Volume 77, issue 9 May 2007 ISSN 0167-7152

ELSEVIER

STATISTICS & PROBABILITY LETTERS

www.elsevier.com/locate/stapro

Available online at www.sciencedirect.com
ScienceDirect
Statistics & Probability Letters 77 (2007) 885–895
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

On sequential detection of parameter changes in linear regression ☆

Lajos Horváth[a,*], Piotr Kokoszka[b], Josef Steinebach[c]

[a] *University of Utah, USA*
[b] *Utah State University, USA*
[c] *Universität zu Köln, Germany*

Received 9 June 2005; received in revised form 11 September 2006; accepted 19 December 2006
Available online 21 January 2007

Abstract

Horvath et al. [2004. Monitoring changes in linear models. J. Statist. Plann. Inference 126, 225–251] developed a family of monitoring procedures to detect a change in the parameters of a linear regression model. These procedures, which are akin to the schemes proposed by Chu et al. [1996. Monitoring structural change. Econometrica 64, 1045–1065], depend on a parameter $0\leqslant\gamma<\frac{1}{2}$. If γ is close to $\frac{1}{2}$, the detection delay is small, so it is desirable to consider the case $\gamma=\frac{1}{2}$, but an extension is not obvious. We show that it can be developed by establishing a Darling–Erdős type limit theorem.

MSC: 62J05; 62L10

Keywords: Darling–Erdős limit theorem; Linear regression model; Sequential monitoring

1. Introduction

Horváth et al. (2004), see also Aue et al. (2006), developed a family of monitoring procedures to detect a change in the parameters of a linear regression model. These procedures, which are akin to the schemes proposed by Chu et al. (1996), depend on a parameter $0\leqslant\gamma<\frac{1}{2}$. The simulations in Horváth et al. (2004) showed that if the change occurs shortly after the monitoring has commenced, then γ must be chosen as close to $\frac{1}{2}$ as possible in order to minimize the detection time. Moreover, in case of a change in the mean, Aue and Horváth (2004) showed that the delay time is proportional to $m^{1/2-\gamma}$, where m is the length of the available training sample. This also indicates that, ideally, $\gamma=\frac{1}{2}$ should be chosen. As explained below, $\gamma=\frac{1}{2}$ cannot, however, be chosen in the procedures of Horváth et al. (2004). The goal of this paper is to develop a procedure which can be viewed as a limit, as $\gamma\to\frac{1}{2}$, of the procedures of Horváth et al. (2004). It is not *a priori* clear how this limit should be defined. It turns out that the extension is not trivial and involves a Darling–Erdős type limit theorem.

☆Research supported by NATO Grant PST.EAP.CLG 980599 and by NSF Grants DMS-0413653 and INT-0223262.
*Corresponding author. Fax: +1 435 797 1822.
E-mail address: Horvath@math.utah.edu (L. Horváth).

0167-7152/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.spl.2006.12.014

In the remainder of this section, we formulate the monitoring problem, motivate the need for a new approach to include the case $\gamma = \frac{1}{2}$, and describe the extension. The main result is stated in Theorem 1.1 which is proved in Section 3. Finite sample performance of the new method is illustrated by a small simulation study in Section 2.

We consider the linear model

$$y_i = \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{\beta}_i + \varepsilon_i, \quad 1 \leqslant i < \infty,$$

where $\boldsymbol{x}_i$ is a $p \times 1$ random vector of the form

$$\boldsymbol{x}_i^{\mathrm{T}} = [1, x_{2i}, \ldots, x_{pi}] \tag{1.1}$$

and $\boldsymbol{\beta}_i$ is a $p \times 1$ parameter vector. It is assumed that there is no change in the historical sample of size m, i.e.

$$\boldsymbol{\beta}_i = \boldsymbol{\beta}_0, \quad 1 \leqslant i \leqslant m. \tag{1.2}$$

Under the null hypothesis, there is no change in the parameters:

$$\mathrm{H}_0: \quad \boldsymbol{\beta}_i = \boldsymbol{\beta}_0, \;\; i = m+1, m+2, \ldots \; . \tag{1.3}$$

Under the alternative hypothesis, the parameters change from $\boldsymbol{\beta}_0$ to $\boldsymbol{\beta}_*$ at an unknown time $m + k^*$:

$$\mathrm{H}_A: \quad \text{there is } k^* \geqslant 1 \text{ such that}$$
$$\boldsymbol{\beta}_i = \boldsymbol{\beta}_0, \;\; m < i \leqslant m + k^* \quad \text{and} \quad \boldsymbol{\beta}_i = \boldsymbol{\beta}_* \neq \boldsymbol{\beta}_0, \;\; i \geqslant m + k^* + 1. \tag{1.4}$$

We emphasize that the values of the parameters $\boldsymbol{\beta}_0, \boldsymbol{\beta}_*$ and the change-point k^* are unknown.

