{m}\sum_{v=1}^{m-i}|\widetilde{r}_{iM}\widetilde{r}_{i+v,M}|\sup_{|a|\le a_0}\big(E(\psi(\varepsilon_0 - a) - \psi(\varepsilon_0^{(v)} - a))^2\big)^{1/2}.$$

Now, a similar estimate as that for I_2 in the proof of Lemma 5.2 gives

$$\sup_{|a|+|b|\le C}|J_2| = O_P\big(m^{-\xi}\big),$$

with some $\xi > 0$, so that altogether we have

$$\sup_{|a|+|b|\le C} var^*\left\{\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\widetilde{r}_{iM}\big(\psi(\varepsilon_i - d_i/\sqrt{m}) - \psi(\varepsilon_i)\big)\right\} = O_P\big(m^{-\xi}\big) \quad (m \to \infty),$$

for some $\xi > 0$. □

Lemma 5.4. *Let the assumptions of Theorem 2.1 be satisfied. Then, for any $T > 0$, as $m \to \infty$,*

$$\max_{1\le k\le\lfloor mT\rfloor}\left(\frac{|\{N_{k,m}(a, b) - E^*N_{km}(a, b)\}_{a=\widehat{\alpha}_{j,m}^*, b=\widehat{\beta}_{j,m}^*}|}{(k/m)^\gamma}\right) = O_P(m^{-\eta}),$$

for some $\eta > 0$, where $\widehat{\alpha}_{j,m}^*$, $\widehat{\beta}_{j,m}^*$ are as in (4.3), and

$$N_{k,m}(a, b) = \frac{1}{\sqrt{m}}\sum_{i=m+1}^{m+k}\widetilde{r}_{iM}\big(\psi(\varepsilon_i - a/\sqrt{m} - b\widetilde{r}_{iM}/\sqrt{m}) - \psi(\varepsilon_i)\big).$$

Proof. Lemma 5.4 is related to Lemma 5.3, but it is somewhat more complicated.

Direct calculations give

$$E^*N_{k,m}(a,b) = -\frac{1}{\sqrt{m}}\sum_{i=m+1}^{m+k}\widetilde{r}_{iM}\lambda\big((a+b\widetilde{r}_{iM})/\sqrt{m}\big)$$
$$= -\lambda'(0)\frac{1}{m}\left(a\sum_{i=m+1}^{m+k}\widetilde{r}_{iM} + b\sum_{i=m+1}^{m+k}\widetilde{r}_{iM}^2\right) + O_P\big(m^{-\eta}\big) \quad (m\to\infty),$$

uniformly for $|a|+|b|\le C$, with some $\eta>0$.

Next, we try to get an upper bound for $var^*\{N_{k,m}(a,b)\}$. We have

$$var^*\{N_{k,m}(a,b)\} = \frac{1}{m}\sum_{i=m+1}^{m+k}\widetilde{r}_{iM}^2E^*\big(\psi(\varepsilon_i-d_i/\sqrt{m})-\psi(\varepsilon_i)-E^*\psi(\varepsilon_i-d_i/\sqrt{m})\big)^2$$
$$+2\frac{1}{m}\sum_{i=m+1}^{m+k}E^*\widetilde{r}_{iM}\Big(\psi(\varepsilon_i-d_i/\sqrt{m})-\psi(\varepsilon_i)-E^*\psi(\varepsilon_i-d_i/\sqrt{m})\Big)$$
$$\times\left(\sum_{v=1}^{m+k-i}\widetilde{r}_{i+v,M}\big(\psi(\varepsilon_{i+v}^{(v)}-d_{i+v}/\sqrt{m})-\psi(\varepsilon_{i+v}^{(v)})-E^*\psi(\varepsilon_{i+v}^{(v)}-d_{i+v}/\sqrt{m})\big)\right)$$
$$= L_{1,k}+2L_{2,k} \quad \text{(say)},$$

