On the Monitoring of Changes in Linear Models with Dependent Errors

Josef G. Steinebach and Alexander Schmitz

Mathematical Institute, University of Cologne,
Weyertal 86-90,
D-50931 Köln, Germany
{jost, schmitza}@math.uni-koeln.de

Abstract. Horváth et al. (2004) developed a monitoring procedure for detecting a change in the parameters of a linear regression model having independent and identically distributed errors. We extend these results to allow for strongly mixing errors, which need not be independent of the stochastic regressors, and we also provide a class of consistent variance estimators. Applications to autoregressive time series and near-epoch dependent regressors are discussed, too.

Keywords. Change in linear model, CUSUM statistic, dependent errors, invariance principle, sequential monitoring procedure, strong mixing, time series.

1 Introduction

In testing time series data for structural stability we have to distinguish between two approaches. Retrospective procedures deal with the detection of a structural break within an observed data set of fixed size, whereas sequential procedures check the stability hypothesis each time a new observation is available. Chu et al. (1996) pointed out that the repeated application of a retrospective procedure each time new data arrive would yield a procedure that rejects a true null hypothesis of no change with probability one, as the number of applications grows. Therefore they derived an alternative (sequential) testing procedure for detecting a change in the parameters of a linear regression model, after a stable training period of size m. Their testing procedure is based on the first excess time of a detector over a boundary function, where the detector is a cumulative sum (CUSUM) type statistic of the residuals. The boundary function is suitably chosen such that the test attains a prescribed asymptotic size (say) α and asymptotic power one as m tends to infinity.

Horváth et al. (2004) extended these results and developed a CUSUM monitoring procedure for detecting a change expected shortly after the monitoring has begun. Since they modeled the errors of the linear regression model to be independent and identically distributed (i.i.d.), Aue et al. (2006) extended this CUSUM monitoring procedure further in order to obtain the right framework for monitoring changes in econometric data. To this end, they developed a testing procedure for monitoring a linear regression model with conditionally heteroskedastic errors. Moreover, in Aue et al. (2009), the delay time associated with the stopping rule is discussed in more detail, that is, the limit distribution under the alternative is derived for multiple time series regression models which allow for stationary regressors satisfying certain moment assumptions, but still being independent of the underlying observation errors.

Recently, Perron and Qu (2007) introduced a retrospective multiple change-point analysis of multivariate regression. They assumed strongly mixing errors, which are not necessarily independent of the stochastic regressors. In this note we show that their dependence conditions permit the application of the CUSUM monitoring procedure as well. In the next section we specify the monitoring procedure and discuss its application in testing parameter stability of an AR(1) process. We can even show that the linear model allows for near-epoch dependent (NED) regressors.

2 Model assumptions and main results

Throughout this article we assume that all random variables are defined on a common probability space $(\Omega, \mathcal{A}, P)$. In order to measure the underlying dependency of the error sequence $\{\varepsilon_i,\ 1 \le i < \infty\}$, we follow the strong mixing concept. The dependency of two sub-σ-algebras $\mathcal{G}$ and $\mathcal{H}$ is measured by

$$\alpha(\mathcal{G}, \mathcal{H}) = \sup\{|P(A \cap B) - P(A)\,P(B)| : A \in \mathcal{G}, B \in \mathcal{H}\}.$$

For consecutive integers k and ℓ the notation $\mathcal{F}_k^\ell$ denotes the σ-algebra generated by $\{\varepsilon_j,\ k \le j \le \ell\}$. The (so-called) mixing coefficient $\alpha(n)$ is defined as

$$\alpha(n) = \sup_{p\in\mathbb{N}} \ \alpha\left(\mathcal{F}_1^p, \mathcal{F}_{p+n}^\infty\right) \quad \text{for} \quad n = 1, 2, \dots .$$

Our aim is to show that the sequential monitoring procedure for the linear model

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta}_i + \varepsilon_i = \beta_{1,i} + x_{2,i}\beta_{2,i} + \cdots + x_{p,i}\beta_{p,i} + \varepsilon_i, \quad 1 \le i < \infty,$$

which was discussed in Horváth et al. (2004), continues to hold under the strong mixing condition

$$\lim_{n\to\infty} \alpha(n) = 0. \tag{1}$$

We denote the $p \times 1$ random regressors by $\mathbf{x}_i = (1, x_{2,i}, ..., x_{p,i})^T$ and set $\boldsymbol{\beta}_i = (\beta_{1,i}, \dots \beta_{p,i})^T$ as the $p \times 1$ parameter vectors, which are assumed to be constant over a training period of length m, the (so-called) "non-contamination assumption", i.e.

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

This period is used as a reference for comparisons with future observations, noticing that a decreasing mixing coefficient indicates that the historical period and the future observation are asymptotically independent in a certain sense.