Horváth et al. (2004) studied the stopping time

$$\tau(m) = \begin{cases} \inf\{k \geqslant 1 : \Gamma(m,k) \geqslant g(m,k)\} \\ \infty \quad \text{if } \Gamma(m,k) < g(m,k) \;\text{ for all } k = 1, 2, \ldots \; . \end{cases}$$

The detector $\Gamma(m,k)$ and the boundary function $g(m,k)$ were chosen so that under the null hypothesis

$$\lim_{m \to \infty} P\{\tau(m) < \infty\} = \alpha,$$

where $0 < \alpha < 1$ is a prescribed number, similar to the significance level in the Neyman–Pearson test, and under the alternative

$$\lim_{m \to \infty} P\{\tau(m) < \infty\} = 1.$$

Following Chu et al. (1996), Horváth et al. (2004) used the detector

$$\Gamma(m,k) = \hat{Q}(m,k) = \frac{1}{\hat{\sigma}_m} \left| \sum_{m < i \leqslant m+k} \hat{\varepsilon}_i \right|,$$

where

$$\hat{\sigma}_m^2 = (m-p)^{-1} \sum_{1 \leqslant i \leqslant m} \hat{\varepsilon}_i^2, \quad \hat{\varepsilon}_i = y_i - \boldsymbol{x}_i^{\mathrm{T}} \hat{\boldsymbol{\beta}}_m,$$

and where

$$\hat{\boldsymbol{\beta}}_m = \left(\sum_{1 \leqslant i \leqslant m} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}} \right)^{-1} \sum_{1 \leqslant i \leqslant m} \boldsymbol{x}_i y_i$$

is the least squares estimator of $\boldsymbol{\beta}_0$. The boundary function was chosen as

$$g(m,k) = c m^{1/2} \left(1 + \frac{k}{m} \right) \left(\frac{k}{m+k} \right)^{\gamma}, \quad 0 \leqslant \gamma < \frac{1}{2}.$$

Horváth et al. (2004) showed that under H_0 (and suitable regularity conditions)

$$\lim_{m\to\infty} P\{\hat{Q}(m,k)\leqslant g(m,k) \text{ for all } 1\leqslant k<\infty\} = P\left\{\sup_{0\leqslant t\leqslant 1} |W(t)|/t^{\gamma}\leqslant c\right\}, \tag{1.5}$$

where $\{W(t),\ 0\leqslant t<\infty\}$ is a Wiener process (Brownian motion). We note that $\gamma<\frac{1}{2}$ is needed to have a finite (nondegenerate) limit in (1.5) according to the law of the iterated logarithm for $W(t)$ at $t=0$.

According to the definition of $\tau(m)$ given above, we continue monitoring until infinity when H_0 is not rejected. This is often a convenient mathematical assumption, but in reality, sooner or later we cease to monitor the observations. We therefore consider the following modification of $\tau(m)$:

$$\tau^*(m) = \begin{cases} \inf\{k : 1\leqslant k\leqslant N, \hat{Q}(m,k)\geqslant c(m;t)g^*(m,k)\} \\ N \quad \text{if } \hat{Q}(m,k)<c(m;t)g^*(m,k) \text{ for all } 1\leqslant k\leqslant N, \end{cases}$$

where $N=N(m)$,

$$g^*(m,k) = m^{1/2}\left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^{1/2},$$

and

$$c(m;t) = \frac{t+D(\log m)}{A(\log m)}, \tag{1.6}$$

$$A(x) = (2\log x)^{1/2}, \quad D(x) = 2\log x + \tfrac{1}{2}\log\log x - \tfrac{1}{2}\log\pi. \tag{1.7}$$

The main difference between $\tau^*(m)$ and $\tau(m)$, apart from using a finite monitoring horizon and $\gamma=\frac{1}{2}$ rather than $\gamma<\frac{1}{2}$, is that the constant c in $\tau(m)$ is now replaced by the sequence $c(m;t)$ in $\tau^*(m)$, and $c(m;t)$ increases to infinity like $(2\log\log m)^{1/2}$. We will see in the following, see condition (1.12), that in order to develop a satisfactory asymptotic theory, we need to assume that N increases with m not slower than a linear function.

We assume that the following conditions are satisfied:

$$\varepsilon_1, \varepsilon_2, \ldots \text{ are independent identically distributed random variables.} \tag{1.8}$$

$$E\varepsilon_i = 0, \quad 0<\sigma^2=E\varepsilon_i^2<\infty \quad \text{and} \quad E|\varepsilon_i|^{\nu}<\infty \text{ with some } \nu>2. \tag{1.9}$$

$$\{\varepsilon_i,\ 1\leqslant i<\infty\} \quad \text{and} \quad \{\boldsymbol{x}_i,\ 1\leqslant i<\infty\} \text{ are independent.} \tag{1.10}$$

There is a positive definite matrix $\boldsymbol{C}$ and a constant $\rho>0$ such that

$$\left|\frac{1}{m}\sum_{1\leqslant i\leqslant m} \boldsymbol{x}_i\boldsymbol{x}_i^{\mathrm{T}} - \boldsymbol{C}\right| = \mathrm{O}(m^{-\rho}) \quad \text{a.s.} \tag{1.11}$$

$$N=\mathrm{O}(m^{\lambda}) \quad \text{with some } 1\leqslant\lambda<\infty \quad \text{and} \quad \liminf_{m\to\infty} N/m>0. \tag{1.12}$$