and, along the lines of the proof of Lemma 5.3, we get

$$L_{1,k}\le\frac{D}{m^{1+\kappa/2}}\left(|a|^\kappa\sum_{i=m+1}^{m+k}\widetilde{r}_{iM}^2+|b|^\kappa\sum_{i=m+1}^{m+k}|\widetilde{r}_{iM}|^{2+\kappa}\right)=\frac{k}{m}m^{-\kappa/2}\big(|a|^\kappa+|b|^\kappa\big)m^{\max(0,(\kappa-\Delta)/(2+\Delta))}O_P(1),$$
$$|L_{2,k}|=\frac{1}{m}\sum_{i=m+1}^{m+k}|\widetilde{r}_{iM}||(a+b\widetilde{r}_{i+v,M})/\sqrt{m}|^{\kappa/2}O_P(1)=\frac{1}{m^{1+\kappa/2}}(|a|^{\kappa/2}k+|b|^{\kappa/2}k)O_P(1) \quad (m\to\infty),$$

uniformly in $|a|+|b|\le C$ and in $1\le k\le\lfloor mT\rfloor$. So, altogether we have

$$var^*\{N_{k,m}(a,b)\}=\frac{k}{m}m^{-\zeta}\max(|a|^\kappa+|b|^\kappa,|a|^{\kappa/2}+|b|^{\kappa/2})O_P(1) \quad (m\to\infty),$$

uniformly in $|a|+|b|\le C$ and in $1\le k\le\lfloor mT\rfloor$, with some $\zeta>0$.

Quite similarly we get, for $v=1,2,\ldots,$

$$var^*\{N_{k+v,m}(a,b)-N_{k,m}(a,b)\}=\frac{v}{m}m^{-\xi}\max(|a|^\kappa+|b|^\kappa,|a|^{\kappa/2}+|b|^{\kappa/2})O_P(1) \quad (m\to\infty).$$

Then, on applying Theorem B.4 of Kirch [20],

$$m^{-1+2\gamma}E^*\max_{1\le k\le\lfloor mT\rfloor}\left(\frac{1}{k^\gamma}|N_{m,k}(a,b)-E^*N_{m,k}(a,b)|\right)^2$$
$$=m^{-1+2\gamma}(\log m)^2\sum_{k=1}^{\lfloor mT\rfloor}\frac{1}{k^{2\gamma}}m^{-\xi}\max(|a|^\kappa+|b|^\kappa,|a|^{\kappa/2}+|b|^{\kappa/2})O_P(1)$$
$$=(\log m)^2m^{-\xi}(1+m^{-2\gamma+1}m^{-1+2\gamma})\max(|a|^\kappa+|b|^\kappa,|a|^{\kappa/2}+|b|^{\kappa/2})O_P(1)$$
$$=(\log m)^2m^{-\xi}\max(|a|^\kappa+|b|^\kappa,|a|^{\kappa/2}+|b|^{\kappa/2})O_P(1) \quad (m\to\infty).$$

We need to replace a,b by the estimators $\widehat{\alpha}_{jm}^*,\widehat{\beta}_{jm}^*$. However our $N_{k,m}(a,b)$ depends on $\varepsilon_1,\ldots,\varepsilon_m$. Therefore we try to replace $N_{k,m}(a,b)$ by something that is asymptotically equivalent, but does not depend on $\varepsilon_1,\ldots,\varepsilon_m$.

