Journal of Econometrics 149 (2009) 174–190

Contents lists available at ScienceDirect

Journal of Econometrics

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

Delay times of sequential procedures for multiple time series regression models[☆]

Alexander Aue[a,*], Lajos Horváth[b], Matthew L. Reimherr[c]

[a] Department of Statistics, University of California, One Shields Avenue, Davis, CA 95616, USA
[b] Department of Mathematics, University of Utah, 155 South 1440 East, Salt Lake City, UT 84112, USA
[c] Department of Statistics, University of Chicago, 5734 S. University Avenue, Chicago, IL 60637, USA

ARTICLE INFO

Article history:
Received 30 April 2007
Received in revised form
15 December 2008
Accepted 22 December 2008
Available online 30 January 2009

JEL classification:
C01
C22

Keywords:
Change-point estimation
CUSUM statistic
Linear models
Sequential tests
Structural stability
Threshold function
Time series regressors

ABSTRACT

We consider a multiple regression model in which the explanatory variables are specified by time series. To sequentially test for the stability of the regression parameters in time, we introduce a detector which is based on the first excess time of a CUSUM-type statistic over a suitably constructed threshold function. The aim of this paper is to study the delay time associated with this detector. As our main result, we derive the limit distribution of the delay time and provide thereby a theory that extends the benchmark average run-length concept utilized in most of the sequential monitoring literature. To highlight the applicability of the limit results in finite samples, we present a Monte Carlo simulation study and an application to macroeconomic data.

1. Introduction

Testing time series data for structural stability is undoubtedly of great importance because estimation and forecast techniques, carried out under the false assumption of stationarity, will inadvertently lead to inaccurate conclusions. Statisticians and econometricians assess the structure of a given set of time series observations with a variety of retrospective and sequential tools. For the retrospective case, testing and change-point estimation procedures have been widely studied and are well established in the literature. Important contributions include Andrews (1993), Hansen (2001), Perron (1989, 1997), and Stock and Watson (1996), among others. The interested reader is also referred to the monographs Brodsky and Darkhovsky (1993), Csörgő and Horváth (1997), and Perron (2006) which comprise a broad range of methods relevant for the retrospective analysis. Sequential procedures seem to be more useful when a decision has to be made on-line, as new data become available. Major developments in sequential change-point detection and diagnosis were initiated by the pioneering works of Shewhart (1931), Page (1954) and Quandt (1958, 1960) on quality control charts which have since resulted in a rich theory with widespread applications. For a detailed review we refer to Lai (2001).

While the body of literature is broad for retrospective tests, the number of publications in econometrics concerned with sequential change-point analysis is comparatively smaller, although growing at a fast rate in recent years. Chu et al. (1996) have developed fluctuation tests that are based on the general paradigm that an initial time period of length m is used to estimate a model with the goal to monitor for parameter changes on-line. The asymptotic analysis is carried out for $m \to \infty$. This approach has been extended and refined in several directions. Berkes et al. (2004) discussed the problem of sequentially detecting change-points in the parameters of GARCH(p, q) processes, while Hušková et al. (2007) analyzed the stability of autoregressive time series. Horváth et al. (2004) introduced residual based on-line monitoring schemes for possible changes in the regression parameters of a linear model. Aue et al. (2006b) generalized their setting to allow for large classes of dependent innovation sequences and, moreover, introduced

☆ We thank John Geweke and two anonymous referees for their constructive comments that have lead to significant improvements in a much revised version of the original paper. This research is partially supported by NSF grants DMS 0604670, DMS 0652420 and DMS 0602219 (VIGRE), and grant RGC–HKUST6428/06H.

* Corresponding author. Tel.: +1 530 752 6475; fax: +1 530 752 6475.
E-mail addresses: alexaue@wald.ucdavis.edu (A. Aue), horvath@math.utah.edu (L. Horváth), mreimherr@galton.uchicago.edu (M.L. Reimherr).

0304-4076/$ – see front matter
doi:10.1016/j.jeconom.2008.12.018

a new scheme based on squared prediction errors. These works have adopted sequential versions of the cumulative sum (CUSUM) statistics.

Other contributions in the literature are based on moving sum (MOSUM) procedures which generally have the advantage of faster detection times when compared to the corresponding CUSUM procedures (see, for example, Chu et al. (1995a,b)). On the other hand, Horváth et al. (2008) have shown that the application of MOSUM procedures requires additional knowledge on the moment structure of the underlying random variables to construct a valid asymptotic procedure. A unifying view on generalized fluctuation tests, incorporating all the above procedures, is offered in Kuan and Hornik (2005) and further in Leisch et al. (2000), and Zeileis et al. (2005).

The aforementioned papers, and most other contributions in the area (Leisch et al. (2000), is an exception), require constant or independent and identically distributed regressors. While this assumption is convenient from an analytical point of view, it is often violated in practice, for example, in regressions of econometric variables. Moreover, research in the area has focused on the level and power of the sequential monitoring procedures, and on the average speed of detection but not on the precise limiting distribution of the associated stopping times. In what follows, we are thus interested in determining the limiting distribution of delay times in linear models with time series regressors under structural instability. The dependence structure of the time series regressors is controlled by a novel moment assumption which has not been used in the econometrics literature before, but which is easy to verify in applications. We can build on a similar result of Aue and Horváth (2004) who studied asymptotic delay times in a simpler location model with random variables changing from one regime to another separated by different means. Related works for sequential change-point detection concerned with the average run-length of CUSUM-type procedures, mainly for normal variates, have been carried out by Basseville (1988), Brook and Evans (1972), Lucas (1982), and Siegmund and Venkatraman (1995), among others. As of now, there are no MOSUM counterparts of the results presented in this paper. This is largely due to the more delicate interplay between the moment structure and the choice of an appropriate stopping time (see Aue et al. (2008b)).

The paper is organized as follows. In Section 2, we specify the underlying multiple regression model and give examples of time series regressors which can be incorporated in the given approach. Section 3 introduces the sequential procedure and presents its limiting distribution. In Section 4 we report the results of a limited simulation study, while we apply our method to a classical economic data set that evaluates the wealth effect on consumption in Section 5. Section 6 contains conclusions and some guidelines for applied work. All proofs are given in Section 7.

2. The multiple time series regression setting

Let $\{y_i\}$ be the sequence of random variables to be observed and assume that they follow the multiple linear regression model

$$y_i = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_i + \varepsilon_i, \quad i = 1, 2, \ldots,$$

where $\{\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^{\mathrm{T}}\}$ is a sequence of p-dimensional random (or deterministic) explanatory variables, $\{\boldsymbol{\beta}_i = (\beta_{i1}, \ldots, \beta_{ip})^{\mathrm{T}}\}$ are p-dimensional parameter vectors, and $\{\varepsilon_i\}$ is an innovation sequence. The superscript $^{\mathrm{T}}$ is used to signify transposition.

Since in this paper we are interested in sequentially monitoring whether or not the regression parameter vectors $\{\boldsymbol{\beta}_i\}$ remain stable over time, we require (see assumption A of Chu et al. (1996)) that there is a non-contaminated reference frame of m observations for which

$$\boldsymbol{\beta}_i = \boldsymbol{\beta}_0, \quad i = 1, \ldots, m. \tag{2.1}$$

Condition (2.1) is referred to as a training period in the literature. It is a seemingly strong assumption. In practice, however, it does not pose a major challenge. If a sufficient amount of historical data is available, it can be analyzed with any of the common retrospective change-point methods available in the literature (see, for example, Hansen (2001)). Depending on the outcome of this pre-test, necessary adjustments can thereafter be performed to ensure structural stability. If the number of observations in the training period is too small to justify the application of asymptotic monitoring procedures, one may apply recursive methods such as Brown et al. (1975) until the training period is of appropriate size. The parameter m is introduced for the technical reason of making asymptotic statements in the monitoring setting precise, as we can now formulate limits "as m tends to infinity". The training period size thus assumes the role of the sample size, say n, in the retrospective setting, where asymptotic results are given "as n tends to infinity". It is nonetheless worthwhile to investigate how well the limit theory applies in finite samples. We shall do so in Section 4. In contrast to our approach, others (see Leisch et al. (2000) and Zeileis et al. (2005), for further discussion) have explicitly introduced a finite monitoring horizon after which the sequential procedure terminates. Limit results are stated in regard to this monitoring horizon requiring that the ratio of monitoring horizon size and training period size is equal to a constant bigger than unity.

Because the model parameters remain stable until observation m, (2.1) can be used for comparisons with future observations to find out if the stability assumption still holds. Thus, we shall introduce in Section 3 a monitoring procedure to discriminate between the no change in the regression parameters null hypothesis

$$H_0 : \boldsymbol{\beta}_i = \boldsymbol{\beta}_0, \quad i = m+1, m+2, \ldots, \tag{2.2}$$

and the break at an unknown time alternative hypothesis

$$\begin{aligned} H_A : &\text{There is a } k^* \geq 1 \quad \text{such that } \boldsymbol{\beta}_i = \boldsymbol{\beta}_0, \\ &i = m+1, \ldots, m+k^*-1, \\ &\text{but } \boldsymbol{\beta}_i = \boldsymbol{\beta}_A, \quad i = m+k^*, m+k^*+1, \ldots, \end{aligned} \tag{2.3}$$

where $\boldsymbol{\beta}_0 \neq \boldsymbol{\beta}_A$. The parameter k^* is called the change-point and is assumed unknown, as are the regression parameters $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_A$.

A standard approach in the linear models methodology is to consider the values of the regressors to be known and, hence, to study a fixed realization. In the present paper, we allow the $\{\mathbf{x}_i\}$ to be modeled by time series. To obtain the main result, we will then condition on a typical outcome of the $\{\mathbf{x}_i\}$. Particularly, we shall assume that the following conditions are satisfied. Let

$$\{\mathbf{x}_i\} \quad \text{be a stationary sequence;} \tag{2.4}$$

and assume that there are a p-dimensional vector $\mathbf{d} = (d_1, \ldots, d_p)^{\mathrm{T}}$ and $K > 0, \mu > 2$ such that, as $k \to \infty$,

$$E\left|\sum_{i=1}^{k}(x_{ij} - d_j)\right|^{\mu} \leq K k^{\mu/2}, \quad j = 1, \ldots, p. \tag{2.5}$$

Condition (2.5) is a novel assumption that has not been used in the econometrics literature before. It is easily verifiable both in theory and in practice as it is, in turn, virtually always implied by moment conditions. With (2.5) we require of the explanatory (time series) variables only that their centered partial sum fluctuations are controlled individually for each coordinate. No additional conditions regulating the inter-coordinate behavior of the vector $\mathbf{x}_i$ need to be imposed and dependence between the coordinates is therefore explicitly allowed. Condition (2.5) replaces the standard but more restrictive assumption which requires that $k^{-1} \sum_{i=1}^{k} \mathbf{x}_i \mathbf{x}_i^{\mathrm{T}}$ converges in probability or almost surely to a non-stochastic, positive definite matrix as $k \to \infty$ (see condition (M2) in Leisch et al. (2000), and condition (A2) in Zeileis et al. (2005); while Horváth et al. (2004), use an even stronger condition). Note also that we have used the optimal $\mu > 2$ in (2.5). In the statement of our main result, Theorem 3.1, however, we need to strengthen this moment condition.

Conditions (2.4) and (2.5) are satisfied for large classes of time series relevant in theory and applications. We are going to discuss some of them as lead examples in the following. The proofs of the statements of Examples 2.1–2.4 are postponed to Section 7.1. It is understood that the notation μ always refers to the corresponding quantity in (2.5) when used subsequently.

Example 2.1 (*Independent, Identically Distributed Random Variables*). Let $1 \le j \le p$. If the coordinate sequence $\{x_{ij}\}$ consists of independent, identically distributed (henceforth iid) random variables with $Ex_{ij} = d_j$ and $E|x_{ij}|^\mu < \infty$, then (2.5) is satisfied if $\mu > 2$.

Example 2.2 (*Linear Processes*). Let $\{z_i : i = 0, \pm 1, \pm 2, \ldots\}$ be a sequence of iid random variables with zero mean such that $E|z_i|^\mu < \infty$ for some $\mu > 2$ and, for $1 \le j \le p$, set

$$x_{ij} = d_j + \sum_{\ell=0}^{\infty} \varphi_\ell(j) z_{i-\ell},$$

where $\{\varphi_\ell(j)\}$ are absolutely summable coefficients, that is $\sum_{\ell=0}^{\infty} |\varphi_\ell(j)| < \infty$. Then, (2.5) is satisfied under the given assumptions.

Example 2.3 (*GARCH-type Sequences*). Let $1 \le j \le p$ and assume that the coordinates of $\{\mathbf{x}_i\}$ are given by the equations

$$x_{ij} = d_j + \sigma_i z_i, \tag{2.6}$$

where $\{z_i\}$ denotes an iid sequence of centered random variables with $E|z_i|^\mu < \infty$ for some $\mu > 2$. The conditional volatility sequence $\{\sigma_i^2\}$ is assumed to be stationary and measurable with respect to the σ-algebra $\sigma(z_\nu, \nu < i)$ generated by the past z_ν. If

$$E|\sigma_i|^\mu < \infty, \tag{2.7}$$

then the jth coordinate of $\{\mathbf{x}_i\}$ satisfies (2.5). Note that the term $\sigma_i z_i$ in the definition in (2.6) allows to include as specifications GARCH(p, q) processes or, more generally, the class of augmented GARCH processes (see Duan (1997), for details). In case of a GARCH(1,1) sequence, Nelson (1990) provided the necessary and sufficient conditions for (2.7) to hold. His result was extended to augmented GARCH(1,1) processes by Aue et al. (2006a), see also Carrasco and Chen (2002). For criteria on the existence of moments for general GARCH(p, q) sequences we refer to Ling and Mcaleer (2002). An alternative way to specify $\{\sigma_i z_i\}$ is provided by stochastic volatility models (see Tsay (2005), for an overview).

Example 2.4 (*Linear Processes with GARCH-type Innovations*). Let $1 \le j \le p$ and assume that the coordinates are given by

$$x_{ij} = d_j + \sum_{\ell=0}^{\infty} \varphi_\ell(j) z_{i-\ell}, \quad z_i = \sigma_i z_i',$$

where $\{z_i'\}$ is a sequence of centered iid random variables satisfying $E|z_i'|^\mu < \infty$ with some $\mu > 2$. It is moreover assumed that $\{\sigma_i\}$ is a stationary sequence and measurable with respect to $\sigma(z_\nu', \nu < i)$. Then, (2.5) holds if (2.7) holds and $\sum_{\ell=0}^{\infty} |\varphi_\ell(j)| < \infty$. As a special case, Example 2.4 includes the important class of ARMA-GARCH models that are frequently studied in econometrics (see Ling and Li (1998) and Ling (2007), for details).