There are random variables ξ and m_0 and a constant $\rho>0$ such that if $m\geqslant m_0$

$$\left|\sum_{m<i<m+k} (\boldsymbol{x}_i-\boldsymbol{c}_1)\right| \leqslant \xi[(m+k)^{1/2-\rho} + (k\log N)^{1/2}], \quad 1\leqslant k\leqslant N. \tag{1.13}$$

The vector $\boldsymbol{c}_1$ in (1.13) is the first column of the matrix $\boldsymbol{C}$ in (1.11). Condition (1.13) is more technical than the other assumptions, but it is satisfied by a large class of random vectors $\boldsymbol{x}_i$. For example, if there are Wiener processes $\{W_j(t),\ 0\leqslant t<\infty\}$ and constants $\kappa_j\geqslant 0$ $(j=1,2,\ldots,p)$ such that

$$\sum_{m\leqslant i\leqslant n} (x_{ji}-c_j) - \kappa_j W_j(n) = \mathrm{O}(n^{1/2-\rho}) \quad \text{a.s.} \tag{1.14}$$

holds, then (1.13) is satisfied. Indeed, by (1.14) we have

$$\left|\sum_{m<i<m+k}(x_{ji}-c_j)-\kappa_j(W_j(m+k)-W_j(m))\right|\leqslant\xi((m+k)^{1/2-\rho})$$

for all $1\leqslant k<\infty$ if $m\geqslant m_0$ and by the modulus of continuity of a Wiener process, cf. Csörgő and Révész (1981), we have

$$\max_{1\leqslant k\leqslant N}|W_j(m+k)-W_j(m)|/(k\log N)^{1/2}=\mathrm{O}(1)\quad\text{a.s.}$$

Throughout this paper we assume that we have a realization of $\{\boldsymbol{x}_i\}$ for which (1.11) and (1.13) hold. We now state the main result of the present paper:

Theorem 1.1. *If* (1.2)–(1.4) *and* (1.8)–(1.13) *hold, then for all* $-\infty<t<\infty$,

$$\lim_{m\to\infty}P\{\hat{Q}(m,k)<c(m;t)g^*(m,k)\quad\textit{for all }1\leqslant k\leqslant N\}=\exp\{-\mathrm{e}^{-t}\}.$$

2. A simulation study

To facilitate the practical application and comparison of the methods proposed in Horváth et al. (2004) and in the present paper, we start by presenting these methods in a way suitable for direct application.

A change in the parameters is signalled at time k such that the detector function

$$D_m(k)=\frac{1}{m\hat{\sigma}_m^2}\left(\frac{m}{k+m}\right)\left(\frac{m+k}{k}\right)^{\gamma}\left|\sum_{m<i\leqslant m+k}\hat{\varepsilon}_i\right|$$

first exceeds a critical value. In Horváth et al. (2004), detectors D_m with $0\leqslant\gamma<\frac{1}{2}$ were considered. Here, we consider $\gamma=\frac{1}{2}$. The main practical difference between these two cases lies in the way the critical values are determined. If $0\leqslant\gamma<\frac{1}{2}$, detection is signalled when $D_m(k)>c_\alpha(\gamma)$, with $c_\alpha(\gamma)$ determined by

$$P\left\{\sup_{0\leqslant t\leqslant 1}|W(t)|/t^{\gamma}>c_\alpha(\gamma)\right\}=\alpha. \tag{2.1}$$

These critical values were computed in Horváth et al. (2004) using simulations. If $\gamma=\frac{1}{2}$, detection is signalled when $D_m(k)>c_\alpha^*(m)$, where now $c_\alpha^*(m)$ depends on m and is determined by

$$c_\alpha^*(m)=c(m;-\log[-\log(1-\alpha)])=\frac{-\log[-\log(1-\alpha)]+D(\log m)}{A(\log m)}, \tag{2.2}$$

with the functions $A(\cdot)$ and $D(\cdot)$ defined in (1.7). In (2.2), as throughout the paper, log denotes the natural logarithm. These critical values increase slowly with m as shown in Table 1.

For moderate values of m, the critical values $c_\alpha^*(m)$ are close to the critical values $c_\alpha(\gamma)$, provided γ is close to $\frac{1}{2}$. For example, $c_{.10}(.45)=2.5437, c_{.10}(.49)=2.8259$ and $c_{.10}^*(300)=2.9139$. In this sense, the detection method proposed in this paper can be seen as a limiting case, as $\gamma\to\frac{1}{2}$, of the detection methods introduced in Horváth et al. (2004).

Since the critical values are close and the realizations of the detector functions with, say, $\gamma=.49$, are, in finite samples, close to those with $\gamma=.50$, it can be expected that the detection times for $\gamma=.49$ and .50 will be similar. This is illustrated in Table 2, which also shows that the method with $\gamma=.50$ dominates methods with $\gamma<.50$ as the training sample size m increases and/or the size of the change increases. The detection times for $\gamma<.45$ are longer than those shown in Table 2. If $m\leqslant 100$, using $\gamma=.50$ does not improve detection time. For such values of m, using $\gamma=.45$ is recommended.

