Econometrics Journal (2006), volume **9**, pp. 373–403.
doi: 10.1111/j.1368-423X.2006.00190.x

Change-point monitoring in linear models

ALEXANDER AUE[0], LAJOS HORVÁTH[1], MARIE HUŠKOVÁ[2]
AND PIOTR KOKOSZKA[3]

[0]*Department of Mathematical Sciences, Clemson University, O-324 Martin Hall, Clemson, SC 29634, USA*
E-mail: `alexaue@clemson.edu`

[1]*Department of Mathematics, University of Utah, 155 South 1440 East, Salt Lake City, UT 84112–0090, USA*
E-mails: `aue@math.utah.edu` and `horvath@math.utah.edu`

[2]*Department of Statistics, Charles University, Sokolovská 83, CZ–18600 Praha, Czech Republic*
E-mail: `huskova@karlin.mff.cuni.cz`

[3]*Department of Mathematics and Statistics, Utah State University, 3900 Old Main Hill, Logan, UT 843322–3900, USA*
E-mail: `Piotr.Kokoszka@usu.edu`

Received: September 9, 2005

Summary We consider a linear regression model with errors modelled by martingale difference sequences, which include heteroskedastic augmented GARCH processes. We develop asymptotic theory for two monitoring schemes aimed at detecting a change in the regression parameters. The first method is based on the CUSUM of the residuals and was studied earlier in the context of independent identically distributed errors. The second method is new and is based on the squares of prediction errors. Both methods use a training sample of size m. We show that, as $m \to \infty$, both methods have correct asymptotic size and detect a change with probability approaching unity. The methods are illustrated and compared in a small simulation study.

Key words: Heteroskedastic errors, Linear model, Prediction errors, On-line change-point detection, Residuals.

1. INTRODUCTION

Changes in economic environment are likely to induce structural instability in the initially chosen underlying econometric model. Therefore, testing for possible structural changes in model parameters has recently become one of the principal objectives of econometric analysis. There are two distinctly different approaches to tackle such problems, namely, (1) retrospective or a posteriori tests, and (2) sequential (on-line) or a priori tests. The present paper is concerned with the second approach which is more useful when a decision has to be made on–line, as new data become available. We follow the general paradigm of Chu *et al.* (1996) in which the initial time period of length m is used to estimate a model, and the goal is to monitor for changes in model parameters, ensuring that the probability of a false detection does not exceed a prespecified level α. In this approach, the asymptotic analysis is carried out assuming that $m \to \infty$, and its goal is to verify that the probability of the detection approaches α under the null of no change and one under the alternative of a change in parameters after the initial time period. Following the work of Chu *et al.* (1996), Horváth *et al.* (2004) developed the asymptotic theory for two new classes of monitoring schemes based on residuals and recursive residuals and appropriately chosen boundary functions. The latter paper, however, considered a linear regression model with independent identically distributed (i.i.d.) errors. While the assumption of i.i.d. errors is convenient from the mathematical point of view, it is typically violated in regressions involving econometric variables.

The goal of the present paper is to develop the theory of sequential change-point testing for two monitoring schemes. The first scheme is the CUSUM of residuals test introduced in Horváth *et al.* (2004). We show that this test is robust, for instance, to heteroskedasticity in the regression errors and inherits all the desirable properties established for regressions with i.i.d. errors. The second scheme is motivated by the work of Clark and McCracken (2005) and is based on squared prediction errors. Allowing heteroskedasticity implies that the squared prediction errors are (even asymptotically) serially correlated, so the asymptotic analysis is much more delicate than for asymptotically uncorrelated residuals.

Even though the literature on sequential monitoring of linear models is still rather small, the methods considered here have a rich background in the retrospective setting. Quandt (1958, 1960) derived the maximum likelihood ratio test under the assumption of normally distributed errors. Further properties and refinements of this maximally selected F-test were developed by Andrews (1991, 1993) and Bai *et al.* (1998) who studied the asymptotic behaviour of a trimmed (restricted) version, and by Horváth (1995) who obtained the exact limit distribution. Wald-type and union-intersection tests were proposed in Hawkins (1987, 1989). Their precise asymptotics were derived in Horváth and Shao (1995). Csörgő and Horváth (1997, p. 298) showed that the union-intersection and the maximally selected tests are asymptotically equivalent. The above methods are based on comparisons of parameter estimates. For further references and details, we refer to Bai and Perron (1998, 2003), and Hansen (1995, 2000), among others. Using residuals to detect changes in regression parameters was advocated by Brown *et al.* (1975) whose approach has been subsequently refined and specialized by many authors (see e.g. Krämer *et al.* 1988; Ploberger *et al.* 1989; Gombay and Horváth 1994). In the context of evaluating out-of-sample forecasts, McCracken (2000) [see also Clark and McCracken (2001, 2005)] advocated tests based on prediction errors.

The paper is organized as follows. In Section 2, we introduce the testing problem and list the assumptions. The test statistics and their limiting behaviour under both null and alternative hypothesis are stated in Section 3. Before proving our results in Section 5, we illustrate them by means of a small simulation study in Section 4.

2. THE TESTING PROBLEM AND THE ASSUMPTIONS

We consider the linear regression model

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta}_i + \varepsilon_i, \qquad 1 \le i < \infty,$$

where $\mathbf{x}_i$ is a $p \times 1$ dimensional random or deterministic vector of the form

$$\mathbf{x}_i^T = (1, x_{2,1}, \ldots, x_{p,i}), \tag{1}$$

$\boldsymbol{\beta}_i$ is a $p \times 1$ dimensional parameter vector and $\{\varepsilon_i\}$ is an error sequence. Our first assumption states that there is no change in the regression parameter during the first m observations, i.e.

$$\boldsymbol{\beta}_i = \boldsymbol{\beta}_0, \qquad 1 \le i \le m. \tag{2}$$

Condition (2) was called the 'non-contamination assumption' in Chu *et al.* (1996). It is particularly important because the test statistics to be defined below use the historical data set as a reference for comparisons with later observations.

We wish to test the no change in the regression parameter null hypothesis

$$H_0: \quad \boldsymbol{\beta}_i = \boldsymbol{\beta}_0, \qquad i = m+1, m+2, \dots, \tag{3}$$

against the alternative hypothesis

$$\begin{aligned} H_A: &\text{there is a } k^* \ge 1 \text{ such that } \boldsymbol{\beta}_i = \boldsymbol{\beta}_0, \quad m < i < m + k^*, \\ &\text{but } \boldsymbol{\beta}_i = \boldsymbol{\beta}_*, \quad i = m + k^*, m + k^* + 1, \dots \text{ with } \boldsymbol{\beta}_0 \ne \boldsymbol{\beta}_*. \end{aligned} \tag{4}$$

The parameters $\boldsymbol{\beta}_0$, $\boldsymbol{\beta}_*$ and k^*, the so-called change-point, are assumed unknown.

The monitoring procedures studied in this paper use a detector function $\Gamma(m, k)$ and a boundary function $g(m, k)$ which together define the stopping time

$$\tau(m) = \inf\{k \ge 1 : |\Gamma(m, k)| \ge g(m, k)\} \tag{5}$$

(with the understanding that $\inf \emptyset = \infty$) which must satisfy

$$\lim_{m\to\infty} P\{\tau(m) < \infty\} = \alpha, \quad \text{under } H_0; \tag{6}$$

$$\lim_{m\to\infty} P\{\tau(m) < \infty\} = 1, \quad \text{under } H_A. \tag{7}$$

The index k labels the time elapsed after the monitoring has commenced. The probability $\alpha \in (0, 1)$ controls the false alarm rate. Condition (6) ensures that the probability of a false alarm is asymptotically bounded by α, while condition (7) means that a change-point is detected with probability approaching one.

In the remainder of this section, we state the assumptions on the regression model.

As mentioned in the 'Introduction', we allow the errors ε_i to exhibit conditional heteroskedasticity. However, we will consider somewhat more general innovation sequences to be defined now. We assume that

$$E\varepsilon_i = 0, \qquad E\varepsilon_i\varepsilon_j = 0 \quad (i \ne j) \qquad \text{and} \qquad E\varepsilon_i^2 \le C \quad \text{with some } C > 0. \tag{8}$$

Under appropriate assumptions, condition (8) is satisfied, for example, for so-called augmented GARCH(1,1) processes introduced by Duan (1997). See Section 6 for a broader discussion.

In order to be able to estimate the asymptotic variance of the squared prediction error, we must assume that

$$E\varepsilon_i^4 < \infty \qquad \text{and} \qquad \mu^2 := \operatorname{Var}\varepsilon_0^2 + 2\sum_{i=1}^{\infty}\operatorname{Cov}\left(\varepsilon_0^2, \varepsilon_i^2\right) > 0. \tag{9}$$

Finally, we impose the usual conditions that there is a positive definite matrix $\mathbf{C}$ and a constant $\kappa > 0$ such that

$$\left|\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^T - \mathbf{C}\right| = \mathcal{O}(n^{-\kappa}) \qquad \text{a.s.} \qquad (n \to \infty) \tag{10}$$

and that

$$\{\varepsilon_i\} \text{ and } \{\mathbf{x}_i\} \text{ are independent.} \tag{11}$$

In (10) and in the following, $|\cdot|$ denotes the maximum norm of both vectors and matrices. Throughout the paper, we work with a realization of the $\mathbf{x}_i$, $1 \le i < \infty$, satisfying the bound in (10).

3. MONITORING PROCEDURES AND THEIR ASYMPTOTICS

In this section, we define the two monitoring procedures and state their asymptotic properties under both the null and the alternative.

The first procedure is based on the residuals

$$\hat{\varepsilon}_i = y_i - \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_m, \tag{12}$$

where

$$\hat{\boldsymbol{\beta}}_m = \left[\sum_{j=1}^{m} \mathbf{x}_j \mathbf{x}_j^T\right]^{-1} \sum_{j=1}^{m} \mathbf{x}_j y_j \tag{13}$$

is the least-squares estimator computed using the initial m observations. The detector function is

$$Q(m,k) = \sum_{i=m+1}^{m+k} \hat{\varepsilon}_i$$

and the boundary function is

$$g(m,k) = c\sqrt{m}\left(1 + \frac{k}{m}\right)\left(\frac{k}{m+k}\right)^{\gamma}, \qquad c = c(\alpha) > 0,\ 0 \le \gamma < \tfrac{1}{2}. \tag{14}$$

The stopping time is

$$\tau_Q(m) = \inf\{k \ge 1 : |Q(m,k)| \ge g(m,k)\}.$$

This method was studied by Horváth *et al.* (2004) under the assumption of i.i.d. errors ε_i. We will show that its asymptotic properties are unaffected by assuming (8) instead, so, for example, under conditional heteroskedasticity.

The second procedure is motivated by the recent work of Clark and McCracken (2005). Denote by

$$\hat{y}_{\ell+i} = \mathbf{x}_{\ell+i}^T \hat{\boldsymbol{\beta}}_\ell, \qquad i = 1, \ldots, r,$$

the linear predictor of $y_{\ell+i}$ based on the first ℓ observations. Then, under H_0, the squared prediction error $(y_{\ell+i} - \hat{y}_{\ell+i})^2$ should be close to the square of the unobservable error $\varepsilon_{\ell+i}^2$. Following the lines of the proofs in Horváth *et al.* (2004), one can then expect that the detector function

$$R(m,k) = \sum_{i=1}^{r} \sum_{\nu=1}^{k-i} (y_{m+\nu+i} - \hat{y}_{m+\nu+i})^2 - \frac{k}{m} \sum_{i=1}^{r} \sum_{\nu=p}^{m-r} (y_{\nu+i} - \hat{y}_{\nu+i})^2$$

will lead to a monitoring procedure satisfying (6), (7) with the boundary function (14). For completeness, we state that the stopping time is defined by

$$\tau_R(m) = \inf\{k \geq 1 : |R(m,k)| \geq g(m,k)\}.$$

The asymptotic properties of the monitoring procedures are formulated in terms of a functional of the standard Brownian motion (Wiener process) which we denote in the following by $\{W(t){:}t \geq 0\}$.

Moreover, the proofs of the theorems to come rely heavily on the following approximations of certain partial sums of the innovations $\{\varepsilon_i\}$. We assume that, for each m, there are independent Wiener processes $\{W_{1,m}(t){:}t \geq 0\}$ and $\{W_{2,m}(t){:}t \geq 0\}$ and a constant $\sigma > 0$ such that

$$\sup_{1 \leq k < \infty} \frac{1}{k^{\Delta}} \left| \sum_{i=m+1}^{m+k} \varepsilon_i - \sigma W_{1,m}(k) \right| = \mathcal{O}_P(1) \qquad (m \to \infty) \tag{15}$$

and

$$\sum_{i=1}^{m} \varepsilon_i - \sigma W_{2,m}(m) = \mathcal{O}_P(m^{\Delta}) \qquad (m \to \infty) \tag{16}$$

with some $\Delta < 1/2$. Conditions (15) and (16) are, in particular, satisfied if $\{\varepsilon_i\}$ is an augmented GARCH sequence. This fact is proved in Lemma 6.1.

We first consider the asymptotics under the null hypothesis.

Theorem 1 *Assume that (2), (8) and (14)–(16) hold, and that* $\gamma < \kappa$. *Then, under* H_0,

$$\lim_{m \to \infty} P\left\{ \frac{1}{\sigma} \max_{1 \leq k < \infty} \frac{|Q(m,k)|}{g(m,k)} \leq 1 \right\} = P\left\{ \sup_{0 < t < 1} \frac{|W(t)|}{t^{\gamma}} \leq c \right\}.$$

The limit distribution of $\sup_{0<t<1} |W(t)| t^{-\gamma}$ is known only in the case $\gamma = 0$. Critical values for other choices of γ were obtained through simulation in Horváth *et al.* (2004).