In the remainder of this section we detail the assumptions on the innovation sequence $\{\varepsilon_i\}$. It is required that

$$\left| \sum_{i=1}^{m} \varepsilon_i \right| = \mathcal{O}_P\left(\sqrt{m}\right) \quad (m \to \infty), \tag{2.8}$$

and that there are a sequence of Brownian motions $\{W_m(t) : t \ge 0\}$ and a constant $\sigma > 0$ such that, for some $\nu > 2$,

$$\sup_{t \ge 1/m} \frac{1}{(mt)^{1/\nu}} \left| \sum_{i=m+1}^{m+mt} \varepsilon_i - \sigma W_m(mt) \right| = \mathcal{O}_P(1) \quad (m \to \infty). \tag{2.9}$$

Assumption (2.8) is similar to the central limit theorem but weaker, since only the order of the partial sums containing the first m innovations is specified and not the convergence in distribution to the normal law. Assumption (2.9) is a uniform weak invariance principle. Observe that the parameter σ can be interpreted as the asymptotic standard deviation of $k^{-1/2} \sum_{i=1}^{k} \varepsilon_i$. For a discussion of specific sequences $\{\varepsilon_i\}$ satisfying conditions (2.8) and (2.9) we refer to Aue and Horváth (2004).

Finally, we do not allow for interaction between the two sequences of random variables and therefore make the standard assumption that

$$\{\mathbf{x}_i\} \text{ and } \{\varepsilon_i\} \text{ are independent.} \tag{2.10}$$

Observe that, later on, we shall fix a realization of $\{\mathbf{x}_i\}$ that satisfies certain regularity conditions (see Propositions 7.1 and 7.2). (These are fulfilled for almost all realizations.)

3. The stopping rule and its limit distribution

To sequentially test the null hypothesis and its alternative introduced in the previous section, we need to define a stopping rule. Usually, these monitoring procedures are given in terms of first excess times of suitably constructed detectors and threshold functions (see Chu et al. (1996), Horváth et al. (2004) and Aue et al. (2006b), among others).

For $i \ge 1$, define the model residuals $\hat{\varepsilon}_i = y_i - \mathbf{x}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}}_m$, where $\hat{\boldsymbol{\beta}}_m$ denotes the least squares estimator for $\boldsymbol{\beta}_0$ based on the first m observations. Here we will work with the stopping time

$$\tau_m = \inf\{k \ge 1 : |\Gamma_m(k)| \ge g_m(k)\} \tag{3.1}$$

(using the convention $\inf \emptyset = \infty$), where

$$\Gamma_m(k) = \sum_{i=m+1}^{m+k} \hat{\varepsilon}_i \quad \text{and} \quad g_m(k) = q\sqrt{m}\left(1 + \frac{k}{m}\right)\left(\frac{k}{m+k}\right)^\gamma \tag{3.2}$$

with $q = q(\alpha) > 0$ and $0 \le \gamma < 1/2$. The quantity $\Gamma_m(k)$ is referred to as a detector. The model residuals $\{\hat{\varepsilon}_i\}$ have under H_0 a probabilistic structure similar to the innovations $\{\varepsilon_i\}$. Assumption (2.9) consequently implies that the detector $\Gamma_m(k)$ exhibits fluctuations akin to those of the Brownian motion $W_m(k)$ if the $\{\boldsymbol{\beta}_i\}$ are indeed stable over time. This relates the stopping time τ_m to crossing probabilities of Brownian motions over curved boundary functions, and motivates thus the choice of $g_m(k)$ in the present context. Our analysis in Section 7 reveals in fact that τ_m is asymptotically equivalent to the stopping time $\tau = \inf\{t \in [0, 1] : W(t) \ge q^* t^\gamma\}$, where $\{W(t) : t \in [0, 1]\}$ denotes a Brownian motion and $q^* = q^*(\alpha)$ a suitably chosen constant linked to $q = q(\alpha)$ in the following way. To ensure a pre-specified asymptotic level α for the sequential procedure, a practitioner can pick $q = \sigma q^*$ such that

$$P\left\{\sup_{0 \le t \le 1} \frac{|W(t)|}{t^\gamma} > q^*\right\} = \alpha \quad \text{under } H_0. \tag{3.3}$$

The boundary functions $g_m(k)$ are chosen due to the simple form they induce for the limit stopping rule τ. They depend, by construction, on a tuning parameter $\gamma \in [0, 1/2)$ that flexibly adjusts the sensitivity of the testing procedure. Note that the right endpoint $1/2$ is excluded, since H_0 would else, due to the law of the iterated logarithm for Brownian motions at zero, be rejected

Table 1
Performance of the monitoring procedures $\tau_m = \tau_m(\gamma)$ for selected values of m, k^* and γ in the case of the independent and identically distributed normal regressors specified in (A).

m	k^*	γ	min	$Q1$	med	$Q3$	max	m	k^*	γ	min	$Q1$	med	$Q3$	max
75	1	0.00	10	24	30	44	191	200	1	0.00	18	33	40	48	118
		0.25	4	14	20	27	156			0.25	5	16	21	28	87
		0.45	1	9	15	24	229			0.45	1	9	13	18	64
	7	0.00	15	32	40	49	216		20	0.00	34	55	63	73	155
		0.25	5	24	33	41	218			0.25	13	41	49	58	139
		0.45	1	20	29	39	498			0.45	1	36	42	52	116
	112	0.00	30	158	185	220	617		300	0.00	46	366	399	436	610
		0.25	11	154	183	219	530			0.25	22	360	394	431	595
		0.45	1	162	197	238	639			0.45	1	372	410	451	679
100	1	0.00	12	27	33	39	100	1000	1	0.00	47	74	82	93	150
		0.25	4	15	21	30	86			0.25	10	31	37	44	86
		0.45	1	8	14	22	99			0.45	1	8	12	20	61
	10	0.00	18	36	43	52	140		100	0.00	131	176	188	202	274
		0.25	12	26	32	42	144			0.25	106	145	156	170	233
		0.45	1	21	28	39	190			0.45	1	131	144	156	211
	150	0.00	41	202	231	229	241		1500	0.00	303	1644	1706	1765	1990
		0.25	15	199	229	266	508			0.25	135	1637	1699	1761	1966
		0.45	1	210	241	285	702			0.45	1	1649	1712	1772	2016

with probability one regardless whether it is true or not. Tabulated critical values q^* for various selections of α and γ are provided in Table 1 of Horváth et al. (2004). We will elaborate further on the practical aspects of τ_m in Section 4. Alternative forms of boundary functions $g_m(k)$ may be entertained as well. The interested reader is in this regard referred to Andreou and and Ghysels (2006) and the references cited in this paper.

Other sequential procedures for the same testing problem were introduced in Leisch et al. (2000), Horváth et al. (2004), Zeileis et al. (2005), and Aue et al. (2006b). Leisch et al. (2000) and Zeileis et al. (2005) investigated, for example, the behavior of moving sum (MOSUM) based detectors, while Horváth et al. (2004) used, next to the stopping rule τ_m from above, procedures based on recursive residuals, an approach advocated by Brown et al. (1975). Aue et al. (2006b) studied on-line schemes which are built on squared prediction errors. Empirical studies in these papers indicate that the CUSUM stopping rule τ_m performs superior in the case of changes soon after the end of the training period, while MOSUM-type procedures perform as well or better in the case of later changes.

The main aim of this paper is to derive the limit distribution of τ_m in the presence of a break in the sequence $\{\boldsymbol{\beta}_i\}$. This will require additional assumptions on the model parameters. Denote by $\boldsymbol{\Delta}_m = \boldsymbol{\beta}_A - \boldsymbol{\beta}_0$ the difference in the regression parameters before and after the change-point k^*. Since we are working under H_A, $\boldsymbol{\Delta}_m$ cannot be the zero vector. Note that we have explicitly allowed for $\boldsymbol{\Delta}_m$ to depend on m, which, in turn, implies that $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_A$ depend on m as well. For notational simplicity this is suppressed. To guarantee that the change can be identified by τ_m we need to relate $\boldsymbol{\Delta}_m$ also to the mean vector $\mathbf{d}$ of the explanatory variables $\{\mathbf{x}_i\}$. Let $\delta_m = \mathbf{d}^{\mathrm{T}}\boldsymbol{\Delta}_m$. It is assumed that there are positive constants C_1, C_2, and $C_3 \le C_4$ such that, as $m \to \infty$,

$$\frac{C_1}{\log m} \le |\boldsymbol{\Delta}_m| \le C_2, \quad \frac{C_1}{\log m} \le |\delta_m| \le C_2 \quad \text{and}$$
$$C_3 \le \frac{|\delta_m|}{|\boldsymbol{\Delta}_m|} \le C_4, \tag{3.4}$$

where $|\cdot|$ denotes the maximum norm of vectors. In (3.4), we allow δ_m to be constant as well as to tend to zero subject to a slow convergence. These conditions emulate corresponding assumptions commonly made in the retrospective change analysis, where limit theorems are often given with reference to fixed changes (the order of magnitude of the parameter differences before and after the breakpoint does not change with increasing sample size) and shrinking changes (the order of magnitude disappears in the long-run). The latter are particularly important if one aims at deriving the limit distribution for the change-point estimator (see Csörgő and Horváth (1997)). In the sequential setting, we show that both fixed and shrinking changes can be handled simultaneously as long as (3.4) is satisfied. The lower bounds for $|\boldsymbol{\Delta}_m|$ and $|\delta_m|$ in (3.4) include the term $\log m$ in the denominator to control for the fluctuations of the time series regressors. This will become apparent from Proposition 7.1 and Lemma 7.2. In the standard case of constant regressors, the additional log term may be dropped.

The final assumption ensures that the change-point k^* occurs shortly after the end of the training period: as $m \to \infty$,

$$k^* = \mathcal{O}\left(m^{\theta}\right) \quad \text{for some } 0 \le \theta < \frac{1-2\gamma}{4(1-\gamma)}. \tag{3.5}$$

Condition (3.5) is not motivated by the fact that CUSUM-type detectors perform better than other available procedures in the case of early changes (see the discussion above) but is rather a technical necessity imposed to ensure that the limit result of Theorem 3.1 holds in fact true (see part (iv) of Lemma 7.1). Given that (3.5) is required, the CUSUM procedure is also the optimal choice in terms of minimized empirical detection time.

The assumptions imposed are sufficient to find normalizing sequences $\{a_m\}$ and $\{b_m\}$ such that the standardized variables $\tau_m^0 = (\tau_m - a_m)/b_m$ converge in distribution to a standard normal random variable, whose distribution function is abbreviated by $\Phi(z)$.

Theorem 3.1. *If conditions (2.1), (2.3)–(2.5), (2.8)–(2.10), (3.4) and (3.5) hold, and if $\mu > 8(1-\gamma)/(1-2\gamma)$ then as $m \to \infty$,*

$$\lim_{m\to\infty} P\{\tau_m \le a_m + b_m z\} = \Phi(z)$$

for all real z, where

$$a_m = \left(c_m^{1-\gamma} - \frac{1}{c_m^{\gamma}|\delta_m|}\sum_{i=m+k^*}^{m+c_m}(\mathbf{x}_i - \mathbf{d})^{\mathrm{T}}\boldsymbol{\Delta}_m\right)^{1/(1-\gamma)},$$

$$b_m = \frac{\sqrt{c_m}\,\sigma}{(1-\gamma)|\delta_m|}$$

and $c_m = (qm^{1/2-\gamma}/|\delta_m|)^{1/(1-\gamma)}$ with $q = \sigma q^*(\alpha)$ determined by (3.3).

Theorem 3.1 establishes the central limit theorem for a suitably standardized version of τ_m. The quantities $\{a_m\}$ are consequently the centering constants which, by assumption (3.5) on k^*, can also conveniently be interpreted as the average delay time of the sequential procedure. If the regressors $\{\mathbf{x}_i\}$ are constant, $a_m = c_m$, whereas in the time series regressor case, an additional correction term is to be included in the definition of a_m. This implies that generally $a_m \neq c_m$. Note, however, that

$$\frac{a_m}{c_m} \xrightarrow{P} 1 \quad (m \to \infty),$$

where $\xrightarrow{P}$ signifies convergence in probability, and that therefore the influence of the time series regressor fluctuations dampens out in the long-run. Their impact to the asymptotic is solely given by their mean component $\mathbf{d}$ which enters through the term $\delta_m = \mathbf{d}^{\mathrm{T}}\mathbf{\Delta}_m$ used in the definition of c_m. Moreover, b_m represents the standard deviation, determining the order of magnitude of the fluctuations around $\tau_m - a_m$.

Even further, we will show next that τ_m satisfies the weak law of large numbers with normalizing sequence c_m. To prove this, note first that

$$\frac{b_m}{c_m} = \frac{\sigma}{1-\gamma}\frac{1}{\sqrt{c_m}|\delta_m|},$$

using the definitions of b_m and c_m in Theorem 3.1. The first ratio on the right-hand side is clearly constant. For the reciprocal of the second, it follows from assumption (3.4) that

$$\sqrt{c_m}|\delta_m| = \left(qm^{1/2-\gamma}|\delta_m|^{1-2\gamma}\right)^{1/[2(1-\gamma)]} \to \infty \quad (m \to \infty)$$

and consequently $b_m/c_m \to 0$. This combined with the display in the previous paragraph implies that $b_m/a_m \xrightarrow{P} 0$ as $m \to \infty$. According to Theorem 3.1, for all $\epsilon > 0$ there are real numbers z_1 and z_2 such that

$$\lim_{m\to\infty} P\left\{1 + \frac{b_m}{a_m}z_1 \le \frac{\tau_m}{a_m} \le 1 + \frac{b_m}{a_m}z_2\right\} = 1 - \epsilon.$$

Since we have just shown that $b_m/a_m \xrightarrow{P} 0$, we have proved that

$$\frac{\tau_m}{a_m} \xrightarrow{P} 1 \quad \text{and} \quad \frac{\tau_m}{c_m} \xrightarrow{P} 1 \quad (m \to \infty)$$

and therefore verified our claim.