Table 1
Critical values $c_\alpha^*(m)$, cf. (2.2)

m	25	50	100	300	500	600
$\alpha=.05$	3.148301	3.197410	3.240825	3.29961	3.323552	3.331652
$\alpha=.10$	2.677540	2.761607	2.828947	2.913876	2.946973	2.958018

Table 2
Five number summary for the detection time for monitoring methods with $\gamma = .45,\ .49,\ .50$

γ	Min	Q1	Med	Q3	Max
	$m = 300\ \Delta = 0.4$				
45	1	25	47	78	449
49	1	25	47	82	1227
50	1	25	48	85	1925
	$m = 300\ \Delta = 0.8$				
45	1	9	13	19	80
49	1	7	12	19	63
50	1	7	12	19	77
	$m = 300\ \Delta = 1.2$				
45	1	4	7	10	29
49	1	4	6	9	27
50	1	4	6	9	30
	$m = 600\ \Delta = 0.4$				
.45	1	27	47	76	303
.49	1	24	45	75	456
.50	1	24	47	79	591
	$m = 600\ \Delta = 0.8$				
.45	1	9	15	21	67
.49	1	7	12	18	63
.50	1	7	13	19	58
	$m = 600\ \Delta = 1.2$				
.45	1	5	7	10	29
.49	1	4	6	9	32
.50	1	4	6	9	31

Controlled size $\alpha = .10$. The mean of unit variance independent normal observations changes at $k^* = 1$ from zero to Δ.

Table 3
Empirical size of monitoring methods with $\gamma = .45, .49, .50$

		$\gamma = .45$		$\gamma = .49$		$\gamma = .50$	
	q	10%	5%	10%	5%	10%	5%
$m = 300$	$2m$	6.16	2.92	4.92	2.04	5.92	1.84
	$4m$	7.68	3.72	5.52	2.40	6.44	2.24
	$6m$	8.12	4.00	5.72	2.40	6.92	2.32
	$9m$	8.36	4.12	5.88	2.40	7.00	2.36
$m = 600$	$2m$	5.84	2.68	5.80	2.92	5.20	1.68
	$4m$	6.88	3.32	6.32	3.20	5.76	2.04
	$6m$	7.12	3.48	6.48	3.20	6.12	2.08
	$9m$	7.28	3.56	6.64	3.28	6.16	2.08

For each m, the percentage of rejections up to time q is reported based on 2500 realizations of independent standard normal random variables of length $9m$. Standard errors are about 1%.

When comparing the performance of monitoring methods of the type discussed here, it is important to keep in mind that the controlled size α has a different interpretation than in the Neyman–Pearson paradigm. The goal is to keep the probability of false rejection below α rather than to make it close to α. Indeed, if approximately 100α percent of realizations of $D_m(k)$ exceeded the critical value up to time q, this percentage could only increase if monitoring were to continue beyond time q. Thus, if controlling the rate of false alarms is the objective, a method with the smallest empirical size is preferred. Table 3 shows that the methods with

$\gamma = .49$ and $\gamma = .50$ dominate in this sense the method with $\gamma = .45$. For $m = 600$, $\gamma = .50$ gives smaller empirical size than $\gamma = .49$.

The limited simulations discussed here should not be viewed as representing a comprehensive comparison of the various methods, but merely as an illustration of their performance. A simulation study specifically designed for a problem at hand is needed to decide which method is most suitable.

3. Proof of Theorem 1.1

The proof of Theorem 1.1 is based on a series of lemmas.

Denote

$$\boldsymbol{C}_m = \frac{1}{m} \sum_{1 \leqslant i \leqslant m} \boldsymbol{x}_i \boldsymbol{x}_i^{\mathrm{T}}.$$

Lemma 3.1. *If* (1.11) *holds, then*

$$|\boldsymbol{C}_m^{-1} - \boldsymbol{C}| = \mathrm{O}(m^{-\rho}) \quad a.s. \tag{3.1}$$

Proof. Relation (3.1) follows directly from assumption (1.11). □

Lemma 3.2. *If the assumptions of Theorem* 1.1 *are satisfied, then, as* $m \to \infty$

$$\max_{1 \leqslant k \leqslant N} \frac{1}{g^*(m,k)} \left| \sum_{m<i\leqslant m+k} \hat{\varepsilon}_i - \left(\sum_{m<i\leqslant m+k} \varepsilon_i - \frac{k}{m} \sum_{1 \leqslant i \leqslant m} \varepsilon_i \right) \right| = \mathrm{O}_{\mathrm{P}}(m^{-\rho}).$$

Proof. Since

$$\hat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_0 = \boldsymbol{C}_m^{-1} \frac{1}{m} \sum_{1 \leqslant j \leqslant m} \boldsymbol{x}_j \varepsilon_j,$$

we obtain

$$\begin{aligned} \sum_{m<i\leqslant m+k} \hat{\varepsilon}_i &= \sum_{m<i\leqslant m+k} [\varepsilon_i - \boldsymbol{x}_i^{\mathrm{T}}(\hat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_0)] \\ &= \sum_{m<i\leqslant m+k} \varepsilon_i - \left(\sum_{m<i\leqslant m+k} \boldsymbol{x}_i \right)^{\mathrm{T}} \boldsymbol{C}_m^{-1} \frac{1}{m} \sum_{1 \leqslant j \leqslant m} \boldsymbol{x}_j \varepsilon_j. \end{aligned}$$