The application of Theorem 1 requires the estimation of σ^2. Since there is no change in the historical data and the errors are orthogonal, the natural estimator is the sample variance of the residuals

$$\hat{\sigma}_m^2 = \frac{1}{m-p} \sum_{i=1}^{m} \left(\hat{\varepsilon}_i - \frac{1}{m} \sum_{\ell=1}^{m} \hat{\varepsilon}_\ell \right)^2.$$

It can be verified that $\hat{\sigma}_m^2$ converges in probability to σ^2, so Theorem 1 remains true when σ is replaced by $\hat{\sigma}_m$.

To prove a corresponding limit theorem for the prediction approach, we need further conditions. Let $\{\mathcal{G}_k\}$ be a filtration, that is $\mathcal{G}_k \subset \mathcal{G}_{k+1}$ for all k, and assume that

$$(\varepsilon_i, \mathcal{G}_i) \quad \text{is an orthogonal martingale difference sequence.} \tag{17}$$

Also, we require that the innovation variances and fourth moments are the same, that is

$$E\varepsilon_i^2 = \sigma^2 \qquad \text{and} \qquad E\varepsilon_i^4 = \kappa > 0, \qquad i \geq 1. \tag{18}$$

Furthermore, we assume that there are independent Wiener processes $\{W_{3,m}(t) : t \geq 0\}$ and $\{W_{4,m}(t) : t \geq 0\}$ such that

$$\max_{1 \leq k < \infty} \frac{1}{k^{\Delta}} \left| \sum_{i=m+1}^{m+k} \left(\varepsilon_i^2 - \sigma^2\right) - \mu W_{3,m}(k) \right| = \mathcal{O}_P(1) \qquad (m \to \infty) \tag{19}$$

and

$$\sum_{i=1}^{m} \left(\varepsilon_i^2 - \sigma^2\right) - \mu W_{4,m}(m) = \mathcal{O}_P\left(m^{\Delta}\right) \qquad (m \to \infty) \tag{20}$$

with some $\Delta < 1/2$, where μ is as defined in (9). Under suitable assumptions, augmented GARCH processes satisfy conditions (19) and (20), as is shown in Lemma A.2.

Theorem 2 *Assume that (2), (9), (10), (14) and (17)–(20) hold. Then, under* H_0,

$$\lim_{m \to \infty} P\left\{\frac{1}{r\mu} \max_{1 \leq k < \infty} \frac{|R(m,k)|}{g(m,k)} \leq 1\right\} = P\left\{\sup_{0 < t < 1} \frac{|W(t)|}{t^{\gamma}} \leq c\right\}.$$

We also need an estimator for μ^2 defined in (9). This parameter is much harder to estimate than σ^2, since the $\{\varepsilon_i^2 - \sigma^2\}$ are not orthogonal. If $\{\varepsilon_i^2\}$ is a strictly stationary sequence, the Bartlett estimator is defined by

$$\hat{\mu}_m^2 = \hat{\delta}_0 + 2 \sum_{j=1}^{q(m)} \omega_j(q(m)) \hat{\delta}_j,$$

where

$$\hat{\delta}_j = \frac{1}{m} \sum_{i=1}^{m-j} \left(\hat{\varepsilon}_i^2 - \frac{1}{m} \sum_{\ell=1}^{m} \hat{\varepsilon}_\ell^2\right) \left(\hat{\varepsilon}_{i+j}^2 - \frac{1}{m} \sum_{\ell=1}^{m} \hat{\varepsilon}_l^2\right) \tag{21}$$

are the sample autocovariances of $\hat{\varepsilon}_1^2, \ldots, \hat{\varepsilon}_m^2$ and

$$\omega_j(q(m)) = 1 - \frac{j}{q(m)+1}$$

are the weights. The maximum lag $q(m)$ is assumed to satisfy

$$q(m) \to \infty \quad \text{and} \quad \frac{q(m)}{\sqrt{m}} \to 0 \qquad (m \to \infty). \tag{22}$$

Under the conditions of Theorem 2, $\hat{\mu_m^2} \xrightarrow{\mathcal{P}} \mu^2$, so the assertion of Theorem 2 remains true if μ is replaced by $\hat{\mu}_m$.

Next, we turn our attention to investigating the behaviour of $\tau_Q(m)$ and $\tau_R(m)$ under the alternative hypothesis. In addition to assumption (10), we also need that the condition

$$\frac{1}{\nu - k^*} \sum_{j=m+k^*+1}^{m+\nu} \mathbf{x}_j \mathbf{x}_j^T \longrightarrow \mathbf{C} \qquad \text{a.s.} \qquad \text{as } \min\{\nu - k^*, m\} \to \infty \tag{23}$$

is satisfied, with $\mathbf{C}$ defined in (10). It turns out that the order of the difference $\hat{\tau}(m) - k^*$ depends on the location of the change-point. Let $\lfloor \cdot \rfloor$ denote the integer part.

Theorem 3 *Assume that (1), (2), (4), (8), (10), (11), (14) and (23) are satisfied,* $|\mathbf{c}_1^T(\boldsymbol{\beta}_0 - \boldsymbol{\beta}_*)| \neq 0$ *and that*

$$k^* = \lfloor \theta m^\beta \rfloor \qquad \textit{with some } \theta > 0,\ \beta \geq 0.$$

Under H_A, *as* $m \to \infty$:

(i) *If* $0 \leq \beta \leq (1-2\gamma)/(2-2\gamma)$, *then*

$$\tau_Q(m) - k^* = \mathcal{O}_P\left(m^{(1-2\gamma)/(2-2\gamma)}\right).$$

(ii) *If* $(1-2\gamma)/(2-2\gamma) \leq \beta < 1$, *then*

$$\tau_Q(m) - k^* = \mathcal{O}_P\left(m^{1/2-\gamma(1-\beta)}\right).$$

(iii) *If* $1 \leq \beta < \infty$, *then*

$$\tau_Q(m) - k^* = \mathcal{O}_P\left(k^* m^{-1/2}\right).$$

Theorem 4 *Suppose that the assumptions of Theorem 3 are satisfied. Then all conclusions of Theorem 3 hold with* $\tau_Q(m)$ *replaced by* $\tau_R(m)$.

Theorems 3 and 4 imply that the difference $\tau_R(m) - k^*$ has, in case (i), an asymptotic order which—depending on the choice of $\gamma \in [0, 1/2)$, but not on that of β—lies between $m^{1/2}$ and m. In case (ii), the order is bounded from below by $m^{1/2}$ and from above by m^β, while in case (iii), the order is at least $m^{1/2}$.

Let us assume that there are several changes in the parameters and the number of observations between changes is at least m. Our monitoring procedures will stop around the first change-point. We use the next m observations after $\tau_{(1)}$, the time of stopping, as the next training sample and start the monitoring procedure from the $(\tau_{(1)} + m)$th observation. If there is a change, we stop again at some time $\tau_{(2)}$. We then continue the monitoring from the $(\tau_{(2)} + m)$th observation. Thus, our procedure might also be used to detect multiple changes.

4. A SMALL SIMULATION STUDY

We illustrate the theory developed in the previous sections in case of the linear regression

$$y_i = \beta_{0i} + \beta_{1i} x_i + \varepsilon_i.$$

The discussion in this section is not intended to provide new insights into research in economics or finance, but merely to illustrate the statistical behaviour of the monitoring procedures and to facilitate their practical application by considering some special cases.

Denote by $\hat{\beta}_0(n)$ and $\hat{\beta}_1(n)$ the least-squares estimators of the intercept and the slope based on the observations (x_i, y_i), $i = 1, 2, \ldots, n$, and consider the residuals

$$\hat{\varepsilon}_i = y_i - \hat{\beta}_0(m) - \hat{\beta}_1(m) x_i, \quad i > 1,$$

the recursive residuals

$$\tilde{\varepsilon}_i = y_i - \hat{\beta}_0(i-1) - \hat{\beta}_1(i-1) x_i, \quad i > 1$$

and the prediction errors

$$\hat{u}_{\ell+i|\ell} = y_{\ell+i} - \hat{\beta}_0(\ell) - \hat{\beta}_1(\ell) x_{\ell+i}, \quad i \geq 1.$$

Table 1. Empirical size (in per cent) for the detectors D_γ and P_γ with $q = 1$; $m = 100$. The left-hand column in each panel reports rejections at level $\alpha = 0.10$, the right at level $\alpha = 0.05$.

	Detector D_γ						Detector $P_\gamma, q = 1$					
Horizon	$\gamma = 0.00$		$\gamma = 0.25$		$\gamma = 0.45$		$\gamma = 0.00$		$\gamma = 0.25$		$\gamma = 0.45$	
$2m$	1.1	0.4	3.7	1.5	5.7	2.7	1.7	0.8	4.2	2.1	7.3	5.3
$4m$	4.9	2.4	6.3	2.9	7.5	3.7	5.3	2.7	7.2	3.6	8.7	6.4
$6m$	7.0	3.5	7.3	3.3	8.2	4.0	7.6	4.0	8.5	4.2	9.2	6.6
$8m$	8.0	4.2	7.8	3.4	8.3	4.1	8.6	4.7	9.2	4.6	9.3	6.7

Since $\hat{u}_{j+1|j} = \tilde{\varepsilon}_{j+1}$, for $r = 1$, the detector $R(m, k)$ becomes

$$R(m,k) = \sum_{j=m+1}^{m+k-1} \hat{u}^2_{j+1|j} - \frac{k}{m}\sum_{j=2}^{m-1} \hat{u}^2_{j+1|j} = \sum_{i=m+2}^{m+k} \tilde{\varepsilon}_i^2 - \frac{k}{m}\sum_{i=3}^{m} \tilde{\varepsilon}_i^2.$$

Recall that

$$Q(m,k) = \sum_{i=m+1}^{m+k} \hat{\varepsilon}_i.$$

It is convenient to work with the normalized detectors

$$D_\gamma(m,k) = \left[m\hat{\sigma}_m^2\right]^{-1/2}\left(\frac{m}{m+k}\right)\left(\frac{m+k}{k}\right)^{\gamma} |Q(m,k)|$$

and

$$P_\gamma(m,k) = \left[m\hat{\mu}_m^2\right]^{-1/2}\left(\frac{m}{m+k}\right)\left(\frac{m+k}{k}\right)^{\gamma} |R(m,k)|,$$

with $\hat{\sigma}_m^2$ being the sample variance of the residuals $\hat{\varepsilon}_i$, $2 \le i \le m$ and $\hat{\mu}_m^2$ the Bartlett estimator defined in Section 3.

The null hypothesis is rejected at level α at the first time k such that $D_\gamma(m, k) > c$ or $P_\gamma(m, k) > c$, where the critical value c is determined by

$$P\left\{\sup_{0<t<1} t^{-\gamma}|W(t)| > c\right\} = \alpha.$$

A table with these critical values is given in Horváth *et al.* (2004).

As in Chu *et al.* (1996), all simulation results below are based on 2500 replications.

Table 1 reports empirical sizes of the procedures with $\beta_{0i} = 0$, $\beta_{1i} = 0$ and independent standard normal errors. This setting is similar to that considered in Chu *et al.* (1996), but both the slope and the intercept are estimated. [Chu *et al.* (1996) reported simulation results only for monitoring a possible change in mean of independent standard normal observations.] The size is well controlled and similar to that of the fluctuation detector of Chu et al. (1996). Methods with small γ are more conservative than methods with γ close to 1/2. We found that size does not depend on the values of β_{0i} and $\beta_{1i} = 0$, the differences are smaller than the chance error. Results for $m = 50$ and 200 are similar and are therefore not reported.

A much more telling insight is provided by a power study reported in Table 2. For $i \le m$, we set $\beta_{0i} = 0$, $\beta_{1i} = 0$. The changes in the regression parameters are such that the maximal power is

Table 2. Empirical power (in percent) for the detectors D_γ and P_γ with $q = 1$; $m = 100$; $k^* = 5$. The left column in each panel reports rejections at level $\alpha = 0.10$, the right at level $\alpha = 0.05$.

	Detector D_γ						Detector $P_\gamma, q = 1$					
	$\gamma = 0.00$		$\gamma = 0.25$		$\gamma = 0.45$		$\gamma = 0.00$		$\gamma = 0.25$		$\gamma = 0.45$	
Horizon	Change in slope from 0 to 1.4											
$2m$	3.1	1.0	8.60	4.00	13.0	7.3	87.1	80.6	92.5	89.1	92.9	90.5
$4m$	16.0	8.5	20.15	11.55	18.3	10.5	89.6	84.0	93.4	90.3	93.7	91.1
$6m$	17.9	10.3	21.25	12.35	18.6	10.8	89.6	84.0	93.4	90.3	93.7	91.1
$8m$	21.7	13.2	24.10	13.95	20.2	12.0	89.6	84.0	93.4	90.3	93.7	91.1
	Change in intercept from 0 to 0.5											
$2m$	77.8	62.7	85.2	75.9	85.3	77.8	7.3	4.7	10.3	6.4	17.4	12.9
$4m$	98.5	96.3	98.5	96.7	97.5	95.3	10.9	6.6	12.8	8.1	18.8	13.7
$6m$	99.4	98.6	99.3	98.3	98.7	97.3	12.2	7.1	13.7	8.8	19.6	13.9
$8m$	99.7	99.0	99.5	98.9	99.1	98.2	13.3	7.5	14.0	8.9	19.9	14.1

Table 3. Summary statistics for the regressions of excess returns on S&P 500 on earnings to price ratio.