The standard approach in sequential analysis is to determine the average run-length (ARL) of the monitoring procedures under consideration only. Theorem 3.1 offers more. For a fixed m, a_m can be interpreted as approximate ARL with b_m giving the average fluctuations around this average. But with the central limit theorem readily available for τ_m other asymptotic quantiles of interest can easily be computed. To the best of our knowledge, the first contribution in the literature presenting such a limit theorem was Aue and Horváth (2004). Their result was stated in the much simpler and less relevant (univariate) location model. It follows as a special case from our Theorem 3.1 on setting $p = 1$ and $\mathbf{x}_i = 1$. It is also worth mentioning that a similar result for MOSUM-type procedure is not to be expected (see Aue et al. (2008b)).

The idea behind the proof of Theorem 3.1 can be explained as follows. By definition of the stopping rule τ_m, the monitoring procedure will not have been terminated by time N if and only if the detector $\Gamma_m(k)$ has remained below the threshold function $g_m(k)$ for all time indices $k \le N$. Consequently, the probabilities of these two events must be the same, that is

$$P\{\tau_m > N\} = P\left\{\max_{1\le k\le N}\frac{|\Gamma_m(k)|}{g_m(k)} \le 1\right\}. \tag{3.6}$$

The main goal is now to define an appropriate sequence $N = N_m(z)$, related to the sequences $\{a_m\}$, $\{b_m\}$ and $\{c_m\}$, so that the convergence in distribution result of Theorem 3.1 follows. That is, we have to prove that, with $\tau_m^0 = (\tau_m - a_m)/b_m$,

$$\lim_{m\to\infty} P\{\tau_m > N_m(z)\} = \lim_{m\to\infty} P\{\tau_m^0 > -z\} = 1 - \Phi(-z) = \Phi(z)$$

holds for all real z. We relegate the exact definition of N to Section 7.

Eq. (3.6) relates exceedance probabilities of τ_m to probabilities of a maximum over a weighted random partial sum. It is somewhat surprising and unusual that the resulting asymptotic in Theorem 3.1 is normal and not of extreme value type. The reason for this fact is that, asymptotically, only those time indices close to N will contribute to the limit (see Lemmas 7.3–7.6) and that therefore $\Gamma_m(N)/g_m(N)$ contains all relevant information.

There are various other lines of research established in the literature dealing with structural breaks. In contrast to our approach which gives a change-point the role of an additional parameter present under the alternative hypothesis, Pesaran et al. (2006), and Pesaran and Timmermann (2007) use a Bayesian framework for forecasting in a Markov switching model utilizing and extending the methodologies developed in Hamilton (1988), Chib (1998) and others. In assuming more structure on the timing and magnitude of breaks via the so-called meta distributions, the derivation of a delay time as in our main result becomes obsolete and makes the two approaches incompatible.

4. A Monte Carlo study

In this section, we provide a simulation study that considers several data generating processes as time series regressors as well as a selection of values for the tuning parameter γ and the training period m to investigate the finite sample behavior of the delay time approximation given in Theorem 3.1.

4.1. The simulation design

We illustrate the applicability of the theoretical result in a simulation study and attempt to provide some loose guidelines for practitioners that may help to ease the data analysis. The objective of this subsection is to explore the normality of τ_m for the training sample sizes $m = 75, 100, 200$ and 1000 which are reasonable to assume in applications. Throughout, we take the innovations $\{\varepsilon_i\}$ as standard normal random variables. The bivariate regression parameters are assumed to switch from $\boldsymbol{\beta}_0 = (1, 1)^{\mathrm{T}}$ to $\boldsymbol{\beta}_A = (1.2, 1.2)^{\mathrm{T}}$ at the change-point k^*. Other specifications for $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_A$ were also considered with similar outcomes, so that the results are not reported here. To explore further the effect of the change-point location on the monitoring procedure, we have considered three values, namely $k^* = 1$, $\lfloor m/10\rfloor$ and $\lfloor 3m/2\rfloor$, representing early, moderate and late changes. The time series regressors $\{\mathbf{x}_i\}$ always satisfy $x_{i1} = 1$, while the second coordinate is chosen from the following list. The sequence $\{x_{i2}\}$ is given by

(A) iid normal random variables with mean $d_2 = 3$ and unit variance;

(B) a first-order autoregressive process with parameter $\varphi = .3$ and mean $d_2 = 3$, that is

$$x_{i2} - d_2 = \varphi(x_{i-1,2} - d_2) + z_i,$$

where $\{z_i\}$ denotes a sequence of iid zero mean normal random variables with variance .84. The specifications imply that x_{i2} has unit variance;

(C) the generalized autoregressive conditionally heteroscedastic GARCH(1,1) process

$$x_{i2} = d_2 + \sigma_i z_i, \quad \sigma_i^2 = \bar{\omega} + \bar{\alpha}z_{i-1}^2 + \bar{\beta}\sigma_{i-1}^2,$$

with parameter specifications $d_2 = 3$, $\bar{\omega} = .5$, $\bar{\alpha} = .2$ and $\bar{\beta} = .3$. Moreover, $\{z_i\}$ are iid standard normal variates. It is ensured that x_{i2} has unit variance.

The number of replications is 2500. We then analyze the data generating processes given by (A)–(C) with the stopping rules $\tau_m = \tau_m(\gamma)$ with the three specifications $\gamma = 0.00, 0.25$ and 0.45, covering the range of admissible values for the tuning parameter.

Table 2
Performance of the monitoring procedures $\tau_m = \tau_m(\gamma)$ for selected values of m, k^* and γ in the case of the first-order autoregressive regressors specified in (B).

m	k^*	γ	min	Q1	med	Q3	max	m	k^*	γ	min	Q1	med	Q3	max
75	1	0.00	10	22	28	37	130	200	1	0.00	20	36	42	49	117
		0.25	3	11	17	24	750			0.25	5	19	24	31	82
		0.45	1	6	10	18	147			0.45	1	7	14	21	92
	7	0.00	13	29	37	48	223		20	0.00	33	59	69	80	154
		0.25	7	22	28	38	141			0.25	20	46	54	63	163
		0.45	1	18	26	35	331			0.45	1	39	48	58	136
	112	0.00	19	155	179	209	465		300	0.00	64	367	402	437	720
		0.25	12	153	180	208	533			0.25	23	363	398	435	584
		0.45	1	159	187	227	748			0.45	1	373	411	453	643
100	1	0.00	13	28	34	42	112	1000	1	0.00	47	73	82	91	152
		0.25	4	17	23	30	131			0.25	12	33	39	47	85
		0.45	1	8	15	24	113			0.45	1	9	14	21	58
	10	0.00	16	36	43	51	171		100	0.00	130	176	187	200	261
		0.25	8	27	34	41	115			0.25	68	143	154	170	237
		0.45	1	23	30	39	146			0.45	1	129	141	152	214
	150	0.00	30	201	228	262	656		1500	0.00	300	1646	1788	1775	1973
		0.25	15	198	226	262	541			0.25	93	1633	1695	1760	1971
		0.45	1	206	238	278	602			0.45	1	1650	1713	1778	1995

Table 3
Performance of the monitoring procedures $\tau_m = \tau_m(\gamma)$ for selected values of m, k^* and γ in the case of the GARCH(1,1) regressors specified in (C).

m	k^*	γ	min	Q1	med	Q3	max	m	k^*	γ	min	Q1	med	Q3	max
75	1	0.00	10	23	28	35	247	200	1	0.00	16	32	37	45	99
		0.25	4	14	19	26	141			0.25	5	15	19	24	95
		0.45	1	9	14	20	256			0.45	1	7	11	16	70
	7	0.00	14	30	38	47	139		20	0.00	30	55	64	72	138
		0.25	9	22	29	38	179			0.25	10	40	48	57	116
		0.45	1	18	26	36	219			0.45	1	34	41	52	134
	112	0.00	29	159	184	217	531		300	0.00	78	371	405	443	619
		0.25	9	157	183	216	558			0.25	29	365	399	438	454
		0.45	1	168	197	240	750			0.45	1	375	415	454	732
100	1	0.00	12	25	31	38	116	1000	1	0.00	51	76	84	94	131
		0.25	4	13	18	25	105			0.25	12	33	39	47	84
		0.45	1	9	12	19	116			0.45	1	11	17	23	76
	10	0.00	17	36	43	53	131		100	0.00	134	182	196	210	288
		0.25	10	25	32	42	161			0.25	103	146	160	177	242
		0.45	1	21	29	38	173			0.45	3	132	144	159	219
	150	0.00	30	200	228	262	500		1500	0.00	339	1649	1702	1763	2008
		0.25	16	197	226	261	459			0.25	113	1634	1693	1755	1988
		0.45	1	204	237	274	606			0.45	1	1651	1711	1776	2017

4.2. Simulation results

The findings of the simulation experiments are reported in Tables 1–3 and Fig. 1. The tables give five number summaries for the various specifications of $\tau_m = \tau_m(\gamma)$ and the three data generating processes (A)–(C). It can be seen that the performance of the monitoring procedures appears very robust with respect to weak dependence and heteroskedasticity in the time series regressors, as the corresponding numerical values in Tables 2 and 3 do not differ significantly from the ones for the iid normal regressors in Table 1.

Second, all procedures perform the worse the later the change occurs after the monitoring commences. This confirms that the condition imposed in (3.5) is not restrictive but necessary, since the asymptotic normality provided in Theorem 3.1 is apparently violated, even for the training period of size $m = 1000$, if $k^* = \lfloor 3m/2 \rfloor$.

Third, there is a strong dependence between the performance of the monitoring procedure and the value of γ chosen to construct the boundary function $g_m(k)$. Selecting $\gamma = 0.00$ yields a rather conservative procedure that requires a comparatively large detection time before rejecting the stability hypothesis. On the other hand, it avoids false early detection which can be seen best from the late change scenario $k^* = \lfloor 3m/2 \rfloor$. The procedure specified by the tuning parameter $\gamma = 0.45$ reacts fastest to the changes but at the cost of an increased false early detection. The procedure with $\gamma = 0.25$ can be interpreted as a compromise between the other two procedures. In general, one would recommend the use of a larger γ in our setting which requires an early change through (3.5), while the recommendation for late changes (if one would use the CUSUM procedure at all) would be to use a small tuning parameter γ. These observations are in accordance with the theoretical findings of Theorem 3.1. They also extend the theoretical and empirical results reported in Horváth et al. (2004) and Aue et al. (2006b).

Fig. 1 illustrates the findings graphically for the GARCH(1,1) model (C). It shows the estimated densities of $\tau_m^0 = (\tau_m - a_m)/b_m$ for the four choices of m and three choices of k^*. The normalizing sequences have been calculated from their definitions in Theorem 3.1. Essentially the same graphs are obtained if instead of the theoretical values the sample mean and variance are used and we refrain from including the graphs. The corresponding graphs for the data generating processes (A) and (B) are similar and hence not included here. Fig. 1 shows that normality is reasonably approached for the early change scenario, especially in case of the large training period with $m = 1000$. Compared to the superimposed standard normal density, there is in all cases some mass missing in the left tail and there is too much mass in the right tail, which is visible also for $m = 1000$. This, however, is to be expected, since τ_m cannot be negative but could, on the other hand, theoretically take an infinite value. For the purpose of constructing reliable confidence intervals for the delay time, Theorem 3.1 provides excellent interval endpoints if γ is sufficiently large and if the change

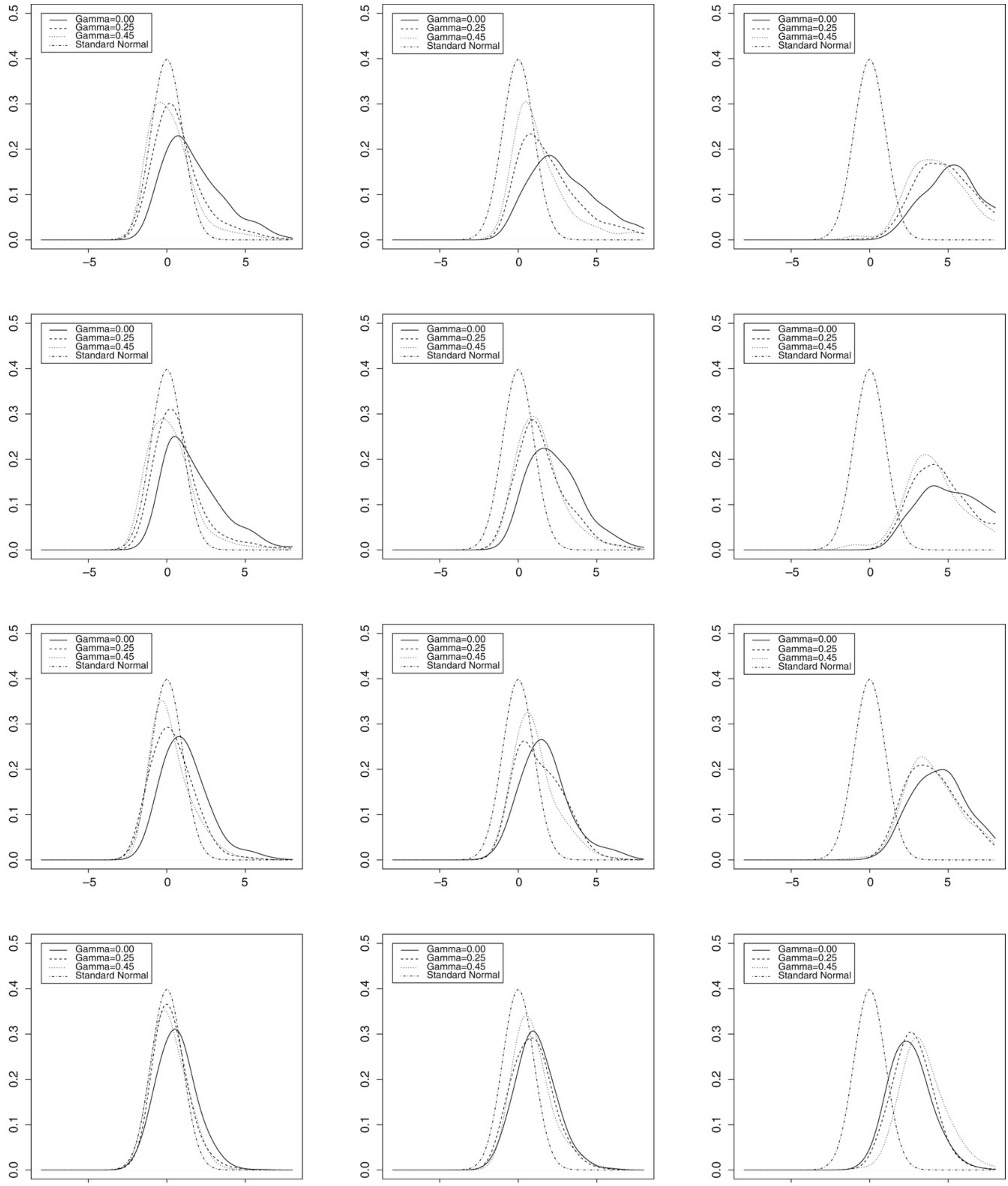

Fig. 1. Estimated density plots for the standardized delay times $\tau_m^0 = \tau_m^0(\gamma)$ using the GARCH(1,1) regressors in (B) and the tuning parameters $\gamma = 0.00, 0.25$ and 0.45. The rows from top to bottom correspond to the respective training period sizes $m = 75, 100, 200$ and 1000. The columns from left to right correspond to the respective change-point locations $k^* = 1, 5$ and 20. The standard normal density is superimposed for comparison.

happens early. Moving from left to right in a given row, it is apparent how later changes distort the normal asymptotic to the extent that it loses its validity if the value of k^* is further increased. Moving from top to bottom in a given column, one can see how an increased training sample improves the bell shape of the density estimator. For more on this topic, we refer to Aue et al. (2008a).