By the Central Limit Theorem and (1.11), we have

$$\left| \sum_{1 \leqslant j \leqslant m} \boldsymbol{x}_j \varepsilon_j \right| = \mathrm{O}_{\mathrm{P}}(m^{1/2}). \tag{3.2}$$

Putting together (3.1), (3.2) and (1.13), we get

$$\begin{aligned} &\max_{1 \leqslant k \leqslant N} \frac{1}{g^*(m,k)} \left| \left[\frac{1}{m} \left(\sum_{m<i\leqslant m+k} \boldsymbol{x}_i \right)^{\mathrm{T}} \boldsymbol{C}_m^{-1} - \frac{k}{m} \boldsymbol{c}_1^{\mathrm{T}} \boldsymbol{C}^{-1} \right] \sum_{1 \leqslant j \leqslant m} \boldsymbol{x}_j \varepsilon_j \right| \\ &= \mathrm{O}_{\mathrm{P}}(m^{1/2}) \max_{1 \leqslant k \leqslant N} \frac{1}{m g^*(m,k)} \left| \sum_{m<i\leqslant m+k} (\boldsymbol{x}_i - \boldsymbol{c}_1)^{\mathrm{T}} \boldsymbol{C}_m^{-1} + k \boldsymbol{c}_1^{\mathrm{T}} (\boldsymbol{C}_m^{-1} - \boldsymbol{C}^{-1}) \right| \\ &= \mathrm{O}_{\mathrm{P}}(1) \max_{1 \leqslant k \leqslant N} \frac{(m+k)^{1/2-\rho} + (k \log N)^{1/2} + k m^{-\rho}}{(m+k)^{1/2} k^{1/2}} = \mathrm{O}_{\mathrm{P}}(m^{-\rho}). \end{aligned} \tag{3.3}$$

By (1.1) we have that $\boldsymbol{c}_1^{\mathrm{T}} \boldsymbol{C}^{-1} = [1, 0, \ldots, 0]$, and therefore Lemma 3.2 follows from (3.3). □

Lemma 3.3. *If the assumptions of Theorem* 1.1 *are satisfied, then there are two independent Wiener processes* $\{W_{1,m}(t),\ 0\leqslant t<\infty\}$ *and* $\{W_{2,m}(t),\ 0\leqslant t<\infty\}$ *such that for any* $a(m)\geqslant 1$

$$\sup_{a(m)\leqslant k<\infty}\frac{1}{g^*(m,k)}\left|\sum_{m<i\leqslant m+k}\varepsilon_i-\frac{k}{m}\sum_{1\leqslant i\leqslant m}\varepsilon_i-\sigma\left(W_{1,m}(k)-\frac{k}{m}W_{2,m}(m)\right)\right|=\mathrm{O_P}(a(m)^{1/v-1/2}).$$

Proof. By (1.10) and the K–M–T approximation (Komlós et al., 1975, 1976; Major, 1976a, b) for each m we can find two independent Wiener processes

$$\{W_{1,m}(t),\ 0\leqslant t<\infty\}\quad\text{and}\quad\{W_{2,m}(t),\ 0\leqslant t<\infty\}$$

such that

$$\sup_{1\leqslant k<\infty}k^{-1/v}\left|\sum_{m<i\leqslant m+k}\varepsilon_i-\sigma W_{1,m}(k)\right|=\mathrm{O_P}(1)\quad(m\to\infty)$$

and

$$\sum_{1\leqslant i\leqslant m}\varepsilon_i-\sigma W_{2,m}(k)=\mathrm{o_P}(m^{1/v}).$$

Hence

$$\sup_{a(m)\leqslant k<\infty}\frac{1}{g^*(m,k)}\left|\sum_{m<i\leqslant m+k}\varepsilon_i-\frac{k}{m}\sum_{1\leqslant i\leqslant m}\varepsilon_i-\sigma\left(W_{1,m}(k)-\frac{k}{m}W_{2,m}(m)\right)\right|$$
$$=\mathrm{O_P}(1)\sup_{a(m)\leqslant k<\infty}\left\{k^{1/v}+\frac{k}{m}m^{1/v}\right\}\left\{m^{1/2}\left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^{1/2}\right\}^{-1}=\mathrm{O_P}(a(m)^{1/v-1/2}),$$

which completes the proof. □

Let

$$t_k=\frac{k}{m}\quad\text{and}\quad s_k=\frac{t_k}{1+t_k}=\frac{k}{k+m}.$$

If $\{W_1(t),\ 0\leqslant t<\infty\}$, $\{W_2(t),\ 0\leqslant t<\infty\}$ and $\{W(t),\ 0\leqslant t<\infty\}$ are independent Wiener processes, then

$$\{W_1(t)-tW_2(1),\ 0\leqslant t<\infty\}\stackrel{\mathrm{d}}{=}\left\{(1+t)W\left(\frac{t}{1+t}\right)\ 0\leqslant t<\infty\right\},$$

so Lemma 3.3 can be rewritten as

$$\sup_{a(m)\leqslant k<\infty}\frac{1}{g^*(m,k)}\left|\sum_{m<i\leqslant m+k}\varepsilon_i-\frac{k}{m}\sum_{1\leqslant i\leqslant m}\varepsilon_i-\sigma(1+t_k)m^{1/2}W_m(s_k)\right|=\mathrm{O_P}(a(m)^{1/v-1/2})\tag{3.4}$$

with suitably chosen Wiener processes $\{W_m(t),\ 0\leqslant t<\infty\}$.