	1954–1971			1971–1997		
	Value	Std. Error	P-value	Value	Std. Error	P-value
Intercept	−0.0284	0.0144	0.0500	0.0024	0.0072	0.7400
Slope	5.5337	2.2320	0.0140	0.4095	0.8651	0.6363

close to 100%, but is never 100%, otherwise the methods could not be compared. The detector D_γ is sensitive to changes in the intercept, but largely insensitive to changes in the slope. The picture is reversed for P_γ. The two detectors thus complement each other. Note that for the change in the slope, the power of P_γ is the same for monitoring horizons of $4m$, $6m$ and $8m$. The realizations of $P_\gamma(m, k)$, $k \geq 1$ look different, but they tend to approach the vicinity of the critical value for $k \leq 100$. After that time, they decrease, and never approach the critical value again. This is due to the *recursive* prediction errors used in P_γ, and this effect is visible to a lesser extent for the power when the intercept changes and for the size.

We will further illustrate the behaviour of the monitoring schemes by simulations based on a part of the data set compiled by McCracken (2000). Our goal is to come closer to a real world application, in which the assumptions of our theory are not perfectly satisfied. As will be seen, the size is no longer strictly controlled. The main reason why the size exceeds the nominal size for long monitoring horizon lies in the nature of the explanatory variables x_i which now form a time series with a complex dependence structure for which the convergence in (10) may be very slow and the stochastic structure of the x_i evolves with time as well. Conditionally heteroskedastic errors, whose marginal distribution has excess kurtosis, also increase the size, but to a lesser extent.

To motivate our simulations, we consider the dependent variables y_i which are the monthly excess returns of the S&P 500 index over the period from January 1954 to March 1997. The explanatory variables x_i are previous month's earnings to price ratios. The detailed description of these quantities is given in McCracken (2000).

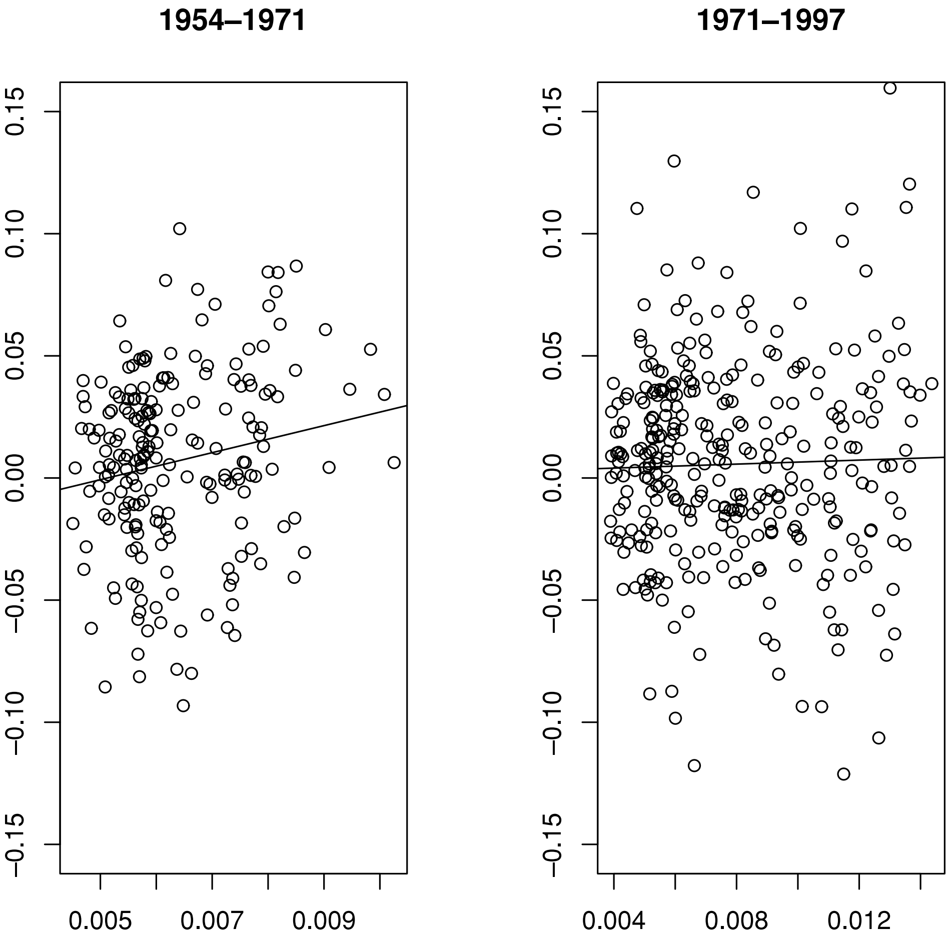

Figure 1. Excess returns on S&P 500 versus earnings to price ratio.

We split the 519 data points (x_i, y_i) into two subsets shown in Figure 1. The first subset contains the initial 200 observations and the second the remaining 319 observations. The break point corresponds to September 1971 and was so chosen because the parameters of the regression appear to change slightly after that date. In the initial 200 months they appear to be constant, so we work in the sequel with $m = 200$. Table 3 gives summary statistics for both time periods and suggests that before 1971 excess returns were positively correlated with the earnings to price ratio, whereas after 1971 the earnings to price ratio had no significant *linear* predictive ability.

We evaluate the performance of the monitoring schemes by reporting cumulative annual empirical rejection frequencies over the period of 10 years which corresponds to 120 observations. While this maximal monitoring period seems appropriate for the macroeconomic data considered in this section, it is somewhat arbitrary: in practice it would be dictated by the timespan of interest.

In our simulations, the errors were generated as realizations of a GARCH(1,1) model fitted to the residuals of the regression for the initial period. The parameters of this GARCH model are $\omega = 0.000144$, $\alpha = 0.095450$ and $\beta = 0.801718$. These parameters change only slightly

Table 4. Empirical size (in per cent) for the detectors D_γ and P_γ for the first 10 years of monitoring. The left-hand column in each panel reports rejections at level $\alpha = 0.10$, the right at level $\alpha = 0.05$.

	Detector D_γ					
Year	$\gamma = 0.00$		$\gamma = 0.25$		$\gamma = 0.45$	
1	0.00	0.00	0.36	0.08	2.93	1.87
2	0.00	0.00	0.60	0.24	4.26	2.93
3	0.04	0.00	0.84	0.28	5.00	3.40
4	0.08	0.00	1.36	0.64	5.60	3.67
5	0.24	0.00	3.08	1.56	7.60	4.53
6	0.76	0.16	4.48	2.60	9.33	5.80
7	1.84	0.52	7.24	4.28	12.60	8.13
8	4.88	2.36	12.28	7.92	17.07	12.00
9	10.40	5.92	19.24	13.16	21.67	16.80
10	17.48	12.08	26.08	20.04	29.00	22.60
	Detector $P_\gamma, q = 1$					
Year	$\gamma = 0.00$		$\gamma = 0.25$		$\gamma = 0.45$	
1	0.20	0.13	2.40	1.40	7.33	6.20
2	0.80	0.60	4.53	2.74	9.80	8.40
3	1.47	1.13	6.13	4.00	12.47	10.13
4	2.73	1.93	7.80	5.40	14.47	12.00
5	4.07	2.87	10.00	6.87	16.67	13.27
6	5.00	3.67	11.67	7.93	18.33	14.80
7	6.00	4.40	13.13	8.87	19.67	15.80
8	7.00	5.00	14.87	10.20	21.27	17.00
9	8.53	5.67	16.47	11.13	22.60	17.80
10	10.07	7.00	17.53	12.20	23.60	18.60

when the GARCH(1,1) model is fitted to the residuals over the remaining sample points, (to $\omega = 0.000117, \alpha = 0.076625, \beta = 0.863166$), so we used the same model for the errors before and after the onset of the monitoring, as required by our theory. The initial model corresponds to the estimated model for the period 1954–1971, i.e. $\beta_0 = -0.0284$ and $\beta_1 = 5.5337$. When studying the power, we consider two regression models after the change:

Model A: $\beta_0 = 0.0000$, $\beta_1 = 0.0000$;
Model B: $\beta_0 = 0.0216$, $\beta_1 = 5.5337$.

Model A represents a very small change in parameters and corresponds to the estimated model for the period 1971–1997 for which neither the intercept nor the slope is significantly different from zero. In Model B, the slope remains unchanged but the intercept increases by 0.05. Visually, this represents a larger separation of the scatter plots than in Figure 1. We performed simulations for $k^* = 3$ and 12, which correspond to a change-point after one quarter and after 1 year since the onset of the monitoring.

Table 4 shows empirical sizes for the detectors D_γ and P_γ. As in the case of a regression with i.i.d. errors, detectors with small γ are more conservative, while detectors with γ close to 1/2

Table 5. Empirical power (in per cent) of the detectors D_γ and P_γ for the first 10 years of monitoring. The left-hand column in each panel reports rejections at level $\alpha = 0.10$, the right at level $\alpha = 0.05$.

Detector D_γ, $\gamma = 0.45$

Year	Change to Model A				Change to Model B			
	After one quarter		After 1 year		After one quarter		After 1 year	
1	3.93	2.33	2.13	1.20	81.60	72.00	3.07	2.07
2	5.13	2.67	3.33	2.00	99.67	99.53	72.80	61.73
3	6.13	3.33	3.80	2.27	100.00	100.00	98.33	97.67
4	8.60	5.13	4.87	2.33	100.00	100.00	99.87	99.80
5	32.27	23.07	9.13	5.07	100.00	100.00	100.00	100.00
6	45.13	36.33	32.00	22.27	100.00	100.00	100.00	100.00
7	58.07	49.87	42.00	33.27	100.00	100.00	100.00	100.00
8	71.47	64.87	54.73	46.60	100.00	100.00	100.00	100.00
9	80.20	75.47	69.67	62.27	100.00	100.00	100.00	100.00
10	86.40	83.27	78.27	73.73	100.00	100.00	100.00	100.00

Detector P_γ, $\gamma = 0.45$, $q = 1$

Year	Change to Model A				Change to Model B			
	After one quarter		After 1 year		After one quarter		After 1 year	
1	8.20	6.07	6.87	5.80	49.33	43.93	9.40	7.87
2	10.53	8.27	9.80	8.33	72.27	66.60	41.80	37.13
3	13.53	11.00	12.67	10.93	82.27	78.20	63.93	58.20
4	16.73	13.40	15.13	12.53	89.00	85.53	74.60	69.80
5	19.47	16.00	18.07	14.87	90.07	87.67	83.07	79.13
6	20.87	17.37	20.67	16.33	90.80	88.40	84.80	81.20
7	22.40	18.67	22.20	17.60	91.07	88.60	85.80	82.73
8	23.80	19.73	23.13	18.80	91.13	88.80	86.13	83.00
9	25.80	21.00	24.20	19.40	91.13	88.80	86.20	83.27
10	26.67	21.87	25.40	20.33	91.13	88.80	86.40	83.33

have a higher rate of false rejections. In the long run, detectors P_γ are more conservative than D_γ and their rejection probabilities increase more gradually with the time of monitoring. Over the period of 10 years, none of the detectors considered in Table 4 has empirical size smaller than α, the detector P_0 comes very close. In a practical application, one would choose to work with a detector with a well-controlled size over the monitoring period of interest.

Table 5 shows that detector D_γ has higher empirical power than the detector P_γ, but even it needs close to 10 years of monitoring to detect a change to Model A, which cannot be effectively detected by detector P_γ. This is not surprising because this change is very small: if only one-thirds of the points were available in the right-hand panel of Figure 1, it would be practically impossible to tell by eye that the parameters of the regressions are different in the right-hand and left-hand panels of that figure. The change to Model B can be readily detected. The change at 3 months is easier to detect than a change at 12 months. Other simulations, not reported here to save space, show that these general properties hold for other choices of γ. In accordance with Table 4, the power is lower for smaller γ. For detector P_γ, it is higher by some 5% for $q = 0$ and decreases for larger q. The practical choice of the detector, including the choice of γ and q, will be dictated by the particular problem at hand, in particular by the time horizon of interest and the minimum size of the change to be detected.

5. PROOFS

Before proving Theorems 1–4, we establish an elementary lemma, which will be extensively used in the sequel. Recall that we work with a fixed realization of the $\mathbf{x}_i$'s.