5. An application to macroeconomic data

5.1. Data description

It is a common understanding in economics that consumption, labor income and household wealth should move together in the long run. Following the pioneering work of Modigliani (1971), almost all contemporary macroeconomic models include the assumption that a dollar increase in wealth (for fixed labor income) implies an increase in consumer spending of roughly four to five cents (see, for instance, the textbook Abel and Bernanke (2001), and the two papers Lettau and Ludvigson (2001, 2004).

To explore the connection between household wealth and consumption in the United States on one side and the predictability of stock returns on the other, the latter authors compiled a data set[1] of 224 4-variate measurements taken quarterly from 1952:1 to 2007:4 of which we used the 200 observations in the period 1958:1–2007:4 for our purposes. The measurements include the variables consumption, asset wealth, labor income and an error correction term due to cointegration. Consumption is measured as total personal consumption expenditures on non-durables and services in billions of dollars, excluding shoes and clothing. Asset wealth is measured as household net worth in billions of dollars. Labor income is described as the annuity value of human wealth. To ensure stationarity, we apply a log difference transformation to the consumption, asset wealth and labor income variables. We shall use the following notation for the quantities involved to built our time series regression model. Let

$y_i =$ log difference of the consumption observations,

$x_{i1} =$ intercept term,

$x_{i2} =$ log difference of the asset wealth observations,

$x_{i3} =$ log difference of the labor income observations,

$x_{i4} =$ the error correction terms.

In the following, we investigate if the assumed fixed linear relation between wealth and consumption is indeed justified via an application of the linear model setting introduced in Section 2. Since we have all data points available at the outset, it is possible to analyze the structural stability with retrospective methods to get a better picture of the data. We pursue this goal in Section 5.2, before we design an appropriate monitoring procedure in Section 5.3 to highlight the usefulness of Theorem 3.1 in a practical environment.

5.2. A preliminary retrospective analysis

In this subsection, we apply the maximally selected likelihood ratio test in the form of Csörgő and Horváth (1997) to check whether or not the regression parameters in the linear regression model $y_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_0 + \varepsilon_i$ is justified for the whole set of observations ($i = 1, \ldots, 200$). Since the data may be confounded by more than one structural break, we adopt the binary segmentation procedure which is similar to the methods discussed in Bai and Perron (1998) and Bai (1999). To this end, denote by

$$\ell_n = -n \max_{p<k<n-p} \left[\ln \hat{\sigma}_k^2 - \ln \hat{\sigma}_n^2\right],$$

the maximally selected log likelihood ratio test statistic, where $\hat{\sigma}_k^2 = \hat{\sigma}_{k,1}^2 + \hat{\sigma}_{k,2}^2$ with $\hat{\sigma}_{k,1}^2$, $\hat{\sigma}_{k,2}^2$ and $\hat{\sigma}_n^2$ being respectively the variance estimator of the residuals from regression fits to the first k, the last k and all n observations. Utilizing Theorem 3.1.1 of Csörgő and Horváth (1997), the critical value for a given asymptotic level $\alpha \in (0, 1)$ of the test can be computed as in their display (3.1.21). Since the application of this procedure is by now standard, we only briefly report the results in the following paragraph.

In the present context, we have used the level $\alpha = 0.01$ in each of the applications of the likelihood ratio test. We have consequently in a first step applied $\ell_{200}^{(1)} = \ell_{200}$ with $p = 4$ to the data points. We have thus found a first change in the regression parameters at time $k_1^* = 123$ (that is, 1988:2). The segmentation procedure requires now to split up the sample into the two subsamples, the first of which contains the initial 123 observations and the second of which the final 77 observations, and to repeat the analysis on the subsamples. Since we are here only interested in the application of the monitoring procedure defined by τ_m, we have only used the first 123 observations. An application of $\ell_{123}^{(2)} = \ell_{123}$ has led to the detection of another change-point at time $k_2^* = 92$ (that is, 1980:3). A third application of the segmentation procedure to the thus obtained subsamples has not found additional structural breaks.

The results of the retrospective analysis are graphically summarized in Fig. 2. The panels show the four time series involved in the regression model building process, namely consumption in the upper left panel, asset wealth (upper right), labor income (lower left) and the correction terms (lower right). Each time series plot depicts k_1^* and k_2^* as dotted lines to indicate the structural breaks detected here.

5.3. Sequential monitoring

In order to illustrate our theory, we apply now the stopping rule τ_m for several choices of the training period size m to the same data set. Since the retrospective analysis indicated two structural breaks in 1980:3 and 1988:2 but structural stability prior to 1980:3, we use the four training periods

$m = 75$, 1958:1–1976:3;

$m = 80$, 1958:1–1977:4;

$m = 85$, 1958:1–1979:1;

$m = 90$, 1958:1–1980:2.

Moreover, for each m, we have selected three values of the boundary function tuning parameter, namely $\gamma = 0$, 0.25 and 0.45. These choices determine now 12 corresponding monitoring procedures given by

$$\tau_m = \tau_m(\gamma), \quad m = 75, 80, 85, 90, \ \gamma = 0, 0.25, 0.45,$$

which will be utilized to see how the parameters m and γ affect the stopping times. At the same time, we can gain valuable information also on how the start of the monitoring is related to the location of the underlying structural break which we expect to happen around the time $k_2^* = 92$ (1980:3) according to the retrospective analysis of Section 5.2. The training period with $m = 90$ ends notably exactly one observation before the estimated time of break.

Given a training period, we started the model building process by fitting the parameters to the multiple regressions

$$y_i = (x_{i1}, x_{i2}, x_{i3}, x_{i4})(\beta_{01}, \beta_{02}, \beta_{03}, \beta_{04})^{\mathrm{T}} + \varepsilon_i, \quad i = 1, \ldots, m, \ m = 75, 80, 85, 90,$$

where the notation of Section 5.1 has been adopted. Utilizing the theoretical contribution of this paper, we started monitoring for changes in the regression parameters in the remaining $200 - m$ observations with the stopping rules $\tau_m = \tau_m(\gamma)$ for the three different values of the tuning parameter given above. The corresponding observed stopping times τ_m, the critical values and the estimated values $\hat{a}_m$, $\hat{b}_m$ and $\hat{c}_m$ are summarized in Table 4, while Fig. 3 depicts the detectors' absolute values $|\Gamma_m(k)|$ for each value of m. In each of these plots, the three boundary functions are superimposed. As the asymptotic level of the test we have chosen $\alpha = 0.1$. Simulations in Horváth et al. (2004) have, however, indicated that the test procedures perform rather conservative for sizes of the training period less than or equal to 100 and suggested that one trims the asymptotic critical values by 10% in an effort

[1] The data can be downloaded at http://faculty.haas.berkeley.edu/lettau/data/cay_q_07Q4.txt.

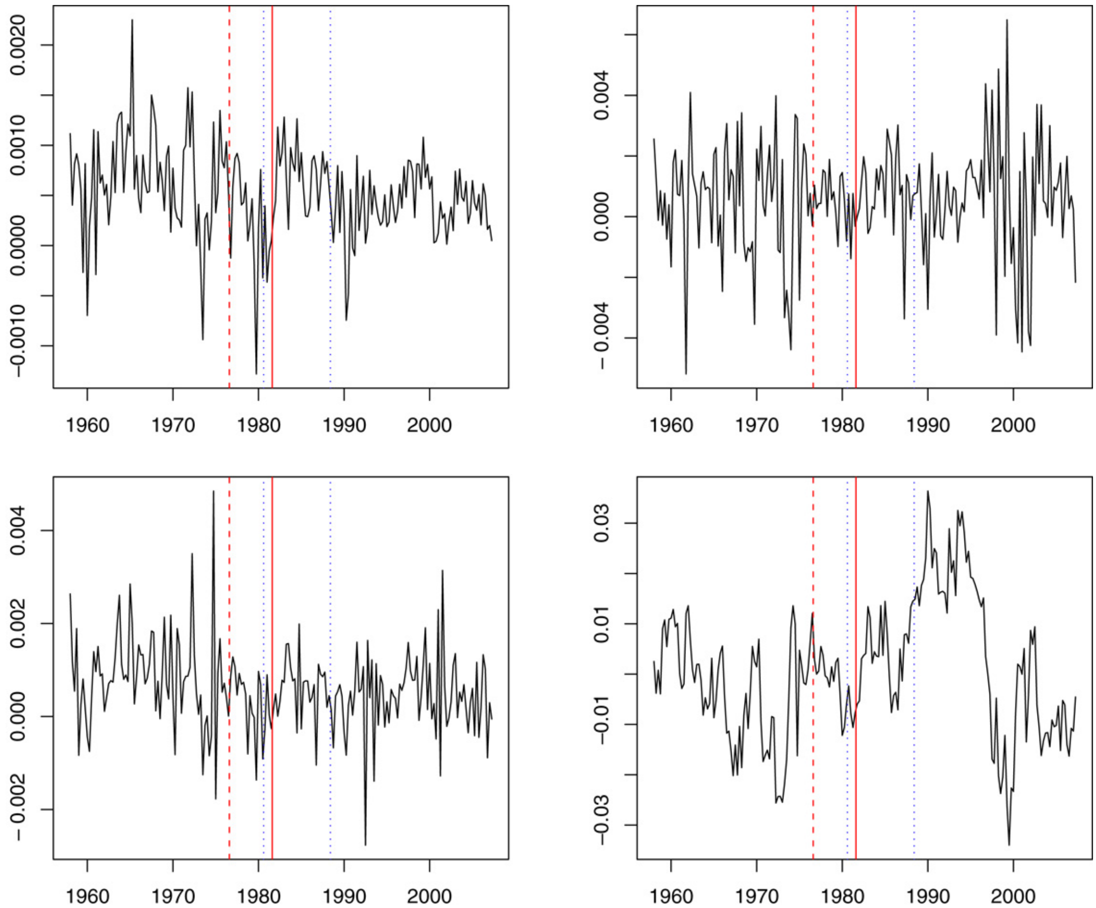

Fig. 2. Log differences of consumption expenditures (upper panel, left), asset wealth (upper panel, right) and labor income (lower panel, left). The error correction terms are plotted in the lower right panel. Each plot depicts the structural changes in the assumed linear relationship detected with the retrospective procedure as dotted lines. Each plot also displays the end of the monitoring period with $m = 75$ as dashed line and the detection of the structural break as solid line for the sequential monitoring procedure with $\gamma = 0.25$.

Table 4
Summary of the nine monitoring procedures given by the training period sizes $m = 75, 80, 85$ and the tuning parameters $\gamma = 0.00, 0.25, 0.45$. The critical values used for the respective values of γ are $q^* = 1.74, 1.90, 2.30$ and are taken from Horváth et al. (2004) subject to an empirical correction.

m	γ	τ_m[date]
75	0.00	66 [1994:1]
	0.25	20 [1982:3]
	0.40	19 [1982:1]
80	0.00	66 [1995:2]
	0.25	15 [1981:3]
	0.45	10 [1980:2]
85	0.00	63 [1995:4]
	0.25	11 [1981:4]
	0.45	4 [1980:1]

to improve the finite sample behavior. We have followed this recommendation here.

Several remarks are in order. First, the monitoring procedure based on the tuning parameter $\gamma = 0$ proves to be too conservative to pick up a structural break in the early 1980s for all choices of m. Unless $m = 90$, however, they reject the null hypothesis of structural stability in the mid-1990s. Given that CUSUM procedures need time to pick up changes that occur later in the sample, it can be argued that, for $m = 75, 80, 85$, the stopping times $\tau_m = \tau_m(0)$ detect the chronologically second change $k_1^* = 123$ [1988:2] indicated by the retrospective analysis. This claim may be further supported by the fact that k_1^* is the more significant break as it was found in the first round of the binary segmentation procedure of Section 5.2.

Second, for $m = 75, 80, 85$, the more sensitive monitoring procedures based on the tuning parameters $\gamma = 0.25, 0.45$ detect a change in the regression parameters in the early 1980s as expected from the retrospective test. That the larger γ reacts faster is in accordance with the findings of the simulation study in Section 4.

Third, the procedures based on $m = 90$ fail to produce a rejection of the structural stability null hypothesis (see the lower right panel of Fig. 3). A possible explanation is the following. Given that the values for the stopping rules $\tau_m = \tau_m(\gamma)$ with $m = 75, 80, 85$, and $\gamma = 0.25, 0.45$ all stop right after or even before k_2^* [1980:3], it seems likely that the actual change happened prior to this estimated break date. An inspection of the log likelihood function (not shown here) reveals that the second highest value of $\ln \hat{\sigma}_k^2 - \ln \hat{\sigma}_n^2$, $k = 5, \ldots, 195$, is reached close to but prior to k_2^*, thus adding to the credibility of our assessment.

In conclusion, we have through the application of our monitoring procedure provided evidence that the standard textbook assumption of parallel growth of consumption and wealth seems questionable. The result shows that suitably calibrated versions of τ_m with values of γ ranging between 0.25 and 0.45 yield reliable tools for practitioners who are interested in real-time monitoring of economic data. The procedures appear to be reasonably robust to the size of the training period and the location of the structural break date relative to the time when monitoring commences. Our findings are additionally supported by a preliminary retrospective analysis.