Lemma 3.4. *If* $\{W(t),\ 0\leqslant t<\infty\}$ *denotes a Wiener process, then*

$$\frac{1}{(2\log\log m)^{1/2}}\max_{1\leqslant k\leqslant N}\frac{|W(s_k)|}{s_k^{1/2}}\xrightarrow{\mathrm{P}}1\quad(m\to\infty).$$

Proof. By (1.12), we can assume that $N>cm$ with some $c>0$. Clearly

$$\sup_{cm\leqslant k<\infty}\frac{|W(s_k)|}{s_k^{1/2}}\leqslant\sup_{c/(1+c)\leqslant s\leqslant 1}\frac{|W(s)|}{s^{1/2}}=\mathrm{O_P}(1).$$

Hence it is enough to prove that

$$\frac{1}{(2\log\log m)^{1/2}}\max_{1\leqslant k\leqslant cm}\frac{|W(s_k)|}{s_k^{1/2}}\xrightarrow{\text{P}}1.$$

We note that

$$|s_k-t_k|\leqslant t_k^2. \tag{3.5}$$

We write

$$\left|\frac{W(s_k)}{s_k^{1/2}}-\frac{W(t_k)}{t_k^{1/2}}\right|\leqslant\frac{|W(s_k)-W(t_k)|}{s_k^{1/2}}+\frac{|W(t_k)|}{t_k^{1/2}}\frac{|s_k-t_k|}{s_k^{1/2}(s_k^{1/2}+t_k^{1/2})}=:T_1(k)+T_2(k). \tag{3.6}$$

Using the modulus of continuity of W and (3.5), we get that

$$\max_{1\leqslant k\leqslant cm/\log m}T_1(k)=O_P(1)\max_{1\leqslant k\leqslant cm/\log m}\frac{t_k(\log m)^{1/2}}{s_k^{1/2}}=o_P(1). \tag{3.7}$$

The law of the iterated logarithm gives

$$\max_{1\leqslant k\leqslant cm/\log m}T_2(k)=O_P(1)(\log\log m)^{1/2}\max_{1\leqslant k\leqslant cm/\log m}\frac{t_k^2}{s_k}=o_P(1). \tag{3.8}$$

By the scale transformation of W, we have

$$\max_{1\leqslant k\leqslant cm/\log m}\frac{|W(t_k)|}{t_k^{1/2}}\stackrel{\text{d}}{=}\max_{1\leqslant k\leqslant cm/\log m}\frac{|W(k)|}{k^{1/2}},$$

so the law of the iterated logarithm for the partial sums yields

$$\frac{1}{(2\log\log m)^{1/2}}\max_{1\leqslant k\leqslant cm/\log m}\frac{|W(k)|}{k^{1/2}}\xrightarrow{\text{P}}1,$$

since

$$\frac{\log\log m}{\log\log(cm/\log m)}\to 1\quad(m\to\infty).$$

Thus we have

$$\frac{1}{(2\log\log m)^{1/2}}\max_{1\leqslant k\leqslant cm/\log m}\frac{|W(s_k)|}{s_k^{1/2}}\xrightarrow{\text{P}}1. \tag{3.9}$$

Clearly, for large m,

$$\max_{cm/\log m\leqslant k\leqslant cm}\frac{|W(s_k)|}{s_k^{1/2}}\leqslant\sup_{c/(2\log m)\leqslant s\leqslant 1}\frac{|W(s)|}{s^{1/2}}.$$

Since

$$\frac{\log\log\log m}{\log\log(2\log m/c)}\to 1\quad(m\to\infty),$$

the law of the iterated logarithm for W (cf. Lemmas 1.1 and 1.2 on pp. 255–256 of Csörgő and Horváth, 1993) implies

$$\frac{1}{(2\log\log\log m)^{1/2}}\sup_{c/(2\log m)\leqslant s\leqslant 1}\frac{|W(s)|}{s^{1/2}}\xrightarrow{\text{P}}1.$$

Hence

$$\frac{1}{(2\log\log m)^{1/2}}\max_{cm/\log m\leqslant k\leqslant m}\frac{|W(s_k)|}{s_k^{1/2}}\xrightarrow{\text{P}}0. \tag{3.10}$$

Lemma 3.4 now follows from (3.9) and (3.10). □

Lemma 3.5. *If* $a(m) = (\log m)^\delta$ *with some* $\delta > 0$, *then*

$$\frac{1}{(2\log\log\log m)^{1/2}}\max_{1\leqslant k\leqslant a(m)}\frac{|W(s_k)|}{s_k^{1/2}}\xrightarrow{\text{P}}1.$$