Lemma 1 *If condition (10) is satisfied for some* $\kappa > 0$, *then*

$$\left| \left(\frac{1}{m} \sum_{i=1}^{m} \mathbf{x}_i \mathbf{x}_i^T \right)^{-1} - \mathbf{C}^{-1} \right| = \mathcal{O}(m^{-\kappa}) \qquad (m \to \infty), \tag{24}$$

$$\sum_{i=1}^{m} \frac{|\mathbf{x}_i|^2}{i} = \mathcal{O}(\log m) \qquad (m \to \infty), \tag{25}$$

and

$$\sum_{i=3}^{\infty} \frac{|\mathbf{x}_i|^2}{i(\log i)^{1+\varepsilon}} < \infty \qquad \textit{for all } \varepsilon > 0. \tag{26}$$

Proof. Horváth *et al.* (2004, lemma 1) proved that (10) implies (24). By Abel's summation formula and standard estimates, we get

$$\sum_{i=1}^{m} \frac{|\mathbf{x}_i|^2}{i} \le \sup_{1 \le k < \infty} \frac{1}{k} \sum_{i=1}^{k} |\mathbf{x}_i|^2 \left(\sum_{j=1}^{m} \frac{1}{j+1} + 1 \right),$$

implying (25), since the supremum on the right-hand side is finite and the latter sum is increasing with order $\log m$. Let $\varepsilon > 0$. Similarly, we also obtain, after an application of the mean-value theorem to the function $f(x) = x(\log x)^{1+\varepsilon}$,

$$\begin{aligned}
\sum_{i=3}^{m} \frac{|\mathbf{x}_i|^2}{i(\log i)^{1+\varepsilon}} &= \sum_{i=3}^{m} \left(\frac{1}{i(\log i)^{1+\varepsilon}} - \frac{1}{(i+1)[\log(i+1)]^{1+\varepsilon}} \right) \sum_{\ell=3}^{i} |\mathbf{x}_\ell|^2 \\
&\quad + \frac{1}{(m+1)[\log(m+1)]^{1+\varepsilon}} \sum_{i=3}^{m} |\mathbf{x}_i|^2 \\
&\le \sum_{i=3}^{m} \frac{1+\varepsilon}{i(\log i)^{1+\varepsilon}} \left(\frac{1}{i} \sum_{\ell=1}^{i} |\mathbf{x}_\ell|^2 \right) + \frac{1}{m} \sum_{i=1}^{m} |\mathbf{x}_i|^2 \\
&\le \sup_{1 \le k < \infty} \frac{1}{k} \sum_{i=1}^{k} |\mathbf{x}_i|^2 \left(\sum_{j=3}^{\infty} \frac{1+\varepsilon}{j(\log j)^{1+\varepsilon}} + 1 \right) < \infty,
\end{aligned}$$

since both the supremum and the sum are finite. Thus the proof is complete. □

5.1. Proof of Theorem 1

The proof of Theorem 1 consists of two lemmas stated and proved in this section.

Lemma 2 *If the conditions of Theorem 1 are satisfied, then, as* $m \to \infty$,

$$\sup_{1\le k<\infty}\frac{1}{g(m,k)}\left|\sum_{i=m+1}^{m+k}\hat{\varepsilon}_i-\left(\sum_{i=m+1}^{m+k}\varepsilon_i-\frac{k}{m}\sum_{i=1}^{m}\varepsilon_i\right)\right|=o_P(1).$$

Proof: Let $\mathbf{C}_m=\frac{1}{m}\sum_{i=1}^{m}\mathbf{x}_i\mathbf{x}_i^T$. Since $\hat{\boldsymbol{\beta}}_m-\boldsymbol{\beta}_0=\mathbf{C}_m^{-1}\frac{1}{m}\sum_{j=1}^{m}\mathbf{x}_j\varepsilon_j$, we have that

$$\sum_{i=m+1}^{m+k}\hat{\varepsilon}_i=\sum_{i=m+1}^{m+k}\varepsilon_i-\left(\sum_{i=m+1}^{m+k}\mathbf{x}_i\right)^T\mathbf{C}_m^{-1}\frac{1}{m}\sum_{j=1}^{m}\mathbf{x}_j\varepsilon_j.$$

Assumption (8) yields that $E\,\mathbf{x}_j\varepsilon_j=0$ and Var $(x_{j,i}\,\varepsilon_j)\le C|\mathbf{x}_i|^2$, so, by Chebyshev's inequality,

$$\left|\sum_{j=1}^{m}\mathbf{x}_j\varepsilon_j\right|=\mathcal{O}_P\left(\sqrt{m}\right)\qquad(m\to\infty).\tag{27}$$

Lemma 5.1 of Horváth *et al.* (2004) and (10) yield that there are random variables k_0 and m_0 such that, for all $k\ge 1$,

$$\left|\sum_{i=m+1}^{m+k}\mathbf{x}_i-k\mathbf{c}_1\right|\le k_0(m^{1-\kappa}+(m+k)^{1-\kappa})\qquad\text{if } m\ge m_0,\tag{28}$$

where $\mathbf{c}_1$ is the first column of $\mathbf{C}$. Putting together equations (24), (27) and (28), we conclude that

$$\begin{aligned}&\sup_{1\le k<\infty}\frac{1}{g(m,k)}\left|\left(\frac{1}{m}\left(\sum_{i=m+1}^{m+k}\mathbf{x}_i\right)^T\mathbf{C}_m^{-1}-\frac{k}{m}\mathbf{c}_1^T\mathbf{C}^{-1}\right)\sum_{j=1}^{m}\mathbf{x}_j\varepsilon_j\right|\\&=\mathcal{O}_P\left(\sqrt{m}\right)\sup_{1\le k<\infty}\frac{1}{g(m,k)}\left(\frac{k}{m^{1+\kappa}}+\frac{(m+k)^{1-\kappa}}{m}+m^{-\kappa}\right).\end{aligned}$$

Since $\gamma<\kappa$, we get that

$$\begin{aligned}&\max_{1\le k\le m}\frac{km^{-1-\kappa}+(1+k/m)m^{-\kappa}}{(1+k/m)[(k/m)/(1+k/m)]^{\gamma}}\\&\le\max_{1\le k\le m}2^{\gamma}\left(\left(\frac{k}{m}\right)^{1-\gamma}m^{-\kappa}+\left(\frac{k}{m}\right)^{-\gamma}m^{-\kappa}\right)=\max_{1\le k\le m}2^{\gamma}(m^{-\kappa}+m^{\gamma-\kappa})=o(1),\end{aligned}$$

as $m\to\infty$. Also,

$$\sup_{m\le k<\infty}\frac{km^{-1-\kappa}+(1+k/m)m^{-\kappa}}{(1+k/m)[(k/m)/(1+k/m)]^{\gamma}}=o(1),$$

thus completing the proof of

$$\sup_{1\le k<\infty}\frac{1}{g(m,k)}\left|\left(\frac{1}{m}\left(\sum_{i=m+1}^{m+k}\mathbf{x}_i\right)^T\mathbf{C}_m^{-1}-\frac{k}{m}\mathbf{c}_1^T\mathbf{C}^{-1}\right)\sum_{j=1}^{m}\mathbf{x}_j\varepsilon_j\right|=o_P(1). \tag{29}$$

Since $\mathbf{c}_1^T\ \mathbf{C}^{-1}=(1,0,\ldots,0)$, the lemma follows from (29).

Lemma 3 *If the conditions of Theorem 1 are satisfied, then there are two independent Wiener processes* $\{W_{1,m}(t):t\ge 0\}$ *and* $\{W_{2,m}(t):t\ge 0\}$ *such that, as* $m\to\infty$,

$$\sup_{1\le k<\infty}\frac{1}{g(m,k)}\left|\sum_{i=m+1}^{m+k}\varepsilon_i-\frac{k}{m}\sum_{i=1}^{m}\varepsilon_i-\sigma\left(W_{1,m}(k)-\frac{k}{m}W_{2,m}(m)\right)\right|=o_P(1).$$

Proof. Using assumptions (15) and (16), we have

$$\begin{aligned}&\sup_{1\le k<\infty}\frac{1}{g(m,k)}\left|\sum_{i=m+1}^{m+k}\varepsilon_i-\frac{k}{m}\sum_{i=1}^{m}\varepsilon_i-\sigma\left(W_{1,m}(k)-\frac{k}{m}W_{2,m}(m)\right)\right|\\&=\mathcal{O}_P\left(\sup_{1\le k<\infty}\left(k^{\Delta}+\frac{k}{m}m^{\Delta}\right)\left[\sqrt{m}\left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^{\gamma}\right]^{-1}\right).\end{aligned}$$

Clearly,

$$\begin{aligned}\sup\nolimits_{1\le k\le m}\frac{k^{\Delta}+km^{\Delta-1}}{\sqrt{m}\left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^{\gamma}}&=\mathcal{O}\left(m^{\gamma-1/2}\max_{1\le k\le m}k^{\Delta-\gamma}+m^{\Delta-1/2}\right)\\&=\mathcal{O}\left(m^{\gamma-1/2}\max_{1\le k\le m}k^{\Delta-\gamma}\right),\end{aligned}$$

since $\gamma<1/2$. Also,

$$m^{\gamma-1/2}\max_{1\le k\le m}k^{\Delta-\gamma}=\left\{\begin{matrix}m^{\gamma-1/2}m^{\Delta-\gamma} & \text{if }\Delta\ge\gamma\\ m^{\gamma-1/2} & \text{if }\Delta<\gamma\end{matrix}\right\}=o(1).$$

Since $\Delta<1/2$, we get similarly

$$\begin{aligned}&\sup_{m<k<\infty}\frac{k^{\Delta}+km^{\Delta-1}}{\sqrt{m}\left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^{\gamma}}\\&\le\sup_{m<k<\infty}\frac{\sqrt{m}k^{\Delta-\gamma}}{(m+k)^{1-\gamma}}+\sup_{m<k<\infty}\frac{k^{1-\gamma}m^{\Delta-1/2}}{(m+k)^{1-\gamma}}\\&=\sup_{m<k<\infty}\sqrt{m}\left(\frac{m+k}{k}\right)^{\gamma-1}k^{1-\Delta}+\sup_{m<k<\infty}m^{\Delta-1/2}\left(\frac{k}{m+k}\right)^{1-\gamma}=o(1),\end{aligned}$$

completing the proof of the lemma. □

Proof of Theorem 1. Putting together Lemmas 2 and 3, it is enough to show that

$$\sup_{1\le k<\infty}\frac{1}{g(m,k)}\left|W_{1,m}(k)-\frac{k}{m}W_{2,m}(m)\right|\xrightarrow{\mathcal{D}}\sup_{0<t<1}\frac{|W(t)|}{t^{\gamma}}.$$

This is established in the proof of theorem 2.1 of Horváth *et al.* (2004).

5.2. *Proof of Theorem 2*

We will use the notation $\mathbf{C}_k = \sum_{j=1}^{k} \mathbf{x}_j \mathbf{x}_j^T$. Elementary calculations, involving the application of (13), show that

$$y_{\ell+i} - \hat{y}_{\ell+i} = \varepsilon_{\ell+i} - \mathbf{x}_{\ell+i}^T \mathbf{C}_\ell^{-1} \sum_{j=1}^{\ell} \mathbf{x}_j \varepsilon_j, \tag{30}$$

and, therefore, for any $N \geq 0$,

$$\begin{aligned} &\sum_{i=1}^{r} \sum_{\nu=1}^{k-i} (y_{N+\nu+i} - \hat{y}_{N+\nu+i})^2 \\ &= \sum_{i=1}^{r} \sum_{\nu=1}^{k-i} \varepsilon_{N+\nu+i}^2 - 2 \sum_{i=1}^{r} \sum_{\nu=1}^{k-i} \varepsilon_{N+\nu+i} \mathbf{x}_{N+\nu+i}^T \mathbf{C}_{N+\nu}^{-1} \sum_{j=1}^{N+\nu} \mathbf{x}_j \varepsilon_j \\ &\quad + \sum_{i=1}^{r} \sum_{\nu=1}^{k-i} \sum_{j=1}^{N+\nu} \mathbf{x}_j^T \varepsilon_j \mathbf{C}_{N+\nu}^{-1} \mathbf{x}_{N+\nu+i} \mathbf{x}_{N+\nu+i}^T \mathbf{C}_{N+\nu}^{-1} \sum_{\ell=1}^{N+\nu} \mathbf{x}_\ell \varepsilon_\ell \\ &= A_{k,1}(N) + A_{k,2}(N) + A_{k,3}(N). \end{aligned} \tag{31}$$

In the following lemmas, we show that appropriately normalized functionals of the second and third terms in (31) are negligible. We then work with the first term to establish the required limit in Theorem 2.