Toward this note that

$$N_{k,m}^{(m)}(a,b)=\frac{1}{\sqrt{m}}\sum_{i=1}^{k}\widetilde{r}_{i+m,M}\big(\psi(\varepsilon_{i+m}^{(i)}-d_i/\sqrt{m})-\psi(\varepsilon_{m+i}^{(i)})\big)$$

has all the properties of $N_{k,m}(a,b)$ above, but it is independent of $\varepsilon_1,\ldots,\varepsilon_m$. This together with the consistency of $\widehat{\alpha}^*_{jm}$ and $\widehat{\beta}^*_{jm}$ implies

$$m^{-1+2\gamma}\max_{1\le k\le\lfloor mT\rfloor}\left(\frac{1}{k^\gamma}\left|\{N^{(m)}_{m,k}(a,b)-E^*N^{(m)}_{m,k}(a,b)\}_{a=\widehat{\alpha}^*_{jm},b=\widehat{\beta}^*_{jm}}\right|\right)^2$$
$$=O_P\big((\log m)^2m^{-\xi}\max(|\widehat{\alpha}^*_{jm}|^\kappa+|\widehat{\beta}^*_{jm}|^\kappa,|\widehat{+}\alpha^*_{jm}|^{\kappa/2}+|\widehat{\beta}^*_{jm}|^{\kappa/2})\big)=O_P\big((\log m)^2m^{-\xi}\big)\quad(m\to\infty).$$

It is still necessary to show the closeness of $N_{k,m}(a,b)$ and $N^{(m)}_{k,m}(a,b)$. Clearly, $N^{(m)}_{k,m}(a,b)$ is independent of $\varepsilon_1,\ldots,\varepsilon_m$ and

$$E^*\big(N_{k,m}(a,b)-N^{(m)}_{k,m}(a,b)\big)=0,$$
$$N_{k,m}(a,b)-N^{(m)}_{k,m}(a,b)=\frac{1}{\sqrt{m}}\sum_{i=1}^k\widetilde{r}_{i+m,M}\left(\left(\psi\left(\varepsilon_{i+m}-\frac{d_i}{\sqrt{m}}\right)-\psi\left(\varepsilon^{(i)}_{i+m}-\frac{d_i}{\sqrt{m}}\right)\right)-\big(\psi(\varepsilon_{i+m})-\psi(\varepsilon^{(i)}_{i+m})\big)\right),$$
$$E^*|N_{k,m}(a,b)-N^{(m)}_{k,m}(a,b)|\le\frac{D}{\sqrt{m}}\sum_{i=1}^k|\widetilde{r}_{i+m,M}|\sup_{|a|\le a_0}E\big(|\psi(\varepsilon_0-a)-\psi(\varepsilon^{(i)}_0-a)|+|\psi(\varepsilon_0)-\psi(\varepsilon^{(i)}_0)|\big)$$
$$\le\frac{D}{\sqrt{m}}\sum_{i=1}^{\lfloor mT\rfloor}|\widetilde{r}_{i+m,M}|\sup_{|a|\le a_0}E|\psi(\varepsilon_0-a)-\psi(\varepsilon^{(i)}_0-a)|,$$

which holds for any $1\le k\le\lfloor mT\rfloor$ and any a,b such that $|d_i|/\sqrt{m}\le a_0$. So, in view of Assumption (B.5),

$$\sup_{|a|+|b|\le C}E^*|N_{k,m}(a,b)-N^{(m)}_{k,m}(a,b)|=O_P(m^{-1/2})\quad(m\to\infty),$$

whence

$$\sup_{1\le k\le\lfloor mT\rfloor}\left(\frac{\sup_{|a|+|b|\le C}E^*|N_{k,m}(a,b)-N^{(m)}_{k,m}(a,b)|}{(k/m)^\gamma}\right)=O_P(m^{-\eta})\quad(m\to\infty),$$

for some $\eta>0$. A combination of the above estimates completes the proof of Lemma 5.4. □

Acknowledgments

The authors would like to thank the referees for a careful reading of the paper and their helpful suggestions. They also gratefully acknowledge support by grants GAUK 162310 (O. Chochola), GAČR 201/09/J006 and GAČR P201/12/1277 (M. Hušková and Z. Prášková), and DFG grant STE 306/22-1 (J.G. Steinebach).