Based on the newly incoming observations $y_{m+1}, y_{m+2}, \dots$, we are interested in sequentially testing the "no change" null hypothesis

$$H_0: \ \boldsymbol{\beta}_{m+i} = \boldsymbol{\beta}_0 \quad \forall\, i \ge 1$$

versus the "change at k^*" alternative, i.e.

$$H_A: \ \exists\, k^* \ \text{such that} \ \boldsymbol{\beta}_{m+i} = \boldsymbol{\beta}_0, \ 1 \le i < k^*, \ \text{but} \ \boldsymbol{\beta}_{m+k^*+i} = \boldsymbol{\beta}_* \ne \boldsymbol{\beta}_0 \quad \forall\, i \ge 0.$$

The parameter k^* is called the change-point, which is assumed to be unknown as well as the values of the parameters $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_*$.

The monitoring procedure is defined via a stopping rule τ_m based on the first exit time of a detector $\hat{Q}_m(\,\cdot\,)$ over a boundary function $g_m^*(\,\cdot\,)$, i.e.

$$\tau_m = \inf\{k \ge 1: \ |\hat{Q}_m(k)| > \sigma\, c\, g_m^*(k)\}, \quad \text{with} \quad \inf\emptyset := \infty,$$

where σ is a positive constant, $c = c(\alpha)$ a critical value, and $g_m^*(\,\cdot\,)$ a certain function to be specified below.

The idea is to determine the detector, the boundary function and the critical constant such that the false alarm rate is asymptotically fixed to a prescribed level α, and that the power of the testing procedure tends to one, i.e.

$$\lim_{m\to\infty} P_{H_0}(\tau_m < \infty) = \alpha \qquad \text{and} \qquad \lim_{m\to\infty} P_{H_A}(\tau_m < \infty) = 1.$$

Let

$$\hat{\boldsymbol{\beta}}_m = \Big(\sum_{i=1}^m \mathbf{x}_i\mathbf{x}_i^T\Big)^{-1} \sum_{j=1}^m \mathbf{x}_j y_j \qquad \text{and} \qquad \hat{\varepsilon}_i = y_i - \mathbf{x}_i^T\hat{\boldsymbol{\beta}}_m$$

denote the least squares estimator for $\boldsymbol{\beta}_0$, solely based on the training period, and the i-th residual, respectively. Following Chu et al. (1996) and Horváth et al. (2004), we use a CUSUM type detector and boundary function as follows:

$$\hat{Q}_m(k) = \sum_{i=m+1}^{m+k} \hat{\varepsilon}_i \qquad \text{and} \qquad g_m^*(k) = m^{1/2}\Big(1 + \frac{k}{m}\Big)\Big(\frac{k}{m+k}\Big)^\gamma,$$

where γ is a certain tuning constant (see below).

The main goal of this note is to derive a limiting distribution (under H_0) in the case of strongly mixing errors. It will also turn out that, if the mean regressor is not orthogonal to the parameter shift, the test has asymptotic power one.

We assume that the following conditions are satisfied: For each m, we can find two standard Wiener processes $\{W_{0,m}(t),\ 0 \le t < \infty\}$ and $\{W_{1,m}(t),\ 0 \le t < \infty\}$ and positive constants σ and δ such that a uniform weak invariance principle holds over the training period, i.e.

$$\sup_{1\le k\le m} k^{-1/(2+\delta)} \Big| \sum_{i=1}^{k} \varepsilon_i - \sigma W_{0,m}(k) \Big| = O_P(1) \quad (m \to \infty), \tag{3}$$

together with a uniform weak invariance principle for the monitoring sequence, i.e.

$$\sup_{1\le k<\infty} k^{-1/(2+\delta)} \Big| \sum_{i=m+1}^{m+k} \varepsilon_i - \sigma W_{1,m}(k) \Big| = O_P(1) \quad (m \to \infty). \tag{4}$$

Furthermore, we assume that there is a positive-definite $p \times p$ matrix $\mathbf{C}$ and a constant $\tau > 0$ such that

$$\Big\| \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T - \mathbf{C} \Big\| = O\left(n^{-\tau}\right) \quad \text{a.s.} \quad (n \to \infty), \tag{5}$$

where $\|\cdot\|$ denotes the maximum norm. We also assume that

$$\Big\| \sum_{j=1}^{m} \mathbf{x}_j \varepsilon_j \Big\| = O_P\left(m^{1/2}\right) \quad (m \to \infty). \tag{6}$$

With the parameters σ and τ introduced above we define the boundary function

$$g_m(k) = \sigma\, c\, g_m^*(k) = \sigma\, c\, m^{1/2} \Big(1 + \frac{k}{m}\Big) \Big(\frac{k}{m+k}\Big)^{\gamma}, \quad 0 \le \gamma < \min\{\tau,\ 1/2\}.$$