Lemma 4 *Let the conditions of Theorem 2 be satisfied. Then, as $m \to \infty$,*

$$\sup_{1 \leq k < \infty} \frac{|A_{k,3}(m)|}{g(m,k)} = o_P(1).$$

Proof. Straightforward estimation gives that

$$|A_{k,3}(m)| \leq \sum_{i=1}^{r} \sum_{\nu=1}^{k-i} \left| \sum_{j=1}^{m+\nu} \mathbf{x}_j \varepsilon_j \right|^2 \left| \mathbf{C}_{m+\nu}^{-1} \right|^2 |\mathbf{x}_{m+\nu+i}|^2 .$$

A combined application of condition (10) and (24) yields

$$\max_{1 \leq \nu < \infty} (m+\nu) \left| \mathbf{C}_{m+\nu}^{-1} \right| = \mathcal{O}(1) \qquad \text{a.s.} \qquad (m \to \infty). \tag{32}$$

Let $\mathbf{x}_j^T = (x_{j,1}, \ldots, x_{j,p})$. It is clear that, for any $1 \leq i \leq p$, $\{\sum_{j=1}^{k} \mathbf{x}_{j,i} \varepsilon_j, \mathcal{G}_k\}$ is a zero mean square integrable martingale and $E[(x_{j,i}\varepsilon_j)^2 \mid \mathcal{G}_{j-1}] = x_{j,i}^2 E[\varepsilon_j^2 | \mathcal{G}_{j-1}]$. Therefore,

$$\sum_{j=3}^{\infty} \frac{1}{j(\log j)^{1+2\delta}} E\left[(x_{j,i}\varepsilon_j)^2 \mid \mathcal{G}_{j-1}\right] \sum_{j=3}^{\infty} \frac{1}{j(\log j)^{1+2\delta}} x_{j,i}^2 E\left[\varepsilon_j^2 \mid \mathcal{G}_{j-1}\right] \qquad \text{a.s.}$$

Using assumption (17) and (26), we get that

$$E\left(\sum_{j=3}^{\infty} \frac{1}{j(\log j)^{1+2\delta}} x_{j,i}^2 E\left[\varepsilon_j^2 \mid \mathcal{G}_{j-1}\right] \right) \leq C \sum_{j=3}^{\infty} \frac{1}{j(\log j)^{1+2\delta}} x_{j,i}^2 < \infty.$$

Thus we have

$$\sum_{j=3}^{\infty} \frac{1}{j(\log j)^{1+2\delta}} E\left[(x_{j,i}\varepsilon_j)^2 \,\middle|\, \mathcal{G}_{j-1}\right] < \infty \qquad \text{a.s.}$$

Using exercise 7.4.10 in Chow and Teicher (1988, p. 249), we get that for all $\delta, \delta^* > 0$ there is a random variable m_0 such that

$$\left|\sum_{j=1}^{m+\nu} \mathbf{x}_j \varepsilon_j\right| \le \delta^* \left(\sqrt{m+\nu}[\log(m+\nu)]^{1/2+\delta}\right) \qquad \text{for all } \nu \ge 1 \text{ if } m \ge m_0. \tag{33}$$

So by (32) and (33), we can find a random variable ξ such that

$$|A_{k,3}(m)| \le \delta^* \xi \sum_{i=1}^{r} \sum_{\nu=1}^{k-i} |\mathbf{x}_{m+\nu+i}|^2 \frac{[\log(m+\nu)]^{1+2\delta}}{m+\nu}.$$

By Abel's summation formula, we get that, for any $1 \le i \le r$,

$$\sum_{j=m+i+1}^{m+k} |\mathbf{x}_j|^2 \frac{[\log j]^{1+2\delta}}{j} \le A \left(\sup_{1 \le j < \infty} \frac{1}{j} \sum_{\ell=1}^{j} |\mathbf{x}_\ell|^2 \right) \left([\log(m+k)]^{1+2\delta} \log\left(\frac{m+k}{m}\right)\right)$$

with some positive constant A, since

$$\begin{aligned} \sum_{j=m+i+1}^{m+k} \frac{[\log j]^{1+2\delta}}{j} &\le [\log(m+k)]^{1+2\delta} \sum_{j=m+i+1}^{m+k} \frac{1}{j} \\ &\le C[\log(m+k)]^{1+2\delta}[\log(m+k) - \log m]. \end{aligned}$$

Hence, there are random variables ξ^* and m^* such that

$$\sup_{1 \le k < \infty} \frac{|A_{k,3}(m)|}{g(m,k)} \le \delta^* \xi^* \sup_{1 \le k < \infty} \frac{[\log(m+k)]^{1+2\delta} \log\left(\frac{m+k}{m}\right)}{\sqrt{m}\left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^{\gamma}}$$

if $m \ge m^*$. It is clear that

$$\sup_{1 \le k < \infty} \frac{[\log(m+k)]^{1+2\delta} \log\left(\frac{m+k}{m}\right)}{\sqrt{m}\left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^{\gamma}} \longrightarrow 0 \qquad (m \to \infty),$$

completing the proof. □

Lemma 5 *Let the conditions of Theorem 2 be satisfied. Then, as $m \to \infty$,*

$$\sup_{1 \le k < \infty} \frac{k}{m} \frac{|A_{m,3}(0)|}{g(m,k)} = o_P(1).$$

Proof. Define the filtration $\{\mathcal{F}_k\}$ by letting $\mathcal{F}_k = \mathcal{G}_{m+k}$. First, note that

$$|A_{m,3}(0)| \le \sum_{i=1}^{r} \sum_{\nu=1}^{m-i} \left|\sum_{j=1}^{\nu} \mathbf{x}_j \varepsilon_j\right|^2 \left|\mathbf{C}_\nu^{-1}\right|^2 |\mathbf{x}_{\nu+i}|^2.$$

Since $\{\sum_{j=1}^{k} \mathbf{x}_j \varepsilon_j\}$ is a mean zero, square integrable martingale with respect to $\{\mathcal{F}_k\}$, via the Hájek–Rényi inequality [see Chow and Teicher (1988, p. 247)] we obtain that, for any $\lambda > 0$,

$$\max_{1 \le \nu \le m} \frac{1}{\nu^{\lambda+1/2}} \left| \sum_{j=1}^{\nu} \mathbf{x}_j \varepsilon_j \right| = \mathcal{O}_P(1) \qquad (m \to \infty).$$

Also, by assumption (10) and Lemma 1, we get $|\mathbf{C}_\nu^{-1}| = \mathcal{O}(1/\nu)$ as $\nu \to \infty$. Consequently, we conclude

$$|A_{m,3}(0)| = \mathcal{O}_P(1) \sum_{i=1}^{r} \sum_{\nu=1}^{m-i} \nu^{2\lambda-1} |\mathbf{x}_{\nu+1}|^2 \qquad (m \to \infty).$$

By (26) we have

$$\sum_{\nu=1}^{m} \nu^{2\lambda-1} |\mathbf{x}_\nu|^2 = \mathcal{O}_P(m^{2\lambda}) \qquad (m \to \infty)$$

and thus we arrive at

$$\sup_{1 \le k < \infty} \frac{k}{m} \frac{|A_{m,3}(0)|}{\sqrt{m}\left(1 + \frac{k}{m}\right)\left(\frac{k}{m+k}\right)^\gamma} = \mathcal{O}_P(1) \sup_{1 \le k < \infty} \left(\frac{k}{m+k}\right)^{1-\gamma} m^{2\lambda - 1/2} = o_P(1)$$

as $m \to \infty$ on choosing λ appropriately small, hence completing the proof. □

Lemma 6 *Let the conditions of Theorem 2 be satisfied. Then, as $m \to \infty$,*

$$\sup_{1 \le k < \infty} \frac{|A_{k,2}(m)|}{g(m,k)} = o_P(1).$$

Proof. Let $1 \le i \le r$ and, for $1 \le k < \infty$, set $D_k = D_k(i) = \sum_{\nu=1}^{k-i} \eta_\nu(i)$, where $\eta_\nu(i) = \varepsilon_{m+\nu+i} \mathbf{x}_{m+\nu+i}^T \mathbf{C}_{m+\nu}^{-1} \sum_{j=1}^{m+\nu} \mathbf{x}_j \varepsilon_j$. Recall that $\mathcal{F}_k = \mathcal{G}_{m+k}$. Then

$$\begin{aligned} E[D_k | \mathcal{F}_{k-1}] &= E\left[\sum_{\nu=1}^{k-i} \varepsilon_{m+\nu+i} \mathbf{x}_{m+\nu+i}^T \mathbf{C}_{m+\nu}^{-1} \sum_{j=1}^{m+\nu} \mathbf{x}_j \varepsilon_j \,\middle|\, \mathcal{F}_{k-1} \right] \\ &= \sum_{\nu=1}^{k-i-1} \varepsilon_{m+\nu+i} \mathbf{x}_{m+\nu+i}^T \mathbf{C}_{m+\nu}^{-1} \sum_{j=1}^{m+\nu} \mathbf{x}_j \varepsilon_j \\ &\quad + E[\varepsilon_{m+k} | \mathcal{F}_{k-1}] \mathbf{x}_{m+k}^T \mathbf{C}_{m+k-i}^T \sum_{j=1}^{m+k-i} \mathbf{x}_j \varepsilon_j \\ &= D_{k-1} \qquad \text{a.s.,} \end{aligned}$$

since $E[\varepsilon_{m+k} | \mathcal{F}_{k-1}] = 0$ a.s. by assumption (17). Hence $\{D_k, \mathcal{F}_k\}$ is a martingale. So by the Hájek–Rényi inequality [see Chow and Teicher (1988, p. 247)] we have, for any $\lambda > 0$,

$$P\left\{ \max_{1 \le k < \infty} \frac{|D_k|}{\sqrt{m}\left(1 + \frac{k}{m}\right)\left(\frac{k}{m+k}\right)^\gamma} \ge \lambda \right\} \le \frac{1}{\lambda^2} \sum_{k=1}^{\infty} \frac{E\eta_k^2(i)}{m\left(1 + \frac{k}{m}\right)^2 \left(\frac{k}{m+k}\right)^{2\gamma}}.$$

Moreover, by (24),

$$E\eta_k^2(i) = \sigma^4 \mathbf{x}_{m+k+i}^T \mathbf{C}_{m+k}^{-1} \mathbf{x}_{m+k+i} \le A \frac{|\mathbf{x}_{m+k+i}|^2}{m+k}$$

with some positive constant A. Hence, for all $0 < \delta < 1 - 2\gamma$,

$$\begin{aligned}
&\sum_{k=1}^{\infty} \frac{E\eta_k^2(i)}{m\left(\frac{m+k}{m}\right)^2 \left(\frac{k}{m+k}\right)^{2\gamma}} \\
&\le Am \sum_{k=1}^{\infty} \frac{|\mathbf{x}_{m+k+i}|^2}{(m+k)^3} \left(\frac{m+k}{k}\right)^{2\gamma} \\
&\le A \sum_{\ell=1}^{\infty} \frac{|\mathbf{x}_{\ell+i}|^2}{\ell^{1+\delta}} \max_{1\le k<\infty} \frac{m}{(m+k)^{2-2\gamma-\delta} k^{2\gamma}} \\
&\le A m^{-1+2\gamma+\delta} \sum_{\ell=1}^{\infty} \frac{|\mathbf{x}_{\ell+i}|^2}{\ell^{1+\delta}} = o(1)
\end{aligned}$$

as $m \to \infty$, since the latter sum is finite according to (26). Choosing λ arbitrarily small, the proof is complete on recognizing that $A_{k,2}(m) = \sum_{i=1}^{r} D_k(i)$. □

Lemma 7 *Let the conditions of Theorem 2 be satisfied. Then, as $m \to \infty$,*

$$\sup_{1\le k<\infty} \frac{k}{m} \frac{|A_{m,2}(0)|}{g(m,k)} = o_P(1).$$

Proof. Observe that, by the orthogonality of the $\mathbf{x}_j \varepsilon_j$ and (25), we have

$$\operatorname{Var}\left(\sum_{\nu=1}^{m-i} \varepsilon_{\nu+i} \mathbf{x}_{\nu+i}^T \mathbf{C}_\nu^{-1} \sum_{j=1}^{\nu} \mathbf{x}_j \varepsilon_j\right) = \sigma^4 \sum_{\nu=1}^{m-i} \mathbf{x}_{\nu+i}^T \mathbf{C}_\nu^{-1} \mathbf{x}_{\nu+i} \le A\sigma^4 \sum_{\nu=1}^{m-i} \frac{|\mathbf{x}_{\nu+i}|^2}{\nu} = \mathcal{O}(\log m)$$

as $m \to \infty$ by (25) with some positive constant A. Thus we have, after applying Chebyshev's inequality,

$$\sup_{1\le k<\infty} \frac{k}{m} \frac{|A_{m,2}(0)|}{\sqrt{m}\left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^{\gamma}} = \mathcal{O}_P(1) \sqrt{\frac{\log m}{m}} \sup_{1\le k<\infty} \left(\frac{k}{m+k}\right)^{1-\gamma} = o_P(1)$$

as $m \to \infty$, finishing the proof.

Proof of Theorem 2. It follows from Lemmas 4–7 that it is enough to prove that

$$\frac{1}{r\mu} \sup_{1\le k<\infty} \frac{1}{g(m,k)} \left| \sum_{i=1}^{r} \sum_{\nu=1}^{k-i} \left(\varepsilon_{m+\nu+i}^2 - \sigma^2\right) - \frac{k}{m} \sum_{i=1}^{r} \sum_{\nu=1}^{m-i} \left(\varepsilon_{\nu+i}^2 - \sigma^2\right) \right| \xrightarrow{\mathcal{D}} \sup_{0<t\le 1} \frac{|W(t)|}{t^\gamma} \qquad (m \to \infty). \tag{34}$$

We note that

$$\sum_{i=1}^{r} \sum_{\nu=1}^{k-i} \left(\varepsilon_{m+\nu+i}^2 - \sigma^2\right) = r \sum_{\nu=1}^{k} \left(\varepsilon_{m+\nu}^2 - \sigma^2\right) - \sum_{i=0}^{r-1} (r-i) \left(\varepsilon_{m+1+i}^2 - \sigma^2\right).$$

Because the number of terms in the second sum on the right-hand side of the latter equation does not depend on m and k, it is clear that

$$\sup_{1\le k<\infty}\frac{1}{g(m,k)}\left|\left[\sum_{i=1}^{r}\sum_{\nu=1}^{k-i}\left(\varepsilon_{m+\nu+i}^2-\sigma^2\right)-\frac{k}{m}\sum_{i=1}^{r}\sum_{\nu=1}^{m-i}\left(\varepsilon_{\nu+i}^2-\sigma^2\right)\right]\right.$$

$$\left.-r\left[\sum_{\nu=1}^{k}\left(\varepsilon_{m+\nu}^2-\sigma^2\right)-\frac{k}{m}\sum_{\nu=1}^{m}\left(\varepsilon_{\nu}^2-\sigma^2\right)\right]\right|=o_P(1)\qquad(m\to\infty).\tag{35}$$