Proof. It follows from (3.6)–(3.8) that

$$\max_{1\leqslant k\leqslant a(m)}\left|\frac{|W(s_k)|}{s_k^{1/2}}-\frac{|W(t_k)|}{t_k^{1/2}}\right| = \text{O}_\text{P}(1).$$

By the scale transformation of W,

$$\max_{1\leqslant k\leqslant a(m)}\frac{|W(t_k)|}{t_k^{1/2}}\overset{\text{d}}{=}\max_{1\leqslant k\leqslant a(m)}\frac{|W(k)|}{k^{1/2}},$$

so the law of the iterated logarithm for partial sums gives

$$\frac{1}{(2\log\log a(m))^{1/2}}\max_{1\leqslant k\leqslant a(m)}\frac{|W(k)|}{k^{1/2}}\xrightarrow{\text{P}}1. \tag{3.11}$$

Since $(\log\log\log m)/(\log\log a(m))\to 1$, Lemma 3.5 follows from (3.11). □

Lemma 3.6. *If* $a(m) = (\log m)^\delta$ *with* $\delta > 1$, *then, for all* t,

$$\lim_{m\to\infty}P\left\{A(\log m)\max_{a(m)\leqslant k\leqslant N(m)}\frac{|W(s_k)|}{s_k^{1/2}}-D(\log m)\leqslant t\right\} = \exp(-\text{e}^{-t}),$$

where $A(x)$ *and* $D(x)$ *are defined in* (1.7).

Proof. By the modulus of continuity of W,

$$\begin{aligned}
&\max_{a(m)\leqslant k\leqslant N(m)}\sup_{0\leqslant s\leqslant 1/m}\left|\frac{|W(s_k)|}{s_k^{1/2}}-\frac{|W(s_k+s)|}{(s_k+s)^{1/2}}\right|\\
&\leqslant\max_{a(m)\leqslant k\leqslant N(m)}\sup_{0\leqslant s\leqslant 1/m}\left[\frac{|W(s_k)-W(s_k+s)|}{s_k^{1/2}}+\frac{|W(s_k+s)|}{(s_k+s)^{1/2}}\frac{|s|}{s_k^{1/2}(s_k^{1/2}+(s_k+s)^{1/2})}\right]\\
&=\text{O}_\text{P}(1)\max_{a(m)\leqslant k\leqslant N(m)}\left[\frac{(1/m)^{1/2}(\log m)^{1/2}}{(k/(m+k))^{1/2}}+(\log\log\log m)^{1/2}\frac{1/m}{k/(m+k)}\right]\\
&=\text{O}_\text{P}(1)((\log m)^{(1-\delta)/2}),
\end{aligned}$$

where we also used the law of the iterated logarithm for W at 0. Thus we have

$$A(\log m)\left|\max_{a(m)\leqslant k\leqslant N(m)}\frac{|W(s_k)|}{s_k^{1/2}}-\sup_{a(m)/m\leqslant t\leqslant N(m)/m}\frac{|W(t/(1+t))|}{(t/(1+t))^{1/2}}\right| = \text{o}_\text{P}(1).$$

Observe that

$$\sup_{a(m)/m\leqslant t\leqslant N(m)/m}\frac{|W(t/(1+t))|}{(t/(1+t))^{1/2}} = \sup_{a(m)/(m+a(m))\leqslant s\leqslant N(m)/(m+N(m))}\frac{|W(s)|}{s^{1/2}}.$$

According to (1.12), we can assume that $N(m) > cm$ with some $c > 0$. Since

$$\sup_{c/(1+c)\leqslant s\leqslant N/(m+N)} \frac{|W(s)|}{s^{1/2}} \leqslant \sup_{c/(1+c)\leqslant s\leqslant 1} \frac{|W(s)|}{s^{1/2}} = \mathrm{O_P}(1),$$

we get that

$$A(\log m) \sup_{c/(1+c)\leqslant s\leqslant 1} \frac{|W(s)|}{s^{1/2}} - D(\log m) \xrightarrow{\mathrm{P}} -\infty.$$

Hence it is enough to prove that

$$\lim_{m\to\infty} P\left\{A(\log m) \sup_{a(m)/(m+a(m))\leqslant s\leqslant c/(1+c)} \frac{|W(s)|}{s^{1/2}} \leqslant t + D(\log m)\right\} = \exp(-\mathrm{e}^{-t}). \tag{3.12}$$

Elementary calculations give

$$A(\log m)\left|A\left(\log \frac{c}{1+c}\frac{m+a(m)}{a(m)}\right) - A(\log m)\right| \to 0 \tag{3.13}$$

and

$$D\left(\log \frac{c}{1+c}\frac{m+a(m)}{a(m)}\right) - D(\log m) \to 0. \tag{3.14}$$