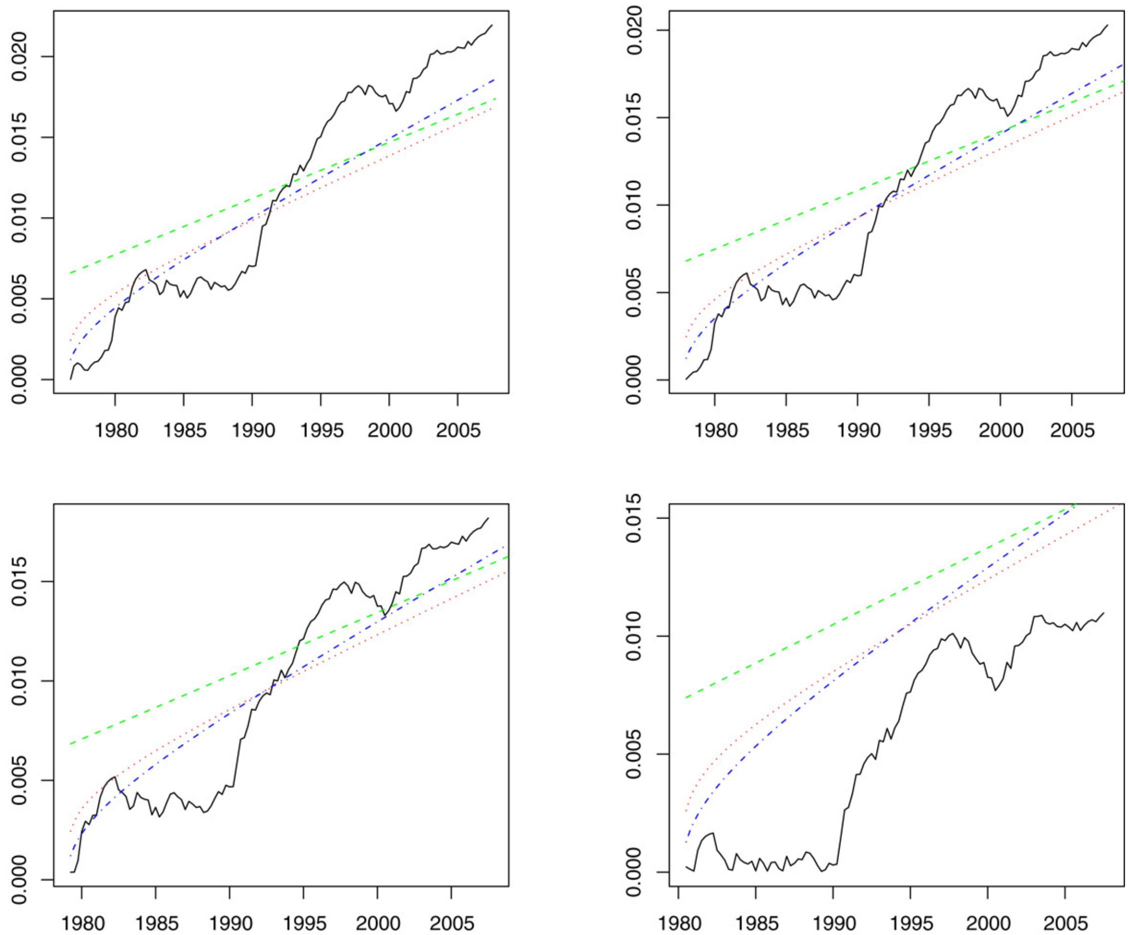

Fig. 3. The performance of the stopping rules $\tau_m = \tau_m(\gamma)$ using training sizes $m = 75$ (upper panel, left), $m = 80$ (upper panel, right), $m = 85$ (lower panel, left) and $m = 90$ (lower panel, right). Each plot depicts the detector values as solid lines (—). The different boundary functions are superimposed with $\gamma = 0$ given by a dashed line (- - -), $\gamma = 0.25$ by a dotted line $(\cdots)$ and $\gamma = 0.45$ by (·- ·). Alarm is caused at the first time the detector exceeds the boundary function value.

6. Conclusions

We have in this paper established the asymptotic normality of delay times associated with a CUSUM-type monitoring procedure in a multiple time series regression setting. This procedure compares the growth of the model residuals with a boundary function that can be tuned via the selection of a sensitivity parameter. The results are given under general and novel conditions. They are of interest in itself but also offer a variety of new tools for econometric practitioners who are interested in the on-line and real-time monitoring of economic and financial data.

The monitoring procedures have been analyzed by means of a Monte Carlo simulation study to assess their finite sample behavior in case of a change of moderate size. We have tested the monitoring procedures on a real macroeconomic data set by first evaluating the whole data set with the help of retrospective tools. These findings were reproduced by an application of the sequential procedure.

A number of recommendations can be made for applied econometricians. These concern mainly the two parameters that determine the performance of the procedures, namely the training period size m and the sensitivity tuning parameter γ. First, the results of Section 5 show that the monitoring works reasonably well also for training period sizes somewhat less than 100 as long as the sensitivity parameter γ is chosen to be at least 0.25 or larger. In the case of even shorter training periods, practitioners may use the techniques described after display (2.1) to reach an appropriate length of, say, $m = 75$ observations.

To ensure that the procedures have reasonable power against detecting changes of moderate size $|\Delta_m|$ (as the ones used in Section 4, where $|\Delta_m| = 0.2$), it is recommended to apply a monitoring procedure with a tuning parameter $\gamma \geq 0.4$. Picking a larger γ generally leads to a more powerful procedure. More pronounced changes are obviously easier detected and allow consequently also for choosing smaller values of γ. Procedures with smaller γ generally behave nicer under the null hypothesis of no change in the regression parameters (which has not been considered in this paper), as they are less prone to cause false alarms.

The results presented here are valid for CUSUM-type procedures that require early changes to ensure optimal performance. If the change, however, occurs late after the monitoring has commenced, the procedures of this paper fail to perform at a satisfactory level. Research to be conducted in the future will focus thus on establishing modified results also for MOSUM-type procedures which generally outperform CUSUM-type procedures for late changes. The technical difficulties that are to be expected have been addressed briefly in the Introduction and in Section 3.

7. Proofs

The proof section is divided into a number of subsections. In Section 7.1, we deal with the time series regressors $\{\mathbf{x}_i\}$ and show how the moment assumption in (2.5) can be utilized to obtain an order of magnitude for deviations of the $\mathbf{x}_i$ from their mean $\mathbf{d}$, and that (2.5) is satisfied for the sequences $\{\mathbf{x}_i\}$ introduced in Examples 2.1–2.4. In Section 7.2, we compare orders of magnitude of the change-point k^*, the training period size m, the sequence N and δ_m. Finally, the limit distribution is derived in Section 7.3.

Throughout, we assume without loss of generality that $\delta_m = \mathbf{d}^T\boldsymbol{\Delta}_m > 0$.

Right at the outset, we specify N and relate it to the normalizing sequences $\{a_m\}$, $\{b_m\}$ and $\{c_m\}$. Let $N = N_m(x)$ be defined via

$$\begin{aligned} N^{1-\gamma} &= \frac{qm^{1/2-\gamma}}{\delta_m} - \frac{1}{c_m^\gamma \delta_m}\sum_{i=m+k^*}^{m+c_m}(\mathbf{x}_i - \mathbf{d})^T\boldsymbol{\Delta}_m \\ &\quad - \sigma x\left(\frac{q^{1/2-\gamma}m^{(1/2-\gamma)^2}}{\delta_m^{3/2-2\gamma}}\right)^{1/(1-\gamma)} \\ &= a_m^{1-\gamma} - x\frac{(1-\gamma)b_m}{c_m^\gamma}, \end{aligned} \tag{7.1}$$

where the second equality follows after elementary calculations using the definitions of a_m, b_m and c_m in Theorem 3.1. A similar definition of N was used by Aue and Horváth (2004) to study the simpler location model setting. The main difference between their sequence and the one used in the present paper is the additional partial sum term in (7.1) which is due to the non-constant time series regressors $\{\mathbf{x}_i\}$.

7.1. Properties of the time series regressors

In this subsection we derive the almost sure order of the magnitude of the maximum of the partial sums obtained from the centered regressors $\mathbf{x}_i - \mathbf{d}$. Subsequently, we will show that the time series specified in Examples 2.1–2.4 satisfy the coordinate-wise moment condition (2.5) and, consequently, also the following proposition.

Proposition 7.1. *If assumptions (2.4) and (2.5) are satisfied, then, for all $j = 1, \ldots, p$ as $m \to \infty$,*

$$\max_{k^*\le k\le c_m}\left|\sum_{i=m+k^*}^{m+k}(x_{ij} - d_j)\right| = o\left(\sqrt{c_m}\left[m(\log m)^2\right]^{1/\mu}\right) \quad a.s. \tag{7.2}$$

and

$$\max_{1\le k\le k^*}\frac{1}{k}\left|\sum_{i=m+1}^{m+k}(x_{ij} - d_j)\right| = o\left(\left[m(\log m)^2\right]^{1/\mu}\right) \quad a.s. \tag{7.3}$$

Proof. By the stationarity assumption (2.4), it holds for any $u > 0$ and $j = 1, \ldots, p$,

$$\begin{aligned} &P\left\{\max_{k^*\le k\le c_m}\left|\sum_{i=m+k^*}^{m+k}(x_{ij} - d_j)\right| > u\right\} \\ &\le P\left\{\max_{1\le k\le c_m}\left|\sum_{i=1}^{k}(x_{ij} - d_j)\right| > u\right\}. \end{aligned}$$

Combining the stationarity of the regressors with the moment condition (2.5) yields that

$$E\left|\sum_{i=\ell}^{k}(x_{\ell j} - d_j)\right|^\mu \le K(k-\ell)^{\mu/2},$$

so Theorem 12.2 of Billingsley (1968, p. 94) implies that, for any $t > 0$, there is a constant $K^* > 0$ such that

$$\begin{aligned} &P\left\{\max_{k^*\le k\le c_m}\left|\sum_{i=m+k^*}^{m+k}(x_{ij} - d_j)\right| > t\sqrt{c_m}\left[m(\log m)^2\right]^{1/\mu}\right\} \\ &\le \frac{K^* c_m^{\mu/2}}{\left(\sqrt{c_m}\left[m(\log m)^2\right]^{1/\mu}\right)^\mu} = \frac{K^*}{m(\log m)^2}, \end{aligned}$$

and (7.2) follows from the Borel–Cantelli lemma.

Similarly, using assumptions (2.4) and (2.5), we obtain for (7.3) that

$$\begin{aligned} &P\left\{\max_{1\le k\le k^*}\frac{1}{k}\left|\sum_{i=m+1}^{m+k}(x_{ij} - d_j)\right| > \left[m(\log m)^2\right]^{1/\mu}\right\} \\ &= P\left\{\max_{1\le k\le k^*}\frac{1}{k}\left|\sum_{i=1}^{k}(x_{ij} - d_j)\right| > \left[m(\log m)^2\right]^{1/\mu}\right\} \\ &\le \sum_{k=1}^{k^*}P\left\{\left|\sum_{i=1}^{k}(x_{ij} - d_j)\right| > k\left[m(\log m)^2\right]^{1/\mu}\right\} \\ &= \sum_{k=1}^{k^*}P\left\{\left|\sum_{i=1}^{k}(x_{ij} - d_j)\right|^\mu > \left(k\left[m(\log m)^2\right]^{1/\mu}\right)^\mu\right\} \\ &\le \sum_{k=1}^{k^*}\frac{k^{-\mu/2}}{m(\log m)^2} \\ &= \mathcal{O}\left(\frac{(k^*)^{1-\mu/2}}{m(\log m)^2}\right). \end{aligned}$$

Since $\mu > 2$, the last expression is summable with respect to m and the Borel–Cantelli lemma finishes the proof. $\square$

The second proposition first establishes that N and c_m are asymptotically equivalent with probability one. In a second part, two almost sure convergence results are derived which are similar to the ones studied in the previous proposition. However, the maximum now has an upper limit N.

Proposition 7.2. *If assumptions (2.4) and (2.5) are satisfied and if $\mu > 8(1-\gamma)/(1-2\gamma)$, then, as $m \to \infty$,*

$$\frac{N}{c_m} \to 1 \quad a.s., \tag{7.4}$$

$$\max_{k^*\le k\le N}\frac{1}{k^\gamma}\left|\sum_{i=m+k^*}^{m+k}(\mathbf{x}_i - \mathbf{d})^T\boldsymbol{\Delta}_m\right| = o\left(\delta_m N^{1-\gamma}\right) \quad a.s. \tag{7.5}$$

and

$$\max_{k^*\le k\le N}\frac{1}{k^\gamma}\left|\sum_{i=m+k^*}^{m+k}(\mathbf{x}_i - \mathbf{d})\right| = o\left(N^{1-\gamma}\right) \quad a.s. \tag{7.6}$$

Proof. To derive (7.4), divide N by c_m to obtain

$$\left(\frac{N}{c_m}\right)^{1-\gamma} = 1 - \frac{1}{c_m\delta_m}\sum_{i=m+k^*}^{m+c_m}(\mathbf{x}_i - \mathbf{d})^T\boldsymbol{\Delta}_m - \frac{\sigma x}{\sqrt{c_m}\delta_m}, \tag{7.7}$$

where the second term on the right-hand side is $o(1)$ almost surely by (7.2) of Proposition 7.1 and the definition of c_m in Theorem 3.1. Indeed, observe that

$$\begin{aligned} &\frac{1}{c_m\delta_m}\sum_{i=m+k^*}^{m+c_m}(\mathbf{x}_i - \mathbf{d})^T\boldsymbol{\Delta}_m = \mathcal{O}\left(\frac{1}{c_m\delta_m}\sqrt{c_m}\left[m(\log m)^2\right]^{1/\mu}|\boldsymbol{\Delta}_m|\right) \\ &= \mathcal{O}\left(\delta_m^{1/(2[1-\gamma])}m^{1/\mu-(1-2\gamma)/(2[1-\gamma])}(\log m)^{2/\mu}\right) \\ &= o(1) \end{aligned}$$

holds with probability one as $m \to \infty$, since $\mu > 8(1-\gamma)/(1-2\gamma) > 2+2\gamma/(1-2\gamma)$. The definition of c_m in Theorem 3.1 implies that the third term on the right-hand side of (7.7) is $o(1)$.