Using the uniform weak invariance principle in (19), we have that

$$\max_{1\le k\le 4m}\frac{1}{\sqrt{m}(k/m)^{\gamma}}\left|\sum_{i=m+1}^{m+k}\left(\varepsilon_i^2-\sigma^2\right)-\mu W_{3,m}(k)\right|=\mathcal{O}_P(1)\max_{1\le k\le 4m}\frac{k^{\Delta}}{\sqrt{m}(k/m)^{\gamma}}$$

and

$$\max_{1\le k\le 4m}\frac{k^{\Delta}}{\sqrt{m}(k/m)^{\gamma}}=\left\{\begin{array}{ll} m^{\gamma-1/2}\max\limits_{1\le k\le 4m}k^{\Delta-\gamma} & \text{if } \gamma\ge\Delta\\ m^{\Delta-1/2}\max\limits_{1\le k\le 4m}\left(\dfrac{k}{m}\right)^{\Delta-\gamma} & \text{if } \gamma<\Delta\end{array}\right\}=o(1),$$

as $m\to\infty$. Hence,

$$\max_{1\le k\le 4m}\frac{1}{\sqrt{m}(k/m)^{\gamma}}\left|\sum_{i=m+1}^{m+k}\left(\varepsilon_i^2-\sigma^2\right)-\mu W_{3,m}(k)\right|=o_P(1)\qquad(m\to\infty).\tag{36}$$

Similarly, (19) also yields

$$\sup_{4m\le k<\infty}\frac{1}{\sqrt{m}(1+k/m)}\left|\sum_{i=m+1}^{m+k}\left(\varepsilon_i^2-\sigma^2\right)-\mu W_{3,m}(k)\right|$$

$$=\mathcal{O}_P(1)\sup_{4m\le k<\infty}\sqrt{m}k^{\Delta-1}=\mathcal{O}_P(1)m^{\Delta-1/2}=o_P(1)\tag{37}$$

as $m\to\infty$, since $\Delta<1/2$. On the other hand, the weak approximation in (20) implies that

$$\sup_{1\le k<\infty}\frac{1}{g(m,k)}\frac{k}{m}\left|\sum_{i=1}^{m}\left(\varepsilon_i^2-\sigma^2\right)-\mu W_{4,m}(m)\right|=o_P(1)\qquad(m\to\infty).\tag{38}$$

On combining equations (35)–(38), we arrive at

$$\sup_{1\le k<\infty}\frac{1}{g(m,k)}\left|\sum_{i=1}^{r}\sum_{\nu=1}^{k-i}\left(\varepsilon_{m+\nu+i}^2-\sigma^2\right)-\frac{k}{m}\sum_{i=1}^{r}\sum_{\nu=1}^{m-i}\left(\varepsilon_{\nu+i}^2-\sigma^2\right)\right|$$

$$-\sup_{1\le k<\infty}\frac{r\mu}{g(m,k)}\left|W_{3,m}(k)-\frac{k}{m}W_{4,m}(m)\right|=o_P(1)\qquad(m\to\infty).$$

Horváth *et al.* (2004, pp. 239–240) showed that for all m,

$$\sup_{1\le k<\infty}\frac{|W_{3,m}(k)-k/mW_{4,m}(m)|}{\sqrt{m}\left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^{\gamma}}\xrightarrow{\mathcal{D}}\sup_{0<t<1}\frac{|W(t)|}{t^{\gamma}}$$

and so Theorem 2 is proved.

5.3. Proofs of Theorems 3 and 4

Proof of Theorem 3. It follows from the definition of $\hat{\varepsilon}_i$ that, for all $k \geq k^*$,

$$\begin{aligned}\sum_{i=m+1}^{m+k} \hat{\varepsilon}_i &= \sum_{i=m+1}^{m+k} \varepsilon_i + \left(\sum_{i=m+1}^{m+k^*} \mathbf{x}_i\right)^T (\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}_m) + \left(\sum_{i=m+k^*}^{m+k} \mathbf{x}_i\right)^T (\boldsymbol{\beta}_* - \hat{\boldsymbol{\beta}}_m) \\ &= \sum_{i=m+1}^{m+k} \varepsilon_i + \left(\sum_{i=m+1}^{m+k} \mathbf{x}_i\right)^T (\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}_m) + \left(\sum_{i=m+k^*}^{m+k} \mathbf{x}_i\right)^T (\boldsymbol{\beta}_* - \boldsymbol{\beta}_0).\end{aligned}$$

By Theorem 1, we have

$$\sup_{m<k<\infty} \frac{1}{g(m,k)} \left| \sum_{i=m+1}^{m+k} \varepsilon_i + \left(\sum_{i=m+1}^{m+k} \mathbf{x}_i\right)^T \left(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}_m\right) \right| = \mathcal{O}_P(1).$$

Condition (14) yields $\sum_{i=m+k^*}^{m+k} \mathbf{x}_i = (k-k^*)\mathbf{c}_1 + o(k-k^*)$ uniformly in m. Thus we have, as $m \to \infty$,

$$\sup_{1\leq k\leq N} \frac{|Q(m,k)|}{g(m,k)} = \mathcal{O}_P(1) + \mathcal{O}(1) \sup_{k^*<k\leq N} \frac{(k-k^*)(1+o(1))}{\sqrt{m}\left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^{\gamma}},$$

where the bounds in $\mathcal{O}_P(1)$ and $\mathcal{O}(1)$ do not depend on N. We need to choose N so large that the supremum on the right-hand side of the last equation is larger than a given constant. Elementary arguments show that choosing N as

$$N - k^* = \begin{cases} Cm^{(1-2\gamma)/(2-2\gamma)} & 0 \leq \beta \leq (1-2\gamma)/(2-2\gamma) \\ Cm^{1/2-\gamma(1-\beta)} & (1-2\gamma)/(2-2\gamma) < \beta \leq 1 \\ Ck^* m^{-1/2} & \beta > 1 \end{cases}$$

we get

$$\lim_{m\to\infty} \sup_{k^*<k\leq N} \frac{k-k^*}{\sqrt{m}\left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^{\gamma}} \geq L(C)$$

and $\lim L(C) = \infty$ as $C \to \infty$. □

The proof of Theorem 4 is more complex and uses a number of technical lemmas.

Similarly to the decomposition (30), which holds true under the null hypothesis H_0, we obtain under H_A, for all $k > k^*$ and $1 \leq i \leq r$,

$$\begin{aligned} y_{m+k+i} - \hat{y}_{m+k+i} &= \varepsilon_{m+k+i} - \mathbf{x}_{m+k+i}^T \mathbf{C}_{m+k}^{-1} \sum_{j=1}^{m+k} \mathbf{x}_j \varepsilon_j \\ &\quad + \mathbf{x}_{m+k+i}^T \Delta - \mathbf{x}_{m+k+i}^T \left(\mathbf{I} - \mathbf{C}_{m+k}^{-1}\mathbf{C}_{m+k^*}\right)\Delta \\ &= \varepsilon_{m+k+i} - \mathbf{x}_{m+k+i}^T \mathbf{C}_{m+k}^{-1} \sum_{j=1}^{m+k} \mathbf{x}_j \varepsilon_j + \mathbf{x}_{m+k+i}^T \mathbf{C}_{m+k}^{-1}\mathbf{C}_{m+k^*}\Delta, \qquad (39)\end{aligned}$$

where $\boldsymbol{\Delta} = \boldsymbol{\beta}_* - \boldsymbol{\beta}_0$ and $\mathbf{I}$ denotes the $p \times p$ identity matrix. We have used that $\boldsymbol{\beta}_i = \boldsymbol{\beta}_0$ if $i \leq m + k^*$ and $\boldsymbol{\beta}_i = \boldsymbol{\beta}_*$ if $i > m + k*$. Introducing the notations

$$\alpha_{m+\nu+i} = \mathbf{x}_{m+\nu+i}^T \mathbf{C}_{m+\nu}^{-1} \mathbf{C}_{m+k^*} \Delta \qquad \text{and} \qquad \eta_{m+\nu} = \mathbf{x}_{m+\nu+1}^T \mathbf{C}_{m+\nu}^{-1} \sum_{j=1}^{m+\nu} \mathbf{x}_j \varepsilon_j,$$

equation (39) yields

$$\begin{aligned}
&\sum_{i=1}^{r} \sum_{\nu=1}^{k-i} (y_{m+\nu+i} - \hat{y}_{m+\nu+i})^2 \\
&= \sum_{i=1}^{r} \sum_{\nu=1}^{k^*-i} (y_{m+\nu+i} - \hat{y}_{m+\nu+i})^2 + \sum_{i=1}^{r} \sum_{\nu=k^*-i+1}^{k-i} (y_{m+\nu+i} - \hat{y}_{m+\nu+i})^2 \\
&= B_{k,1}(m) + \cdots + B_{k,4}(m),
\end{aligned}$$

where we have used the further refined abbreviations

$$\begin{aligned}
B_{k,1}(m) &= \sum_{i=1}^{r} \sum_{\nu=1}^{k^*-i} (y_{m+\nu+i} - \hat{y}_{m+\nu+i})^2 + \sum_{i=1}^{r} \sum_{\nu=k^*-i+1}^{k-i} (\varepsilon_{m+\nu+i} - \eta_{m+\nu})^2, \\
B_{k,2}(m) &= 2 \sum_{i=1}^{r} \sum_{\nu=k^*-i+1}^{k-i} \alpha_{m+\nu+i} \varepsilon_{m+\nu+i}, \\
B_{k,3}(m) &= -2 \sum_{i=1}^{r} \sum_{nu=k^*-i+1}^{k-i} \alpha_{m+\nu+i} \eta_{m+\nu}, \\
B_{k,4}(m) &= \sum_{i=1}^{r} \sum_{\nu=k^*-i+1}^{k-i} \alpha_{m+\nu+i}^2.
\end{aligned}$$

The first term, $B_{k,1}(m)$, does not depend on the parameter subject of change, $\boldsymbol{\Delta}$. Therefore, it is bounded in probability.

Lemma 8 *Let the assumptions of Theorem 4 be satisfied. Then, as* $m \to \infty$,

$$\sup_{1 \leq k < \infty} \frac{1}{g(m,k)} \left| B_{k,1}(m) - \frac{k}{m} \sum_{i=1}^{r} \sum_{\nu=p}^{m-r} (y_{\nu+i} - \hat{y}_{\nu+i})^2 \right| = \mathcal{O}_P(1).$$

Proof: The assertion follows directly from Theorem 2, since the process under consideration is exactly the one under the null hypothesis H_0.

Next, we show that $B_{k,4}(m)$ will be asymptotically 'large'. Subsequently, it will turn out that this term determines the asymptotics under the alternative H_A, since it dominates also $B_{k,2}(m)$ and $B_{k,3}(m)$.

Lemma 9 *Let the assumptions of Theorem 4 be satisfied. Then, for any* $s \in (0, 1)$, *there are* k_0 *and* m_0 *such that*

$$B_{k,4}(m) \geq (1-s) r \Delta^T \mathbf{C} \Delta \frac{m+k^*}{m+k} \left(k - k^*\right)$$

if $m \geq m_0$ and $k \geq k^* + k_0$.

Proof: By assumption (14) and Lemma 1, for any $1 \le i \le r$ and for a given $s \in (0, 1)$, we can find an integer $m_0 = m_0(s)$ such that

$$\begin{aligned}\sum_{\nu=k^*-i+1}^{k-i} \alpha^2_{m+\nu+i} &= \textstyle\sum_{\nu=k^*-i+1}^{k-i} \Delta^T \mathbf{C}_{m+k^*}\mathbf{C}^{-1}_{m+\nu}\mathbf{x}_{m+\nu+i}\mathbf{x}^T_{m+\nu+i}\mathbf{C}^{-1}_{m+\nu}\mathbf{C}_{m+k^*}\Delta \\ &\ge (1-s)\sum_{\nu=k^*-i+1}^{k-i} \Delta^T \left(\frac{m+k^*}{m+\nu}\right)^2 \mathbf{x}_{m+\nu+i}\mathbf{x}^T_{m+\nu+i}\Delta\end{aligned} \tag{40}$$

if $m \ge m_0$ for all $k \ge 1$. By Abel's summation formula, we get that

$$\begin{aligned}&\sum_{\nu=k^*-i+1}^{k-i} \frac{1}{(m+\nu)^2}\mathbf{x}_{m+\nu+i}\mathbf{x}^T_{m+\nu+i} \\ &= \sum_{\nu=k^*-i+1}^{k-i}\left(\frac{1}{[m+\nu]^2} - \frac{1}{[m+\nu+1]^2}\right)\sum_{j=k^*-i+1}^{\nu} \mathbf{x}_{m+j+i}\mathbf{x}^T_{m+j+i} \\ &\quad + \frac{1}{(m+k-i+1)^2}\sum_{\nu=k^*-i+1}^{k-i}\mathbf{x}_{m+\nu+i}\mathbf{x}^T_{m+\nu+i}.\end{aligned} \tag{41}$$

Using (14) and again Abel's summation formula, there is k_0 such that, for any $s > 0$,

$$\begin{aligned}&\sum_{\nu=k^*-i+1}^{k-i}\left(\frac{1}{[m+\nu]^2} - \frac{1}{[m+\nu+1]^2}\right)(\nu - k^* + i) + \frac{k-k^*}{(m+k-i+1)^2} \\ &= \sum_{\nu=k^*-i+2}^{k-i}\frac{1}{(m+\nu)^2} + \frac{1}{(m+k^*-i+1)^2} \\ &= \sum_{\nu=k^*-i+1}^{k-i}\frac{1}{(m+\nu)^2} \\ &\ge (1-s)\tfrac{k-k^*}{(m+k)(m+k^*)}\end{aligned}$$

for all $1 \le i \le r$ if $k \ge k^* + k_0$. So by (14), (40) and (41), we have

$$B_{k,4}(m) = \sum_{i=1}^{r}\sum_{\nu=k^*-i+1}^{k-i}\alpha^2_{m+\nu+i} \ge (1-s)r\Delta^T\mathbf{C}\Delta\frac{m+k^*}{m+k}(k-k^*)$$

if $k \ge k^* + k_0$ and $m \ge m_0$, completing the proof.