References

[1] A. Aue, S. Hörmann, L. Horváth, M. Hušková, J.G. Steinebach, Sequential testing for the stability of high-frequency portfolio betas, Econom. Theory 28 (2012) 804–837.
[2] A. Aue, L. Horváth, M. Hušková, P. Kokoszka, Change-point monitoring in linear models, Econom. J. 9 (2006) 373–403.
[3] Z.D. Bai, X.R. Chen, B.Q. Miao, C. Radhakrishna Rao, Y. Wu, Asymptotic theory of least distances estimate in multivariate linear models, Statistics 21 (1990) 503–519.
[4] Z.D. Bai, C. Radhakrishna Rao, Y. Wu, M-estimation of multivariate linear regression parameters under convex discrepancy, Statistica Sinica 2 (1992) 237–254.
[5] I. Berkes, S. Hörmann, J. Schauer, Split invariance principles for stationary processes, Ann. Probab. 39 (2011) 2441–2473.
[6] P. Billingsley, Convergence of Probability Measures, Wiley, New York, 1968.
[7] T. Caporale, Time varying CAPM betas and banking sector risk, Econom. Lett. 115 (2012) 293–295.
[8] A. Černíková, M. Hušková, Z. Prášková, J.G. Steinebach, Delay time in monitoring jump changes in linear models, Statistics (2011) 25. online available; http://dx.doi.org/10.1080/02331888.2011.577895.
[9] C.-S.J. Chu, M. Stinchcombe, H. White, Monitoring structural change, Econometrica 64 (1996) 1045–1065.
[10] M. Csörgő, L. Horváth, Limit Theorems in Change-point Analysis, Wiley, Chichester, 1997.
[11] J. Davidson, Stochastic Limit Theory, Oxford University Press, New York, 1994.
[12] M.G. Genton, E. Ronchetti, Robust prediction of beta, in: E.J. Kontoghiorghes, B. Rustem, P. Winker (Eds.), Computational Methods in Financial Engineering, in: Essays in Honour of Manfred Gilli, Springer, New York, 2008, pp. 147–161.
[13] E. Ghysels, On stable factor structures in the pricing of risk: do time-varying betas help or hurt? J. Finance 53 (1998) 549–573.
[14] S. Hörmann, P. Kokoszka, Weakly dependent functional data, Ann. Statist. 38 (2010) 1845–1884.
[15] L. Horváth, M. Hušková, P. Kokoszka, J.G. Steinebach, Monitoring changes in linear models, J. Statist. Plann. Inference 126 (2004) 225–251.
[16] L. Horváth, P. Kokoszka, Inference for Functional Data with Applications, Springer, New York, 2012.
[17] P.J. Huber, Robust Statistics, Wiley, New York, 1981.
[18] J. Jurečková, P.K. Sen, Robust Statistical Procedures. Asymptotics and Interrelations, Wiley, New York, 1996.
[19] R. Koenker, S. Portnoy, M estimation of multivariate regressions, J. Amer. Statist. Assoc. 85 (1990) 1060–1068.
[20] C. Kirch, Resampling Methods for the Change Analysis of Dependent Data, Ph.D. Thesis, University of Cologne, 2006.
[21] A. Koubková, Sequential Change-Point Analysis, Ph.D. Thesis, Charles University, Prague, 2006.
[22] J. Lintner, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, Rev. Econ. Stat. 47 (1965) 13–37.
[23] R.D. Martin, T.T. Simin, Outlier-resistant estimates of beta, Financ. Anal. J. 59 (2003) 56–69.
[24] R. Merton, An intertemporal asset pricing model, Econometrica 41 (1973) 867–880.
[25] W.F. Sharpe, Capital asset prices: a theory of market equilibrium under conditions of risk, J. Finance 19 (1964) 424–442.
[26] W.B. Wu, M-estimation of linear models with dependent errors, Ann. Statist. 35 (2007) 495–521.