By (7.4) it suffices to show the assertion of (7.5) if N is replaced by c_m. Now arguments similar to the ones used in the proof of

Proposition 7.1 imply that, as $m \to \infty$,

$$P\left\{\max_{k^* \le k \le 2c_m} \frac{1}{k^\gamma}\left|\sum_{i=m+k^*}^{m+k}(\mathbf{x}_i - \mathbf{d})^{\mathrm{T}}\boldsymbol{\Delta}_m\right| > t\delta_m c_m^{1-\gamma}\right\}$$
$$\le \sum_{i=k^*}^{2c_m} P\left\{\left|\sum_{i=m+k^*}^{m+k}(\mathbf{x}_i - \mathbf{d})^{\mathrm{T}}\boldsymbol{\Delta}_m\right| > tk^\gamma\delta_m c_m^{1-\gamma}\right\}$$
$$\le K_1\sum_{k=k^*}^{2c_m}(tk^\gamma\delta_m c_m^{1-\gamma})^{-\mu}(k-k^*)^{\mu/2}|\boldsymbol{\Delta}_m|^\mu$$
$$\le \frac{K_2}{(tc_m^{1-\gamma})^\mu}\sum_{k=1}^{2c_m} k^{\mu(1/2-\gamma)}$$
$$\le \frac{K_3}{t^\mu}c_m^{1-\mu/2}$$
$$\le \frac{K_4}{t^\mu}\left(\frac{m^{1/2-\gamma}}{\delta_m}\right)^{(1-\mu/2)/(1-\gamma)}$$
$$\le \frac{K_5}{t^\mu}(\log m)^{(\mu/2-1)/(1-\gamma)}m^{(1-\mu/2)(1-2\gamma)/(2[1-\gamma])},$$

which is summable since $\mu > 8(1-\gamma)/(1-2\gamma)$. Consequently, (7.5) follows from the Borel–Cantelli lemma. The claim in (7.6) is proved in the same fashion. □

Propositions 7.1 and 7.2 ensure that the deviations $\mathbf{x}_i - \mathbf{d}$ remain so small uniformly that their impact does not influence the limit behavior of τ_m. In particular, both propositions are satisfied for the classical constant regressors $\mathbf{x}_i = \mathbf{d}$. Note that we only needed to impose stationarity and the moment condition (2.5) and that more restrictive assumptions (such as the existence of a density or mixing conditions) that would limit the applicability of our approach or would be harder to verify can be avoided.

The subsection is concluded by studying the specific $\{\mathbf{x}_i\}$ of Examples 2.1–2.4.

Proposition 7.3. *Under the conditions of Example 2.1, the resulting iid sequence $\{\mathbf{x}_i\}$ satisfies (2.5).*

Proof. It follows from Rosenthal's inequality (see Petrov (1995), p. 59). □

Proposition 7.4. *If the conditions of Example 2.2 are satisfied, then the resulting multivariate linear process $\{\mathbf{x}_i\}$ satisfies (2.5).*

Proof. An application of Minkowski's inequality yields that

$$E\left|\sum_{i=1}^k(x_{ij}-d_j)\right|^\mu = E\left|\sum_{i=1}^k\sum_{\ell=0}^\infty \varphi_\ell(j)z_{i-\ell}\right|^\mu$$
$$= E\left|\sum_{\ell=0}^\infty \varphi_\ell(j)\sum_{i=1}^k z_{i-\ell}\right|^\mu$$
$$\le \left(\sum_{\ell=0}^\infty\left[E\left|\varphi_\ell(j)\sum_{i=1}^k z_{i-\ell}\right|^\mu\right]^{1/\mu}\right)^\mu$$
$$\le \left(\sum_{\ell=0}^\infty|\varphi_\ell(j)|\left[E\left|\sum_{i=1}^k z_{i-\ell}\right|^\mu\right]^{1/\mu}\right)^\mu.$$

Now, Rosenthal's inequality (see Petrov (1995), p. 59) gives

$$E\left|\sum_{i=1}^k z_{i-\ell}\right|^\mu \le K_1 k^{\mu/2} \quad \text{for all } \ell,$$

where $K_1 > 0$ is a constant, and therefore, with two further constants $K_2, K_3 > 0$,

$$E\left|\sum_{i=1}^k(x_{ij}-d_j)\right|^\mu \le K_2 k^{\mu/2}\sum_{\ell=0}^\infty|\varphi_\ell(j)| \le K_3 k^{\mu/2},$$

since the sequence $\{\varphi_\ell(j)\}$ is absolutely summable. This completes the proof. □

Proposition 7.5. *If the conditions of Example 2.3 are satisfied, then the resulting multivariate GARCH-type sequence $\{\mathbf{x}_i\}$ satisfies (2.5).*

Proof. Note that $\{\sum_{i=1}^k x_{ij}, \sigma(z_i, i < k)\}$ is a martingale, so Rosenthal's inequality in the form given by Hall and Heyde (1980, p. 23) implies that

$$E\left|\sum_{i=1}^k(x_{ij}-d_j)\right|^\mu \le K_1\left(E\left[\sum_{i=1}^k\sigma_i^2 Ez_i^2\right]^{\mu/2} + \sum_{i=1}^k E|\sigma_i z_i|^\mu\right)$$
$$\le K_2\left(E\left[\sum_{i=1}^k\sigma_i^2\right]^{\mu/2} + k\right)$$

with two constants $K_1, K_2 > 0$. Since, by Minkowski's inequality with some $K_3 > 0$,

$$\left(E\left[\sum_{i=1}^k\sigma_i^2\right]^{\mu/2}\right)^{2/\mu} \le \sum_{i=1}^k(E|\sigma_i|^\mu)^{\mu/2} \le K_3 k,$$

the proof of Proposition 7.5 is complete. □

Proposition 7.6. *If the conditions of Example 2.4 are satisfied, then the resulting linear process with GARCH-type innovations $\{\mathbf{x}_i\}$ satisfies (2.5).*

Proof. By Proposition 7.5, it holds that, for some constant $K > 0$,

$$E\left|\sum_{i=1}^k z_i\right|^\mu \le Kk^{\mu/2},$$

so that the assertion follows along the lines of the proof of Proposition 7.3. □

7.2. Preliminary estimates

In view of the previous subsection, from now on only a fixed realization of $\{\mathbf{x}_i\}$ is considered which satisfies the results of Propositions 7.1 and 7.2. In the following, we compare the orders of magnitude of m, k^*, δ_m and N in two lemmas. Recall the definition of N in (7.1).

Lemma 7.1. (i) *If (3.5) holds, then $k^*/m \to 0$ as $m \to \infty$.*
(ii) *If conditions (2.4), (2.5) and (3.4) are satisfied, then $N/m \to 0$ as $m \to \infty$.*
(iii) *If conditions (2.4), (2.5) and (3.4) are satisfied, then $\sqrt{N}\delta_m \to \infty$ as $m \to \infty$.*
(iv) *If conditions (2.4), (2.5), (3.4) and (3.5) are satisfied, then $k^*/\sqrt{N} \to 0$ as $m \to \infty$.*

Proof. (i) By assumption (3.5), $k^* = \mathcal{O}(m^\theta)$. Since $\theta < 1$, the assertion follows readily.

(ii) It follows from (7.4) that instead of N it suffices to prove the assertion for c_m. It holds then that

$$\left(\frac{c_m}{m}\right)^{1-\gamma} = \frac{q}{\sqrt{m}\delta_m} \le \frac{q\log m}{C_1\sqrt{m}} = o(1) \quad (m \to \infty)$$

on account of assumption (3.4).

(iii) Applying assumption (3.4) it can be verified that

$$(\sqrt{c_m}\delta_m)^{2(1-\gamma)} = \frac{qm^{1/2-\gamma}}{\delta_m^{3-2\gamma}} \ge \frac{qm^{1/2-\gamma}}{C_2^{3-2\gamma}} \to \infty \quad (m \to \infty).$$

In view of (7.4), the proof is complete.

(iv) On account of (7.4), we give the proof only for c_m again. Utilizing (3.4), (3.5) and the definition of c_m, we get that, as $m \to \infty$,

$$\frac{k^*}{\sqrt{c_m}} = \mathcal{O}\left(m^{\theta-(1-2\gamma)/(4[1-\gamma])}\right) = o(1).$$

Thus, by (7.4), part (iv) of Lemma 7.1 is proved. □

Lemma 7.2. *If conditions (2.4) and (2.5) are satisfied, then, for all real x,*

$$\lim_{m\to\infty} \frac{1}{\sigma}\left(\frac{N}{m}\right)^{\gamma-1/2}\left(q - \frac{1}{\sqrt{m}(N/m)^\gamma}\sum_{i=m+k^*}^{m+N}\mathbf{x}_i^{\mathrm{T}}\mathbf{\Delta}_m\right) = x.$$

Proof. We show first that it suffices to prove the claim for the term

$$x + \frac{N^{\gamma-1/2}}{\sigma}\left(\frac{1}{c_m^\gamma}\sum_{i=m+k^*}^{m+c_m}(\mathbf{x}_i-\mathbf{d})^{\mathrm{T}}\mathbf{\Delta}_m - \frac{1}{N^\gamma}\sum_{i=m+k^*}^{m+N}(\mathbf{x}_i-\mathbf{d})^{\mathrm{T}}\mathbf{\Delta}_m\right). \tag{7.8}$$

Indeed, it holds for $m \to \infty$ that

$$\begin{aligned}
&\left(\frac{N}{m}\right)^{\gamma-1/2}\frac{1}{\sqrt{m}(N/m)^\gamma}\sum_{i=m+k^*}^{m+N}\mathbf{x}_i^{\mathrm{T}}\mathbf{\Delta}_m \\
&= \left(\frac{N}{m}\right)^{\gamma-1/2}\left(\frac{(N-k^*+1)\delta_m}{\sqrt{m}(N/m)^\gamma} + \frac{1}{\sqrt{m}(N/m)^\gamma}\sum_{i=m+k^*}^{m+N}(\mathbf{x}_i-\mathbf{d})^{\mathrm{T}}\mathbf{\Delta}_m\right) \\
&= \left(\frac{N}{m}\right)^{\gamma-1/2}\left(\frac{N\delta_m}{\sqrt{m}(N/m)^\gamma} + \frac{1}{\sqrt{m}(N/m)^\gamma}\sum_{i=m+k^*}^{m+N}(\mathbf{x}_i-\mathbf{d})^{\mathrm{T}}\mathbf{\Delta}_m\right) + o(1),
\end{aligned}$$

where the last equality follows from part (iv) of Lemma 7.1. Hence, (7.8) is established if we can show that, as $m \to \infty$,

$$\begin{aligned}
&\frac{1}{\sigma}\left(\frac{N}{m}\right)^{\gamma-1/2}\left(q - \frac{N\delta_m}{\sqrt{m}(N/m)^\gamma}\right) \\
&= x + \frac{1}{\sigma}\left(\frac{N}{m}\right)^{\gamma-1/2}\frac{1}{\sqrt{m}(c_m/m)^\gamma}\sum_{i=m+k^*}^{m+c_m}(\mathbf{x}_i-\mathbf{d})^{\mathrm{T}}\mathbf{\Delta}_m + o(1).
\end{aligned} \tag{7.9}$$

To show the latter, inserting the definition of N given in (7.1) and subsequently rewriting q with the definition of c_m in Theorem 3.1, we obtain that

$$\frac{N\delta_m}{\sqrt{m}(N/m)^\gamma} = q - \frac{1}{\sqrt{m}(c_m/m)^\gamma}\sum_{i=m+k^*}^{m+c_m}(\mathbf{x}_i-\mathbf{d})^{\mathrm{T}}\mathbf{\Delta}_m,$$

so that the left-hand side of (7.9) reduces to

$$\begin{aligned}
&\frac{1}{\sigma}\left(\frac{N}{m}\right)^{\gamma-1/2}\left(\sigma x\left(\frac{m}{c_m}\right)^{\gamma-1/2} + \frac{1}{\sqrt{m}(c_m/m)^\gamma}\sum_{i=m+k^*}^{m+c_m}(\mathbf{x}_i-\mathbf{d})^{\mathrm{T}}\mathbf{\Delta}_m\right) \\
&= x + \frac{1}{\sigma\sqrt{N}(c_m/N)^\gamma}\sum_{i=m+k^*}^{m+c_m}(\mathbf{x}_i-\mathbf{d})^{\mathrm{T}}\mathbf{\Delta}_m + o(1),
\end{aligned}$$

since $N/c_m \to 1$ by (7.4), and Eq. (7.8) follows readily. Introduce

$$R(y,z) = \frac{1}{y^\gamma}\sum_{i=m+k^*}^{m+z}(\mathbf{x}_i-\mathbf{d})^{\mathrm{T}}\mathbf{\Delta}_m.$$

Using this definition, (7.8) can be rewritten as $x + \sigma^{-1}N^{\gamma-1/2}(R(c_m,c_m) - R(N,N))$. Consequently, an application of the triangle inequality implies that the lemma is proved if we can show that, as $m \to \infty$,

$$N^{\gamma-1/2}\left|R(c_m,c_m) - R(N,c_m)\right| + N^{\gamma-1/2}\left|R(N,c_m) - R(N,N)\right| = o(1). \tag{7.10}$$

We start by investigating the first sum of the right-hand side in (7.10). The mean-value theorem yields, by applying (7.4),

$$\begin{aligned}
\left|N^\gamma - c_m^\gamma\right| &= \left|(N^{1-\gamma})^{\gamma/(1-\gamma)} - (c_m^{1-\gamma})^{\gamma/(1-\gamma)}\right| \\
&= \frac{\gamma}{1-\gamma}\left|N^{1-\gamma} - c_m^{1-\gamma}\right| c_m^{2\gamma-1}(1+o(1))
\end{aligned}$$

as $m \to \infty$. Therein, for $m \to \infty$,

$$\begin{aligned}
\left|N^{1-\gamma} - c_m^{1-\gamma}\right| &= \frac{1}{\delta_m}\left|R(c_m,c_m) + \sigma x c_m^{1/2-\gamma}\right| \\
&= o\left(\frac{c_m^{1/2-\gamma}\left[m(\log m)^2\right]^{1/\mu}}{\delta_m}\right),
\end{aligned}$$

where we have used the definition of N given in (7.1) and display (7.2) of Proposition 7.1 to obtain the rate. Combining the results of the last two equations with (7.2), we arrive for the first term of (7.10) at

$$\begin{aligned}
&N^{\gamma-1/2}|R(c_m,c_m) - R(N,c_m)| \\
&= \frac{N^{\gamma-1/2}|N^\gamma - c_m^\gamma|}{c_m^\gamma N^\gamma}\left|\sum_{i=m+k^*}^{m+c_m}(\mathbf{x}_i-\mathbf{d})^{\mathrm{T}}\mathbf{\Delta}_m\right| \\
&= o\left(\frac{\left[m(\log m)^2\right]^{2/\mu}}{\sqrt{c_m}\delta_m}\right).
\end{aligned}$$

By assumption (3.4), $\delta_m \geq C_1/\log m$, so inserting the definition of c_m into the rate on the right-hand side of the previous equation implies that

$$\begin{aligned}
\frac{\left[m(\log m)^2\right]^{2/\mu}}{\sqrt{c_m}\delta_m} &= \mathcal{O}\left(m^{2/\mu-(1/2-\gamma)/(2[1-\gamma])}(\log m)^{1+4/\mu}\right) \\
&= o(1) \quad (m \to \infty)
\end{aligned}$$

because, by assumption, $\mu > 8(1-\gamma)/(1-2\gamma)$. This completes the proof for the first sum in (7.10). The proof for the second sum follows similarly and is, hence, omitted here. □

7.3. Deriving the limit distribution

In this subsection, we will complete the proof of Theorem 3.1 through a series of lemmas. We start by investigating the maximum prior to the change-point k^*.