It remains to investigate the asymptotics of the mixed terms $B_{k,2}(m)$ and $B_{k,3}(m)$. The first term is studied in the following lemma.

Lemma 10 *Let the assumptions of Theorem 4 be satisfied. Then, for all integers* $1 \le a \le m$ *and* $\lambda > 0$, *there is a constant c such that*

$$P\left\{\sup_{k^*+a\le k<\infty}\frac{(m+k)|B_{k,2}(m)|}{\Delta^T\mathbf{C}\Delta(m+k^*)(k-k^*)} \ge \lambda\right\} \le \frac{c}{\Delta^T\mathbf{C}\Delta}\frac{\sigma^2}{\lambda^2}.$$

Proof: Choose $1 \le a \le m$ and $\lambda > 0$. The Hájek–Rényi inequality yields

$$P\left\{\sup_{k^*+a\le k<\infty} \frac{(m+k)}{\Delta^T\mathbf{C}\Delta(m+k^*)(k-k^*)}\left|\sum_{\nu=k^*-i+1}^{k-i}\alpha_{m+\nu+i}\varepsilon_{m+\nu+i}\right| \ge \lambda\right\}$$
$$\le \frac{\sigma^2}{\lambda^2}\left[\left(\frac{m+k^*+a}{\Delta^T\mathbf{C}\Delta(m+k^*)a}\right)^2\sum_{\nu=k^*-i+1}^{k^*+a-1}\alpha^2_{m+\nu+i}\right.$$
$$\left.+\sum_{\nu=k^*+a}^{\infty}\left(\frac{m+\nu}{\Delta^T\mathbf{C}\Delta(m+k^*)(\nu-k^*)}\right)^2\alpha^2_{m+\nu+i}\right].$$

Using Lemma 1, there is a constant c_1 such that

$$\sum_{\nu=k^*+a}^{\infty}\left(\frac{m+\nu}{\Delta^T\mathbf{C}\Delta(m+k^*)(\nu-k^*)}\right)^2\alpha^2_{m+\nu+i}$$
$$\le \frac{c_1}{(\Delta^T\mathbf{C}\Delta)^2}\sum_{\nu=k^*+a}^{\infty}\frac{1}{(\nu-k^*)^2}\Delta^T\mathbf{x}_{m+\nu+i}\mathbf{x}^T_{m+\nu+i}\Delta.$$

Elementary computations give

$$\sum_{\nu=k^*+a}^{\infty}\left(\frac{1}{[\nu-k^*]^2}-\frac{1}{[\nu-k^*+1]^2}\right)(\nu-k^*-a) \le \frac{c_2}{a}$$

and therefore (14) and Abel's summation formula imply that

$$\sum_{\nu=k^*+a}^{\infty}\frac{1}{(\nu-k^*)^2}\Delta^T\mathbf{x}_{m+\nu+i}\mathbf{x}^T_{m+\nu+i}\Delta \le \frac{c_3}{a}\Delta^T\mathbf{C}\Delta.$$

Similar arguments also give

$$\left(\frac{m+k^*+a}{\Delta^T\mathbf{C}\Delta(m+k^*)(k-k^*)}\right)^2\sum_{\nu=k^*-i+1}^{k^*+a-1}\alpha^2_{m+\nu+i} \le \frac{c_4}{a}\Delta^T\mathbf{C}\Delta,$$

hence completing the proof of Lemma 10.

Finally, we verify that $B_{k,3}(m)$ is small compared to $B_{k,4}(m)$.

Lemma 11 *Let the assumptions of Theorem 4 be satisfied. Then, for any $M > k^*$, there is a constant C such that*

$$P\left\{\max_{k^*<k\le M}\frac{(m+k)|B_{k,3}(m)|}{\Delta^T\mathbf{C}\Delta(m+k^*)(k-k^*)} \ge \lambda\right\} \le \frac{C}{\lambda^2}\frac{M-k^*}{(m+k^*)^2}$$

for all $\lambda > 0$.

Proof: Using Abel's summation formula, we obtain that

$$\begin{aligned}
\sum_{\nu=k^*-i+1}^{k-i} \alpha_{m+\nu+i}\eta_{m+\nu} &= \sum_{\nu=k^*+-i+1}^{k-i} \alpha_{m+\nu+i}\mathbf{x}_{m+\nu+1}^T \mathbf{C}_{m+\nu}^{-1} \sum_{j=1}^{m+\nu} \mathbf{x}_j \varepsilon_j \\
&= \sum_{\nu=1-m}^{k^*-i+1} \left(\sum_{\ell=k^*-i+1}^{k-i} \alpha_{m+\ell+i}\mathbf{x}_{m+\ell+1}^T \mathbf{C}_{m+\ell}^{-1} \right) \mathbf{x}_{m+\nu}\varepsilon_{m+\nu} \\
&\quad + \sum_{\nu=k^*-i+2}^{k-i} \left(\sum_{\ell=\nu}^{k-i} \alpha_{m+\ell+i}\mathbf{x}_{m+\ell+1}^T \mathbf{C}_{m+\ell}^{-1} \right) \mathbf{x}_{m+\nu}\varepsilon_{m+\nu} \\
&= \sum_{\nu=1-m}^{k^*-i+1} \mathbf{R}_k(m)\mathbf{x}_{m+\nu}\varepsilon_{m+\nu} + \sum_{\nu=k^*-i+2}^{k-i} \mathbf{S}_{k,\nu}(m)\mathbf{x}_{m+\nu}\varepsilon_{m+\nu}.
\end{aligned}$$

Note that $\mathbf{R}_k(m)$ does not depend on the summation index ν. The Hájek–Rényi inequality yields

$$\begin{aligned}
&P\left\{ \max_{k^*<k\le M} \frac{m+k}{\Delta^T \mathbf{C}\Delta(m+k^*)(k-k^*)} \left| \sum_{\nu=1-m}^{k^*-i+1} \mathbf{R}_k(m)\mathbf{x}_{m+\nu}\varepsilon_{m+\nu} \right| \ge \lambda \right\} \\
&\le \frac{\sigma^2}{\lambda^2} \sum_{k=k^*+1}^{M} \left(\frac{(m+k)\mathbf{R}_k(m)\mathbf{x}_{k+1}}{\Delta^T \mathbf{C}\Delta(m+k^*)(k-k^*)} \right)^2 .
\end{aligned}$$

Applying Lemmas 1 and 14, we get that, for all i,

$$\begin{aligned}
|\mathbf{R}_k(m)| &= \left| \sum_{\ell=k^*-i+1}^{k-i} \alpha_{m+\ell+i}\mathbf{x}_{m+\ell+1}^T \mathbf{C}_{m+\ell}^{-1} \right| \\
&= \left| \sum_{\ell=k^*-i+1}^{k-i} \mathbf{x}_{m+\ell+i}^T \mathbf{C}_{m+\ell}^{-1} \mathbf{C}_{m+k^*} \Delta \mathbf{x}_{m+\ell+1}^T \mathbf{C}_{m+\ell}^{-1} \right| \\
&\le C \sum_{\ell=k^*-i+1}^{k} |\mathbf{x}_{m+\ell+i}||\mathbf{x}_{m+\ell+1}| \frac{m+k^*}{(m+\ell)^2} |\Delta| \\
&= C(m+k^*)|\Delta| \sum_{\ell=k^*-i+1}^{k} (m+\ell)^{-2} |\mathbf{x}_{m+\ell+i}||\mathbf{x}_{m+\ell+1}| \\
&= C(m+k^*) \frac{k-k^*}{(m+k)(m+k^*)},
\end{aligned}$$

where C denotes a universal constant which may vary from line to line. We arrive at

$$\begin{aligned}
&\sum_{k=k^*+1}^{M}\left(\frac{m+k}{\Delta^T\mathbf{C}\Delta(m+k^*)(k-k^*)}\right)^2\left(\left[\sum_{\ell=k^*-i+1}^{k-i}\alpha_{m+\ell+i}\mathbf{x}_{m+\ell+1}^T\mathbf{C}_{m+l}^{-1}\right]\mathbf{x}_{k+1}\right)^2\\
&\le C\sum_{k=k^*+1}^{M}\left(\frac{m+k}{\Delta^T\mathbf{C}\Delta(m+k^*)(k-k^*)}\frac{(m+k^*)(k-k^*)^2}{(m+k)(m+k^*)}|\mathbf{x}_{k+1}|\right)^2\\
&\le \frac{C}{(m+k^*)^2}\sum_{k=k^*+1}^{M}|\mathbf{x}_{k+1}|^2\\
&\le \frac{C(M-k^*)}{(m+k^*)^2}.
\end{aligned}$$

Similar arguments apply also in case of $\sum_{\nu=k^*-i+2}^{k-i}\mathbf{S}_{k,\nu}(m)\mathbf{x}_{m+\nu}\varepsilon_{m+\nu}$, thus completing the proof.

Using the previous auxiliary results, we are in a position to prove the limiting behaviour of our test statistics under H_A.

Proof of Theorem 4. By Lemmas 8–11, it is enough to investigate the quantity

$$r(k^*,k)=\frac{\frac{m+k^*}{m+k}(k-k^*)}{\sqrt{m}\left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^{\gamma}}.$$

(i) Let $0\le\beta\le(1-2\gamma)/(2-2\gamma)=\zeta$ and set $\tilde{k}=k^*+Cm^{\zeta}$. Note that, by definition of k^* and $\tilde{k}$, $\frac{m+k^*}{m+\tilde{k}}\to 1$ and $1+\frac{\tilde{k}}{m}\to 1$ as $m\to\infty$. Furthermore, $\tilde{k}/(m+\tilde{k})\sim\tilde{k}/m\sim Cm^{\zeta-1}$ as $m\to\infty$, since $\beta\le\zeta$. [Let $\{a_n\}$ and $\{b_n\}$ be sequences of real numbers. We say that a_n $\sim b_n$ if $a_nb_n^{-1}\to 1$ as $n\to\infty$.] Hence,

$$r(k^*,\tilde{k})\sim\frac{Cm^{\zeta}}{\sqrt{m}C^{\gamma}m^{\gamma(\zeta-1)}}=C^{1-\gamma}$$

resulting in $\liminf_{m\to\infty}r(k^*,\tilde{k})=C^{1-\gamma}$, which can be as large as we want choosing C appropriately large.

(ii) Let $\zeta\le\beta<1$ and set $\tilde{k}=k^*+Cm^{1/2-\gamma(1-\beta)}$. Now

$$\lim_{C\to\infty}\liminf_{m\to\infty}r(k^*,\tilde{k})=\lim_{C\to\infty}\frac{C}{C^{\gamma}}=\infty.$$

(iii) Similar arguments work also in the third case.

ACKNOWLEDGMENTS

We thank Michael W. McCracken for making available the data used in Section 4. We are grateful to the co-editor Pierre Perron and three referees for useful suggestions for improving the paper. This research was partially supported by NATO grant PST.EAP.CLG 980599, NSF grants INT–0223262, DMS–0413653, DMS–0604670 and grant MSM 0021620839.

REFERENCES

Andrews, D. W. K. (1991). Asymptotic normality of series estimators for nonparametric and semiparametric regression models. *Econometrica 59*, 307–45.

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica 61*, 821–56.

Aue, A., I. Berkes and L. Horváth (2006). Strong approximation for the sum of squares of augmented GARCH sequences. Bernoulli 12, 583–608.

Bai, J. R. L. Lumsdaine and J. H. Stock (1998). Testing for and dating common breaks in multivariate time series. *Review of Economic Studies 65*, 395–432.

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

Bai, J. and P. Perron, (2003). Critical values for multiple structural change tests. *The Economometrics Journal 6*, 72–8.

Brown, R. L., J. Durbin, and J. M. Evans, (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society Series B. 37*, 149–92.

Carrasco, M., and X. Chen, (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory 18*, 17–39.

Chow, Y. S., and H. Teicher (1988). *Probability Theory* (*2nd ed.*). Springer, New York.

Chu, C.–S., J. Stinchcombe and H. White (1996). Monitoring structural change. *Econometrica 64*, 1045–65.

Clark, T. E. and M. W. McCracken (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics 105*, 85–110.

Clark, T. E. and M. W. McCracken (2005). The power of tests of predictive ability in the presence of structural breaks. *Journal of Econometrics 124*, 1–31.

Csörg, M. and L. Horváth (1997). *Limit Theorems in Change–Point Analysis*. Wiley, Chichester.

Csörg, M. and P. Révész (1981). *Strong Approximations in Probability and Statistics*. Acacemic Press, New York.

Duan, J.-C. (1997). GARCH (p, q) process and its diffusion limit. *Journal of Econometrics 79*, 97–127.

Eberlein, E. (1986). On strong invariance principles under dependence assumptions. *Annals of Probability 14*, 260–70.

Geweke, J. (1986). Modeling persistence of conditional variances: A comment. *Econometric Reviews 5*, 57–61.

Gombay, E. and L. Horváth (1994). An application of the maximum likelihood test to the change-point problem. *Stochastic Processes and their Applications 50*, 161–71.

Hansen, B. E. (1995). Regression with nonstationary volatility. *Econometrica 63*, 1113–32.