The Darling–Erdős (1956) law for $W(t)/\sqrt{t}$ (cf. Csörgő and Horváth, 1993, p. 256) gives

$$\lim_{m\to\infty} P\left\{A\left(\log \frac{c}{1+c}\frac{m+a(m)}{a(m)}\right) \sup_{a(m)/(m+a(m))\leqslant s\leqslant c/(1+c)} \frac{|W(s)|}{s^{1/2}} \right.$$
$$\left. \leqslant t + D\left(\log \frac{c}{1+c}\frac{m+a(m)}{a(m)}\right)\right\} = \exp(-\mathrm{e}^{-t}),$$

and therefore (3.12) follows via (3.13) and (3.14). □

Lemma 3.7. *If the conditions of Theorem* 1.1 *are satisfied, then, for all real t*,

$$\lim_{m\to\infty} P\left\{A(\log m)\,\frac{1}{\sigma} \max_{1\leqslant k\leqslant N(m)} \frac{1}{[k(m+k)]^{1/2}}\left|\sum_{m<i\leqslant m+k} \hat{\varepsilon}_i\right| \leqslant t + D(\log m)\right\} = \exp(-\mathrm{e}^{-t}).$$

Proof. Putting together Lemmas 3.2–3.5 and (3.4) with $a(m) = (\log m)^2$, we obtain

$$\frac{1}{(2\log\log\log m)^{1/2}}\frac{1}{\sigma} \max_{1\leqslant k\leqslant a(m)} \frac{1}{[k(m+k)]^{1/2}}\left|\sum_{m<i\leqslant m+k} \hat{\varepsilon}_i\right| \xrightarrow{\mathrm{P}} 1 \tag{3.15}$$

and

$$\frac{1}{(2\log\log m)^{1/2}}\frac{1}{\sigma} \max_{1\leqslant k\leqslant N} \frac{1}{[k(m+k)]^{1/2}}\left|\sum_{m<i\leqslant m+k} \hat{\varepsilon}_i\right| \xrightarrow{\mathrm{P}} 1.$$

Now, (3.15) gives

$$\lim_{m\to\infty} P\left\{A(\log m)\frac{1}{\sigma} \max_{a(m)\leqslant k\leqslant N} \frac{1}{[k(m+k)]^{1/2}}\left|\sum_{m<i\leqslant m+k} \hat{\varepsilon}_i\right| \leqslant t + D(\log m)\right\} = \exp(-\mathrm{e}^{-t}). \tag{3.16}$$

Lemma 3.3 gives

$$A(\log m)\left|\frac{1}{\sigma}\max_{a(m)\leqslant k\leqslant N}\frac{1}{[k(m+k)]^{1/2}}\sum_{m<i\leqslant m+N}\hat{\varepsilon}_i-\max_{a(m)\leqslant k\leqslant N}\frac{|W(s_k)|}{s_k^{1/2}}\right|$$
$$=\mathrm{O_P}((\log\log m)^{1/2}(\log m)^{2(1/\nu-1/2)})=\mathrm{o_P}(1),$$

and therefore (3.16) follows from Lemma 3.6. □

Proof of Theorem 1.1. In light of Lemma 3.7, it is enough to prove that $|\hat{\sigma}_m^2-\sigma^2|=\mathrm{o_P}((\log\log m)^{-1})$. This follows from the result on p. 228 of Csörgő and Horváth (1997). □

References

Aue, A., Horváth, L., 2004. Delay time in sequential detection of change. Statist. Probab. Lett. 67, 221–231.
Aue, A., Horváth, L., Hušková, M., Kokoszka, P., 2006. Change-point monitoring in linear models with conditionally heteroskedastic errors. Econometrics J. 9, 373–403.
Chu, C.-S.J., Stinchcombe, M., White, H., 1996. Monitoring structural change. Econometrica 64, 1045–1065.
Csörgő, M., Horváth, L., 1993. Weighted Approximations in Probability and Statistics. Wiley, New York.
Csörgő, M., Horváth, L., 1997. Limit Theorems in Change-Point Analysis. Wiley, New York.
Csörgő, M., Révész, P., 1981. Strong Approximations in Probability and Statistics. Academic Press, New York.
Darling, D.A., Erdős, P., 1956. A limit theorem for the maximum of normalized sum of independent random variables. Duke Math. J. 23, 143–155.
Horváth, L., Hušková, M., Kokoszka, P., Steinebach, J., 2004. Monitoring changes in linear models. J. Statist. Plann. Inference 126, 225–251.
Komlós, J., Major, P., Tusnády, G., 1975. An approximation of partial sums of independent R.V.'s and the sample DF.I. Z. Wahrscheinlichkeitstheorie und verwandte Gebiete 32, 111–131.
Komlós, J., Major, P., Tusnády, G., 1976. An approximation of partial sums of independent R.V.'s and the sample DF.II. Z. Wahrscheinlichkeitstheorie und verwandte Gebiete 34, 33–58.
Major, P., 1976a. The approximation of partial sums of independent r.v.'s. Z. Wahrscheinlichkeitstheorie und verwandte Gebiete 35, 213–220.
Major, P., 1976b. Approximation of partial sums of i.i.d.r.v.'s when the summands have only two moments. Z. Wahrscheinlichkeitstheorie und verwandte Gebiete 35, 221–230.