Lemma 7.3. *If the assumptions of Theorem 3.1 are satisfied, then*

$$\left(\frac{N}{m}\right)^{\gamma-1/2}\left(\max_{1\leq k<k^*}\frac{|\Gamma_m(k)|}{h_m(k)} - \frac{1}{\sqrt{m}(N/m)^\gamma}\sum_{i=m+k^*}^{m+N}\mathbf{x}_i^{\mathrm{T}}\mathbf{\Delta}_m\right) \xrightarrow{P} -\infty,$$

where $h_m(k) = g_m(k)/q$ with $q = \sigma q^*(\alpha)$ given in (3.3).

Proof. We begin by expanding

$$\Gamma_m(k) = \sum_{i=m+1}^{m+k} \varepsilon_i + \sum_{i=m+1}^{m+k} \mathbf{x}_i^{\mathrm{T}}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_m) + \sum_{i=m+k^*}^{m+k} \mathbf{x}_i^{\mathrm{T}} \Delta_m. \qquad (7.11)$$

The equation is valid for all k (letting the empty sum equal to zero if $k < k^*$). It thus suffices to consider each term separately to prove the assertion of Lemma 7.3. As pointed out, the third sum in (7.11) vanishes for the $k < k^*$ under consideration. The first sum on the right-hand side of (7.11) has already been dealt with in Aue and Horváth (2004, Lemma 3.2). Restating their display (3.11), it holds that

$$\max_{1\le k<k^*} \frac{1}{h_m(k)} \left| \sum_{i=m+1}^{m+k} \varepsilon_i \right| = o_P(1), \quad (m \to \infty). \qquad (7.12)$$

For the second sum on the right-hand side in (7.11), note first that $\hat{\boldsymbol{\beta}}_m$ is a $\sqrt{m}$-consistent estimator of $\boldsymbol{\beta}$. Therefore, using the definition of $h_m(k)$ and Eq. (7.5) of Proposition 7.2, a subsequent rearrangement of terms shows that the second partial sum can, for $m \to \infty$, be estimated as

$$\begin{aligned}
&\max_{1\le k<k^*} \frac{1}{h_m(k)} \left| \sum_{i=m+1}^{m+k} \mathbf{x}_i^{\mathrm{T}}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_m) \right| \\
&= \mathcal{O}_P(1) \left(\max_{1\le k\le k^*} \frac{k|\mathbf{d}|}{\sqrt{m} h_m(k)} \right. \\
&\quad \left. + \max_{1\le k\le k^*} \frac{1}{\sqrt{m} h_m(k)} \left| \sum_{i=m+1}^{m+k} (\mathbf{x}_i - \mathbf{d}) \right| \right) \\
&= \mathcal{O}_P(1) \max_{1\le k<k^*} \left(\frac{k}{m}\right)^{1-\gamma} \left(1 + \frac{k}{m}\right)^{-1} \left(\frac{m}{m+k}\right)^{-\gamma} \\
&\quad + \mathcal{O}_P(1) \max_{1\le k<k^*} \left(\frac{k}{m}\right)^{1-\gamma} \left(1 + \frac{k}{m}\right)^{-1} \\
&\quad \times \left(\frac{m}{m+k}\right)^{-\gamma} \left[m(\log m)^2\right]^{1/\mu} \\
&= \mathcal{O}_P(1) \left(\frac{k^*}{\sqrt{m}}\right)^{1-\gamma} m^{1/\mu-(1-\gamma)/2} (\log m)^{2/\mu} \\
&= o_P(1),
\end{aligned}$$

since $k^*/\sqrt{m} \to 0$ by parts (i) and (iv) of Lemma 7.1 and $1/\mu - (1-\gamma)/2 < 0$ by assumption on μ. The previous calculations show that the detector $\Gamma_m(k)$ does not contribute asymptotically and that the convergence result will therefore be determined by investigating the remaining trend part for which we can write, applying (7.5),

$$\begin{aligned}
&\frac{1}{\sqrt{m}(N/m)^\gamma} \sum_{i=m+k^*}^{m+N} \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\Delta}_m \\
&= \frac{1}{\sqrt{m}(N/m)^\gamma} \sum_{i=m+k^*}^{m+N} \mathbf{d}^{\mathrm{T}} \boldsymbol{\Delta}_m + \frac{1}{\sqrt{m}(N/m)^\gamma} \sum_{i=m+k^*}^{m+N} (\mathbf{x}_i - \mathbf{d})^{\mathrm{T}} \boldsymbol{\Delta}_m \\
&= \frac{(N-k^*+1)\delta_m}{\sqrt{m}(N/m)^\gamma} + o\left(m^{\gamma-1/2}\delta_m N^{1-\gamma}\right)
\end{aligned}$$

as $m \to \infty$. By (7.4),

$$m^{\gamma-1/2}\delta_m N^{1-\gamma} = \mathcal{O}(1) m^{\gamma-1/2} \delta_m c_m^{1-\gamma} = \mathcal{O}(1),$$

using the definition of c_m in Theorem 3.1. Hence,

$$\frac{1}{\sqrt{m}(N/m)^\gamma} \sum_{i=m+k^*}^{m+N} \mathbf{x}_i^{\mathrm{T}} \Delta_m = \frac{(N-k^*+1)\delta_m}{\sqrt{m}(N/m)^\gamma} + o(1) \quad (m \to \infty).$$

The limit of the right-hand side can be computed as

$$\begin{aligned}
\lim_{m\to\infty} \frac{(N-k^*+1)\delta_m}{\sqrt{m}(N/m)^\gamma} &= \lim_{m\to\infty} \frac{N^{1-\gamma}\delta_m}{m^{1/2-\gamma}} \\
&= \lim_{m\to\infty} \frac{c_m^{1-\gamma}\delta_m}{m^{1/2-\gamma}} = q = q(\alpha) > 0,
\end{aligned}$$

where the first equality sign follows from part (i) of Lemma 7.1, the second from Eq. (7.7) and the third from the definition of c_m in Theorem 3.1. The proof is complete after recognizing that, as a consequence of Lemma 7.1 (ii) and $\gamma < 1/2$, $(N/m)^{\gamma-1/2} \to \infty$ as $m \to \infty$. □

The next step consists of approximating the maximum over the remaining range of indices k ranging from k^* to N with Brownian motions. The precise formulation follows.

Lemma 7.4. *If the assumptions of Theorem 3.1 are satisfied, it holds that*

$$\begin{aligned}
&\left(\frac{N}{m}\right)^{\gamma-1/2} \max_{k^*\le k\le N} \frac{1}{h_m(k)} \left| \Gamma_m(k) - \left(\sigma W_m(k) + \sum_{i=m+k^*}^{m+k} \mathbf{x}_i^{\mathrm{T}} \Delta_m\right) \right| \\
&= o_P(1) \quad (m \to \infty),
\end{aligned}$$

where $h_m(k) = q^{-1} g_m(k)$ and $\{W_m(t) : t \ge 0\}$ denotes a standard Brownian motion.

Proof. We use the decomposition for $\Gamma_m(k)$ established in (7.11). We first show that the partial sums of the ε_i ranging between $m+1$ and $m+k$ can be approximated with Brownian motions. In a second step we show that the time series regressors corresponding to the detector and the Brownian motion are close as well. Using assumption (2.9), we can write that, as $m \to \infty$,

$$\begin{aligned}
&\left(\frac{N}{m}\right)^{\gamma-1/2} \max_{k^*\le k\le N} \frac{1}{h_m(k)} \left| \sum_{i=m+1}^{m+k} \varepsilon_i - \sigma W_m(k) \right| \\
&= \mathcal{O}_P(1) \left(\frac{N}{m}\right)^{\gamma-1/2} \max_{k^*\le k\le N} \frac{k^{1/\nu}}{h_m(k)}.
\end{aligned}$$

The right-hand side of the latter equation is $o_P(1)$ as $m \to \infty$, which has been proved in Lemma 3.3 of Aue and Horváth (2004). It remains therefore to investigate the second sum in (7.11). To do so, note that as $m \to \infty$, by the $\sqrt{m}$-consistency of $\hat{\boldsymbol{\beta}}_m$ and (7.6),

$$\begin{aligned}
&\left(\frac{N}{m}\right)^{\gamma-1/2} \max_{k^*\le k\le N} \frac{1}{h_m(k)} \left| \sum_{i=m+1}^{m+k} \mathbf{x}_i^{\mathrm{T}}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_m) \right| \\
&= \mathcal{O}_P(1) \left(\frac{N}{m}\right)^{\gamma-1/2} \\
&\quad \times \left(\max_{k^*\le k\le N} \frac{k|\mathbf{d}|}{\sqrt{m} h_m(k)} + \max_{k^*\le k\le N} \frac{1}{\sqrt{m} h_m(k)} \left| \sum_{i=m+1}^{m+k} (\mathbf{x}_i - \mathbf{d}) \right| \right) \\
&= \mathcal{O}_P(1) \left(\frac{N}{m}\right)^{\gamma-1/2} \max_{k^*\le k\le N} \left(\frac{k}{m}\right)^{1-\gamma} \left(1 + \frac{k}{m}\right)^{-1} \left(\frac{m}{m+k}\right)^{-\gamma} \\
&\quad + \mathcal{O}_P(1) \left(\frac{N}{m}\right)^{\gamma-1/2} \max_{k^*\le k\le N} \left(\frac{N}{m}\right)^{1-\gamma} \left(1 + \frac{k}{m}\right)^{-1} \left(\frac{m}{m+k}\right)^{-\gamma} \\
&= \mathcal{O}_P(1) \left(\frac{N}{m}\right)^{1/2} \\
&= o_P(1).
\end{aligned}$$

Since the third term in the decomposition (7.11) clearly cancels with the drift term associated to the Brownian motion in the statement of Lemma 7.4, its proof is complete. □

Lemma 7.5. *If the assumptions of Theorem 3.1 are satisfied, then*

$$\left(\frac{N}{m}\right)^{\gamma-1/2}\max_{k^*\le k\le N}\left|\frac{\sigma W_m(k)}{h_m(k)}-\frac{\sigma W_m(k)}{\sqrt{m}(k/m)^\gamma}\right|=o_P(1)\quad(m\to\infty),$$

where $h_m(k)=g_m(k)/q$ and $\{W_m(t):t\ge 0\}$ denotes a standard Brownian motion.

Proof. Since it follows along the lines of the proof of Lemma 3.3 in Aue and Horváth (2004), the details are omitted.

To lighten notation, we introduce an abbreviation for the drift term induced by the time series $\{\mathbf{x}_i\}$, namely we let

$$S_m(k^*,k)=\sum_{i=m+k^*}^{m+k}\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\Delta}_m,\quad k\ge 1,$$

where we apply the usual convention and set $S_m(k^*,k)=0$ if $k<k^*$. We can prove the following lemma.

Lemma 7.6. *If the assumptions of Theorem 3.1 are satisfied, then*

$$\lim_{m\to\infty}P\left\{\frac{\alpha_m(\gamma)}{\sigma}\left(\max_{k^*\le k\le N}\frac{\sigma W_m(k)+S_m(k^*,k)}{\sqrt{m}(k/m)^\gamma}-\frac{S_m(k^*,N)}{\sqrt{m}(N/m)^\gamma}\right)\le\frac{\beta_m(\gamma)}{\sigma}\right\}=\Phi(x),$$

where Φ denotes the distribution function of a standard normal random variable and

$$\alpha_m(\gamma)=\left(\frac{N}{m}\right)^{\gamma-1/2}\quad\textit{and}\quad\beta_m(\gamma)=\alpha_m(\gamma)\left(q-\frac{S_m(k^*,N)}{\sqrt{m}(N/m)^\gamma}\right).$$

Proof. *First step:* We prove that, asymptotically, the maximum will be reached close to N. To be precise, we shall show that, for all $\delta\in(0,1)$,

$$\lim_{m\to\infty}P\left\{\max_{k^*\le k\le N}\frac{\sigma W_m(k)+S_m(k^*,k)}{\sqrt{m}(k/m)^\gamma}=\max_{(1-\delta)N\le k\le N}\frac{\sigma W_m(k)+S_m(k^*,k)}{\sqrt{m}(k/m)^\gamma}\right\}=1.\tag{7.13}$$

To do so, rewrite the first maximum in Eq. (7.13) as

$$\max_{k^*\le k\le N}\frac{1}{\sqrt{m}(k/m)^\gamma}\times\left(\sigma W_m(k)+\sum_{i=m+k^*}^{m+k}(\mathbf{x}_i-\mathbf{d})^{\mathrm{T}}\boldsymbol{\Delta}_m+(k-k^*+1)\delta_m\right).$$

Then, the last term can be bounded from above by

$$\max_{k^*\le k\le N}\frac{(k-k^*+1)\delta_m}{\sqrt{m}(k/m)^\gamma}=m^{\gamma-1/2}\delta_m\max_{k^*\le k\le N}(k^{1-\gamma}-k^*k^{-\gamma})\le m^{\gamma-1/2}\delta_m N^{1-\gamma}.$$

On the other hand it is bounded from below by

$$\max_{k^*\le k\le N}\frac{(k-k^*+1)\delta_m}{\sqrt{m}(k/m)^\gamma}\ge\frac{(N-k^*+1)\delta_m}{\sqrt{m}(N/m)^\gamma}\sim m^{\gamma-1/2}\delta_m N^{1-\gamma}$$

which hence gives the precise order of the third term. Next, we compare this order of magnitude with the order of magnitude of the Brownian motion term. It follows from Lemma 3.4 of Aue and Horváth (2004) that, as $m\to\infty$,

$$\max_{k^*\le k\le N}\frac{\sigma|W_m(k)|}{\sqrt{m}(k/m)^\gamma}=\mathcal{O}_P\left(\left(\frac{N}{m}\right)^{1/2-\gamma}\right)$$

$$=\mathcal{O}_P\left(\left(\frac{N}{m}\right)^{1/2-\gamma}\frac{\sqrt{N}\delta_m}{\sqrt{N}\delta_m}\right)=o_P\left(m^{\gamma-1/2}\delta_m N^{1-\gamma}\right)$$

where we applied Lemma 7.1(iii). It remains to observe that the second term has already been dealt with in (7.5) of Proposition 7.2. We have therefore shown that

$$\lim_{m\to\infty}P\left\{\max_{k^*\le k\le N}\frac{|\sigma W_m(k)+S_m(k^*,k)|}{\sqrt{m}(k/m)^\gamma}=\max_{(1-\delta)N\le k\le N}\frac{|\sigma W_m(k)+S_m(k^*,k)|}{\sqrt{m}(k/m)^\gamma}\right\}=1,$$

which implies (7.13), since we have assumed without loss of generality that $\delta_m>0$.