Hansen, B. E. (2000). Testing for structural change in conditional models. *Journal of Econometrics 97*, 93–115.

Hawkins, D. L. (1987). A test for a change point in a parametric model based on a maximal Wald-type statistic. *Sankhyā. The Indian Journal of Statistics Series A 49*, 368–76.

Hawkins, D. L. (1989). A U–I approach to retrospective testing for shifting parameters in a linear model. *Communications in Statistics Theory and Methods 18*, 3117–34.

Horváth, L. (1995). Detecting changes in linear regressions. *Statistics 26*, 189–208.

Horváth, L., M. Hušková, P. Kokoszka, and J. Steinebach (2004). Monitoring changes in linear models. *Journal of Statistical Planning and Inference 126*, 225–51.

Horváth, L. and Q. M. Shao (1995). Limit theorems for the union-intersection test. *Journal of Statistical Planning and Inference 44*, 133–48.

Krämer, W., W. Ploberger and R. Alt (1988). Testing for structural change in dynamic models. *Econometrica 56*, 1355–69.

McCracken, M. W. (2000). Robust out–of–sample inference. *Journal of Econometrics 99*, 195–223.

Nelson, D. B. (1991). Conditional heteroskedastisticity in asset returns: A new approach. *Econometrica 59*, 347–70.

Ploberger, W., W. Krämer and K. Kontrus (1989). A new test for structural stability in the linear regression model. *Journal of Econometrics 40*, 307–18.

Quandt, R. E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association 53*, 873–80.

Quandt, R. E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American Statistical Association 55*, 324–30.

APPENDIX A:
AUGMENTED GARCH PROCESSES

As mentioned in the Introduction, we wish to allow the errors ε_i to exhibit conditional heteroskedasticity. We can achieve this by modelling them as an augmented GARCH(1,1) sequence (see Duan 1997). Most conditionaly heteroskedastic models used in practice are augmented GARCH(1,1) processes, (see Carrasco and Chen 2002 and Aue *et al.* 2006 for specific examples). The goal of this section is to verify that conditions (15), (16) and (19), (20) hold under appropriate assumptions, if

$$\{\varepsilon_i\} \text{ is an augmented GARCH(1,1) sequence} \tag{A1}$$

defined by the recursions

$$\varepsilon_k = \sigma_k \zeta_k, \tag{A2}$$

$$\Lambda(\sigma_k^2) = h(\zeta_{k-1})\Lambda(\sigma_{k-1}^2) + g(\zeta_{k-1}), \tag{A3}$$

where $-\infty < k < \infty$; $\Lambda(x)$, $h(x)$ and $g(x)$ are real-valued functions defined on the non-negative real numbers and the real numbers, respectively, and

$$\begin{gathered}\{\zeta_k : -\infty < k < \infty\} \text{ are independent, identically distributed} \\ \text{random variables with } E\zeta_0 = 0.\end{gathered} \tag{A4}$$

To solve for σ_k^2, we must assume that the inverse function

$$\Lambda^{-1}(x) \quad \text{exists.} \tag{A5}$$

Following Aue *et al.* (2006), we further assume

$$E|\zeta_0|^{8+\delta} < \infty \text{ with some } \delta > 0; \tag{A6}$$

$$E|g(\zeta_0)|^{\nu} < \infty \text{ with some } \nu > 4(1+\max\{0, z\}); \tag{A7}$$

$$\Lambda(\sigma_0^2) \geq \omega \qquad \text{a.s.} \qquad \text{with some } \omega > 0; \tag{A8}$$

Λ' exists and is non-negative, and there are C and z such that

$$\left| \frac{1}{\Lambda\prime(\Lambda^{-1}(x))} \right| \leq Cx^{z} \text{ for all } x \geq \omega. \tag{A9}$$

Assumptions (A.6)–(A.9) cover the case when Λ is a polynomial-type function (i.e., $\Lambda(x) = x^b$). The case of $\Lambda(x) = \log x$, the so-called exponential GARCH model, received special attention in the literature (cf. Geweke 1986 and Nelson 1991) and requires a stronger assumption:

$$\begin{array}{c}\text{if } \Lambda(x) = \log x, \text{ then } E \exp(t|g(\zeta_0)|) \text{ exists with some } t > 4 \\ \text{and } |h(x)| < h \text{ with some } 0 < h < 1.\end{array} \tag{A10}$$

Throughout the sequel, we assume that the augmented GARCH sequence in (A.2) and (A.3) satisfies (A.4)–(A.6), (9) and, in case of polynomial-type functions, (A.7)–(A.9) or, for exponential GARCH models, (6.10). We note that under these conditions $\{\varepsilon_k\}$ is a stationary and ergodic sequence with $E\varepsilon_0 = 0$ and $E\varepsilon_0^4 < \infty$ (see Aue *et al.* 2006).

Lemma A.1 *If the conditions of Theorem 1 are satisfied, then, for each m, there are independent Wiener processes* $\{W_{1,m}(t){:}t \geq 0\}$ *and* $\{W_{2,m}(t){:}t \geq 0\}$ *such that (15) and (16) are satisfied.*

Proof: First, we show that the partial sums $\sum_{i=1}^{m} \varepsilon_i$ and $\sum_{i=m+1}^{m+k} \varepsilon_i$ can be approximated with independent Wiener processes. Following the proof of lemma 5.4 in Aue *et al.* (2006), we can define random variables $\bar{\varepsilon}_i \in \sigma(\zeta_j : i - i^{\rho} \leq j \leq i)$ with some $0 < \rho < 1/10$ such that

$$\max_{1 \leq k < \infty} \left| \sum_{i=1}^{k} \varepsilon_i - \sum_{i=1}^{k} \bar{\varepsilon}_i \right| = \mathcal{O}(1) \qquad \text{a.s.} \tag{A11}$$

So it is sufficient to approximate $\sum_{i=1}^{m} \bar{\varepsilon}_i$ and $\sum_{i=m+1}^{m+k} \bar{\varepsilon}_i$. It follows from the definition of $\bar{\varepsilon}_i$ that $\sum_{i=1}^{m-m^{\rho}} \bar{\varepsilon}_i$ and $\{\sum_{i=m+1}^{m+k} \bar{\varepsilon}_i : 1 \leq k < \infty\}$ are independent. We show below that there is a $0 < \Delta < 1/2$ such that

$$\sum_{i=m-m^{\rho}+1}^{m} \bar{\varepsilon}_i \overset{\text{a.s.}}{=} \mathcal{O}\left(m^{\Delta}\right) \qquad (m \to \infty). \tag{A12}$$

Hence, we can approximate $\sum_{i=1}^{m} \bar{\varepsilon}_i$ and $\{\sum_{i=m+1}^{m+k} \bar{\varepsilon}_i : 1 \leq k < \infty\}$ with two independent Wiener processes. In light of (A.11) and the stationarity of the ε_i, we need to show only that there is a Wiener process $\{W(t){:}t \geq 0\}$ such that

$$\sup_{1 \leq k < \infty} \frac{1}{k^{\Delta}} \left| \sum_{i=1}^{k} \varepsilon_i - \sigma W(k) \right| = \mathcal{O}_P(1). \tag{A13}$$

It is clear that (A.13) and the upper bounds for the increments of $\{W(t){:}t \geq 0\}$ in Csörg and Révész (1981) imply (A.12).

The proof of (A.13) is based on the strong approximation in Eberlein (1986). Let $\mathcal{F}_m = \sigma(\zeta_i : i \leq m)$. First, it follows from the definition of the ε_i that

$$E\left(\sum_{i=m+1}^{m+n} \varepsilon_i \,\middle|\, \mathcal{F}_m \right) = 0 \qquad \text{a.s.}$$

Since $\{\varepsilon_i\}$ is a sequence of orthogonal martingales, we get that

$$\operatorname{Var}\left(\sum_{i=m+1}^{m+n}\varepsilon_i\right)=n\sigma^2.$$

Similarly, it holds a.s.,

$$E\left[\left(\sum_{i=m+1}^{m+n}\varepsilon_i\right)^2\middle|\mathcal{F}_m\right]=E\left[\sum_{i=m+1}^{m+n}\sum_{j=m+1}^{m+n}\zeta_i\zeta_j\sigma_i\sigma_j\middle|\mathcal{F}_m\right]=E[\zeta_0^2]\sum_{i=m+1}^{m+n}E\left[\sigma_i^2|\mathcal{F}_m\right].$$

Next, we show that, uniformly in m,

$$E\left|\sum_{i=m+1}^{m+n}E\left[\sigma_i^2|\mathcal{F}_m\right]-\sum_{i=m+1}^{m+n}E\sigma_i^2\right|=\mathcal{O}\left(n^{1-\Delta^*}\right)$$

with some $\Delta^*>0$. By stationarity, it is enough to show

$$E\left|\sum_{i=1}^{n}E\left[\sigma_i^2|\mathcal{F}_0\right]-\sum_{i=1}^{n}E\sigma_i^2\right|=\mathcal{O}\left(n^{1-\Delta^*}\right). \tag{A14}$$

The proof of (A.14) is based on the truncation method of Aue *et al.* (2006). Therein, random variables $\sigma_i^*\in\sigma(\zeta_j\,:\,i-i^{\rho^*}<j\le i)$ are defined with $\rho^*>1/4$. They showed that

$$\sum_{i=1}^{\infty}E\left|E\left[\sigma_i^2|\mathcal{F}_0\right]-E\left[\sigma_i^{*2}|\mathcal{F}_0\right]\right|<\infty$$

and therefore

$$\sum_{i=1}^{n}E\left[\sigma_i^2|\mathcal{F}_0\right]=\textstyle\sum_{i=1}^{n}E\left[\sigma_i^{*2}|\mathcal{F}_0\right]+\mathcal{O}(1)=\sum_{i=1}^{n}E[\sigma_i^2]+\mathcal{O}(1).$$

The approximation in (A.13) now follows from Eberlein (1986).

Lemma A.2 *If the conditions of Theorem 2 are satisfied then, for each m, there are independent Wiener processes* $\{W_{3,m}(t){:}t\ge 0\}$ *and* $\{W_{4,m}(t){:}t\ge 0\}$ *such that (19) and (20) are satisfied.*

Proof: Approximations of sums of $\varepsilon_i^2-\sigma^2$ were considered by Aue *et al.* (2006). Since $\{\varepsilon_i^2\}$ is a stationary sequence with exponentially decreasing autocorrelation, their method also gives the independence of $W_{3,m}$ and $W_{4,m}$. By lemma 5.4 of Aue *et al.* (2006), there are random variables $\bar\varepsilon_i\in\sigma(\zeta_j:i-i^\rho\le j\le i)$ with some $0<\rho<1/10$ such that

$$\max_{1\le k<\infty}\left|\sum_{i=1}^{k}\varepsilon_i^2-\sum_{i=1}^{k}\bar\varepsilon_i^2\right|=\mathcal{O}(1)\qquad\text{a.s.} \tag{A15}$$

Hence, it is enough to approximate $\sum_{i=1}^{m}\bar\varepsilon_i^2$ and $\sum_{i=m+1}^{m+k}\bar\varepsilon_i^2$. It follows from the definition of $\bar\varepsilon_i$ that $\sum_{i=1}^{m-m^\rho}\bar\varepsilon_i^2$ and $\{\sum_{i=m+1}^{m+k}\bar\varepsilon_i^2:k\ge 1\}$ are independent. Thus, we can establish the approximations for the sums independently of each other. By theorem 2.4 of Aue *et al.* (2006), we can define a

Wiener process $\{W_{4,m}(t):t \geq 0\}$ such that

$$\sum_{i=1}^{m-m^\rho} \left(\bar{\varepsilon}_i^2 - \sigma^2\right) - W_{4,m}(m - m^\rho) = \mathcal{O}_P\left(m^\Delta\right)$$

with any $\Delta > 3/8$,

$$\sum_{i=1}^{m-m^\rho} \left(\bar{\varepsilon}_i^2 - \sigma^2\right) - \sum_{i=1}^{m} \left(\bar{\varepsilon}_i^2 - \sigma^2\right) = \mathcal{O}_P\left(m^\Delta + \sqrt{m^\rho \log m}\right),$$

and

$$W_{4,m}(m - m^\rho) - W_{4,m}(m) = \mathcal{O}\left(\sqrt{m^\rho \log m}\right)$$

by the strong law for the increments of a Wiener process (cf. Csörg and Révész 1981). Hence, (20) is proved.

In light of (A.15), we can define a sequence $\{\tilde{\varepsilon}_i : -\infty < i < \infty\}$ independent of $\sum_{i=1}^{m-m^\rho} \bar{\varepsilon}_i^2$ such that $\{\tilde{\varepsilon}_i : -\infty < i < \infty\} \overset{\mathcal{D}}{=} \{\varepsilon_i : -\infty < i < \infty\}$ and

$$\max_{1 \leq k < \infty} \left| \sum_{i=1}^{k} \tilde{\varepsilon}_i^2 - \sum_{i=1}^{k} \bar{\varepsilon}_i^2 \right| = \mathcal{O}(1) \qquad \text{a.s.}$$

Since $\{\tilde{\varepsilon}_i\}$ is a stationary sequence, theorem 2.4 of Aue *et al.* (2006) implies

$$\sup_{1 \leq k < \infty} \frac{1}{k^\Delta} \left| \sum_{i=m+1}^{m+k} \tilde{\varepsilon}_i^2 - W_{3,m}(k) \right| = \mathcal{O}_P(1)$$

with a suitably chosen Wiener process $W_{3,m}(t)$ and with any $\Delta > 3/8$. Thus, (19) is also proved.