Second step: We will determine the behavior for $\delta\to 0$. It follows from the first step of the proof that, for any $\epsilon>0$, there is m large enough such that

$$\left(\frac{N}{m}\right)^{\gamma-1/2}\max_{(1-\delta)N\le k\le N}\frac{1}{\sqrt{m}(k/m)^\gamma}\left|\sum_{i=m+k^*}^{m+k}(\mathbf{x}_i-\mathbf{d})^{\mathrm{T}}\boldsymbol{\Delta}_m\right|\le\epsilon\left(\frac{N}{m}\right)^{\gamma-1/2}\frac{(N-k^*+1)\delta_m}{\sqrt{m}(N/m)^\gamma}.$$

Clearly,

$$\left(\frac{N}{m}\right)^{\gamma-1/2}\max_{(1-\delta)N\le k\le N}\frac{(k-k^*+1)\delta_m}{\sqrt{m}(k/m)^\gamma}=\left(\frac{N}{m}\right)^{\gamma-1/2}\frac{(N-k^*+1)\delta_m}{\sqrt{m}(N/m)^\gamma}$$

and the drift term takes its maximum in N. Next, we note that

$$\left(\frac{N}{m}\right)^{\gamma-1/2}\max_{(1-\delta)N\le k\le N}\frac{|W_m(k)-W_m(N)|}{\sqrt{m}(k/m)^\gamma}\le\sup_{(1-\delta)N\le t\le N}\frac{|W_m(t)-W_m(N)|}{\sqrt{N}(t/N)^\gamma}\overset{\mathcal{D}}{=}\sup_{1-\delta\le s\le 1}\frac{|W_m(s)-W_m(1)|}{s^\gamma}\xrightarrow{P}0$$

as $\delta\to 0$, where we have used the scale transformation and the almost sure continuity of Brownian motions, respectively. Therefore,

$$\lim_{m\to\infty}P\left\{\frac{a_m(\gamma)}{\sigma}\left(\max_{k^*\le k\le N}\frac{\sigma W_m(k)+S_m(k^*,k)}{\sqrt{m}(k/m)^\gamma}-\frac{S_m(k^*,N)}{\sqrt{m}(k/m)^\gamma}\right)\le\frac{b_m(\gamma)}{\sigma}\right\}$$
$$=\lim_{m\to\infty}P\left\{\frac{a_m(\gamma)}{\sigma}\left(\frac{\sigma W_m(N)+S_m(k^*,N)}{\sqrt{m}(k/m)^\gamma}-\frac{S_m(k^*,N)}{\sqrt{m}(k/m)^\gamma}\right)\le\frac{b_m(\gamma)}{\sigma}\right\}$$
$$=\lim_{m\to\infty}P\left\{\frac{W(N)}{\sqrt{N}}\le\frac{b_m(\gamma)}{\sigma}\right\}=\Phi(x),$$

where the last equality follows from the fact that $W(N)/\sqrt{N}$ is a standard normal random variable and that, by Lemma 7.2, $b_m(\gamma)/\sigma\to x$ as $m\to\infty$. The proof is now complete. □

Proof of Theorem 3.1. Combining the results of Lemmas 7.3–7.6 with Eq. (3.6), we arrive at

$$\lim_{m\to\infty}P\{\tau_m\ge N_m(x)\}$$
$$=\lim_{m\to\infty}P\left\{\alpha_m(\gamma)\left(\max_{1\le k\le N}\frac{|\Gamma_m(k)|}{g_m(k)}-\frac{1}{\sqrt{m}(N/m)^\gamma}\sum_{i=m+k^*}^{m+N}(\mathbf{x}_i-\mathbf{d})^{\mathrm{T}}\boldsymbol{\Delta}_m\right)\le\beta_m(\gamma)\right\}$$

$$= \lim_{m\to\infty} P\left\{\frac{\alpha_m(\gamma)}{\sigma}\left(\max_{k^*\le k\le N}\frac{\sigma W_m(k)+S_m(k^*,k)}{\sqrt{m}(k/m)^\gamma} - \frac{S_m(k^*,N)}{\sqrt{m}(N/m)^\gamma}\right)\le\frac{\beta_m(\gamma)}{\sigma}\right\}$$
$$= \Phi(x). \tag{7.14}$$

Using the symmetry relation of the standard normal distribution function Φ and subsequently the definition of N given in (7.1), we have that

$$\begin{aligned}\Phi(x) &= 1-\Phi(-x)\\ &= 1-\lim_{m\to\infty} P\{\tau_m \ge N(m,-x)\}\\ &= 1-\lim_{m\to\infty} P\left\{\tau_m^{1-\gamma}\ge a_m^{1-\gamma}+x\frac{(1-\gamma)b_m}{c_m^\gamma}\right\}\\ &= \lim_{m\to\infty} P\left\{\frac{c_m^\gamma}{1-\gamma}\frac{\tau_m^{1-\gamma}-a_m^{1-\gamma}}{b_m}\le x\right\}.\end{aligned}$$

The last series of equations implies that the proof of Theorem 3.1 will be complete if we can establish that the random quantities

$$\tau_m^0=\frac{\tau_m-a_m}{b_m}\quad\text{and}\quad \tau_m^1=\frac{c_m^\gamma}{1-\gamma}\frac{\tau_m^{1-\gamma}-a_m^{1-\gamma}}{b_m}$$

have the same limit distribution. To provide this relation, note that from the definition of N in (7.1), Eq. (7.14), and the proof of Proposition 7.1 we obtain that

$$\frac{\tau_m}{a_m}\xrightarrow{P}1\quad\text{and}\quad\frac{a_m}{c_m}\longrightarrow 1\quad(m\to\infty).$$

Applying the latter two convergence results and the mean-value theorem yields, as $m\to\infty$,

$$\begin{aligned}\tau_m^0 &= \frac{1}{b_m}\left[(\tau_m^{1-\gamma})^{1/(1-\gamma)}-(a_m^{1-\gamma})^{1/(1-\gamma)}\right]\\ &= \frac{a_m^\gamma}{1-\gamma}\frac{\tau_m^{1-\gamma}-a_m^{1-\gamma}}{b_m}(1+o_P(1))\\ &= \frac{c_m^\gamma}{1-\gamma}\frac{\tau_m^{1-\gamma}-a_m^{1-\gamma}}{b_m}(1+o_P(1))\\ &= \tau_m^1(1+o_P(1)).\end{aligned}$$

Therefore Slutsky's theorem implies that

$$\lim_{m\to\infty}P\left\{\tau_m^0\le x\right\}=\lim_{m\to\infty}P\left\{\tau_m^1\le x\right\}$$

and, referring to Eq. (7.14), Theorem 3.1 is established. □

References

Abel, A.B., Bernanke, B.S., 2001. Macroeconomics, 4th ed. Addison Wesley, New York.

Andreou, E., Ghysels, E., 2006. Monitoring disruptions in nancial markets. Journal of Econometrics 135, 77–124.

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

Aue, A., Berkes, I., Horváth, L., 2006a. Strong approximation for the sums of squares of augmented GARCH sequences. Bernoulli 12, 583–608.

Aue, A., Horváth, L., 2004. Delay time in sequential detection of change. Statistics & Probability Letters 67, 221–231.

Aue, A., Horváth, L., Hušková, M., Kokoszka, P., 2006b. Change-point monitoring in linear models. Econometrics Journal 9, 373–403.

Aue, A., Horváth, L., Kokoszka, P., Steinebach, J., 2008a. Detecting changes in mean: asymptotic normality of stopping times. Test 17, 515–530.

Aue, A., Horváth, L., Kühn, M., Steinebach, J., 2008b. On the reaction time of moving sum detectors. Preprint.

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

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

Basseville, M., 1988. Detecting changes in signals and systems—A survey. Automatica 24, 309–326.

Berkes, I., Gombay, E., Horváth, L., Kokoszka, P., 2004. Sequential change-point detection in GARCH(p, q) models. Econometric Theory 20, 1140–1167.

Billingsley, P., 1968. Convergence of Probability Measures. Wiley, New York.

Brodsky, B.E., Darkhovsky, B.S., 1993. Nonparametric Methods in Change-Point Problems. Kluwer, Dordrecht.

Brook, D., Evans, D.A., 1972. An approach to the probability distribution of CUSUM run lengths. Biometrika 59, 539–549.

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

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

Chib, S., 1998. Estimation and comparison of multiple change point models. Journal of Econometrics 86, 221–241.

Chu, C.-S.J., Hornik, K., Kuan, C.-M., 1995a. MOSUM tests for parameter constancy. Biometrika 82, 603–617.

Chu, C.-S.J., Hornik, K., Kuan, C.-M., 1995b. The moving-estimates test for parameter stability. Econometric Theory 11, 699–720.

Chu, C.-S.J., Stinchcombe, J., White, H., 1996. Monitoring structural change. Econometrica 64, 1045–1065.

Csörgő, M., Horváth, L., 1997. Limit Theorems in Change-Point Analysis. Chichester, Wiley.

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

Hall, P., Heyde, C.C., 1980. Martingale Limit Theory and its Application. Academic Press, New York, London.

Hamilton, J.D., 1988. Rational expectations econometric analysis of changes in regime. An investigation of the term structure of interest rates. Journal of Economic Dynamics and Control 12, 385–423.

Hansen, B.E., 2001. The new econometrics of structural change: Dating breaks in US labor productivity. Journal of Economic Perspectives 15, 17–128.

Horváth, L., Hušková, M., Kokoszka, P., Steinebach, J., 2004. Monitoring changes in linear models. Journal of Statistical Planning and Inference 126, 225–251.

Horváth, L., Kühn, M., Steinebach, J., 2008. On the performance of the fluctuation test for structural change. Sequential Analysis 27, 126–140.

Hušková, M., Prašková, Z., Steinebach, J., 2007. On the detection of changes in autoregressive time series, I. Asymptotics. Journal of Statistical Planning and Inference 137, 1243–1259.

Kuan, C.-M., Hornik, K., 2005. The generalized fluctuation test: A unifying view. Econometric Reviews 82, 603–617.

Lai, T.L., 2001. Sequential analysis: Some classical problems and new challenges. Statistica Sinica 11, 303–351.

Leisch, F., Hornik, K., Kuan, C.M., 2000. Monitoring structural changes with the generalized fluctuation test. Econometric Theory 16, 835–854.

Lettau, M., Ludvigson, S.C., 2001. Consumption, aggregate wealth and expected stock returns. Journal of Finance 56, 815–849.

Lettau, M., Ludvigson, S.C., 2004. Understanding trend and cycle in asset values: Reevaluating the wealth effect on consumption. The American Economic Review 94, 276–299.

Ling, S., 2007. Self-weighted and local quasi-maximum likelihood estimation for ARMA-GARCH/IGARCH models. Journal of Econometrics 140, 849–873.

Ling, S., Li, W.K., 1998. Limiting distributions of maximum likelihood estimators for unstable autoregressive moving-average time series with general autoregressive heteroscedastic errors. The Annals of Statistics 26, 84–125.

Ling, S., Mcaleer, M., 2002. Necessary and sufficient moment conditions for the GARCH(r, s) and asymmetric power GARCH(r, s) models. Econometric Theory 18, 729–772.

Lucas, J.M., 1982. Combined Shewhart-CUSUM quality control schemes. Journal of Quality Technology 14, 51–59.

Modigliani, F., 1971. Consumer spending and monetary policy: The linkages. Federal Reserve Bank of Boston Conference Series 5, 9–84.

Nelson, D.B., 1990. Stationarity and persistence in the GARCH(1,1) model. Econometric Theory 6, 318–334.

Page, E.S., 1954. Continuous inspection schemes. Biometrika 41, 100–114.

Perron, P., 1989. The great crash, the oil-price shock and the unit-root hypothesis. Econometrica 57, 1361–1401.

Perron, P., 1997. Further evidence on breaking trend functions in macroeconomic variables. Journal of Econometrics 80, 355–385.

Perron, P., 2006. Dealing with structural breaks. In: Patterson, K., Mills, T.C. (Eds.), Palgrave Handbook of Econometrics 1. Econometric Theory. Macmillan, London, pp. 278–352.

Pesaran, M.H., Pettenuzzo, D., Timmermann, A., 2006. Forecasting time series subject to multiple structural breaks. Review of Economic Studies 73, 1057–1084.

Pesaran, M.H., Timmermann, A., 2007. Selection of estimation window in the presence of breaks. Journal of Econometrics 137, 134–161.

Petrov, V.V., 1995. Limit Theorems of Probability Theory. Clarendon, Oxford, UK.

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, 873–880.

Quandt, R.E., 1960. Tests of the hypothesis that a linear regression system obeys two separate regimes. Journal of the American Statistical Association 55, 324–330.
Shewhart, W.A., 1931. Economic Control of Manufactured Products. Van Nostrand Reinhold, New York.
Siegmund, D., Venkatraman, E.S., 1995. Using the generalized likelihood ratio statistic for sequential detection of a change-point. The Annals of Statistic 23, 255–271.
Stock, J.H., Watson, M.W., 1996. Evidence on structural instability in macroeconomic time series relations. Journal of Business and Economic Statistics 14, 11–30.
Tsay, R.S., 2005. Analysis of Financial Time Series, 2nd edition. Wiley, New York.
Zeileis, A., Leisch, F., Kleiber, C., Hornik, K., 2005. Monitoring structural change in dynamic econometric models. Journal of Applied Econometrics 20, 99–121.