The Wiener process in (4) is constructed from the innovations of the monitoring sequence, whereas the first Wiener process in (3) relies only on the innovations of the training period. Except for the case of an independent error sequence, both processes are typically dependent. We point out that this dependency is also influenced by the specific construction among the various methods to obtain a strong invariance principle (cf., e.g., Philipp, 1986). We allow the approximating Wiener processes to depend on m, because we do not impose strict stationarity of the error sequence. Moreover, no rate in the decay of the mixing coefficient is assumed. The parameter σ^2 is the asymptotic variance of $m^{-1/2} \sum_{i=1}^{m} \varepsilon_i$. The choice of the parameter δ is closely related to a moment condition, e.g., $\sup_{1\le i<\infty} E\varepsilon_i^{2+\delta} < \infty$.

Now we state our main results. For details and proofs we refer to Schmitz and Steinebach (2008).

Theorem 1. *Assume that the conditions* (1) – (6) *hold. Then, under* H_0*, we have*

$$\lim_{m\to\infty} P\Big(\frac{1}{\sigma} \sup_{1\le k<\infty} \frac{|\hat{Q}_m(k)|}{g_m^*(k)} > c\Big) = P\Big(\sup_{0<t\le 1} \frac{|W(t)|}{t^{\gamma}} > c\Big),$$

where $\{W(t),\ 0 \le t < \infty\}$ *is a standard Wiener process.*

Remark 1. The limit distribution in Theorem 1 is a functional of the Wiener process and allows for an asymptotic choice of the critical value $c = c(\alpha)$. Selected quantiles are given in Horváth et al. (2004). According to Horváth et al. (2007), in the case of $\gamma = 1/2$, which is excluded here, an asymptotic extreme value distribution can be derived by proving a Darling-Erdős type limit theorem.

An application of Theorem 1 in practice requires the estimation of the unknown parameter σ. As a consequence of the invariance principles for dependent random variables, σ^2 is the long run variance, i.e.

$$0 < \sigma^2 = E\varepsilon_1^2 + 2\sum_{k=2}^{\infty} E\varepsilon_1\varepsilon_k < \infty.$$

Consistent estimators are, for example, available by using squares of sums of residuals from "non-overlapping blocks" as follows. Let $\{\ell_m,\ 1 \le m < \infty\}$ be a non-decreasing sequence of positive integers with $1 \le \ell_m \le m$ such that $\ell_m/m \to 0$ as $m \to \infty$, but

$$\liminf_{m\to\infty} \frac{\ell_m}{m^\theta} > 0 \quad \text{for some} \quad \max\Big\{1 - 2\tau, 1 - \tau - \frac{\delta}{2(2+\delta)}\Big\} < \theta < 1. \tag{7}$$

We propose the estimator

$$\hat{\sigma}_m^2 = \frac{1}{k}\sum_{j=1}^{k}\Big\{\frac{1}{\sqrt{\ell}}\sum_{i=(j-1)\ell+1}^{j\ell} \hat{\varepsilon}_i\Big\}^2, \tag{8}$$

with $\ell = \ell_m$ and $k = k_m = [m/\ell]$.

Remark 2. The estimator above is not the same but motivated by the class of consistent estimators introduced in Peligrad and Shao (1995) for the case of a ρ-mixing sequence. Our approach here is to prove the consistency of $\hat{\sigma}_m^2$ from (8) in the case of α-mixing errors solely via the approximating Wiener process.

Theorem 2. *Assume that the conditions* (1) – (8) *hold. Then, under H_0, we have*

$$\lim_{m\to\infty} P\Big(\frac{1}{\hat{\sigma}_m}\sup_{1\le k<\infty}\frac{|\hat{Q}_m(k)|}{g_m^*(k)} > c\Big) = P\Big(\sup_{0<t\le 1}\frac{|W(t)|}{t^\gamma} > c\Big),$$

where $\{W(t),\ 0 \le t < \infty\}$ is a standard Wiener process.

Theorem 3. *Let $\mathbf{c}_1^T(\boldsymbol{\beta}_0 - \boldsymbol{\beta}_*) \ne 0$, where $\mathbf{c}_1$ denotes the first column of $\mathbf{C}$ from* (5). *Assume that the conditions* (1) – (8) *hold. Then, under H_A, we have*

$$\lim_{m\to\infty} P\Big(\frac{1}{\hat{\sigma}_m}\sup_{1\le k<\infty}\frac{|\hat{Q}_m(k)|}{g_m^*(k)} > M\Big) = 1 \quad \textit{for all} \quad M > 0.$$

2.1 Monitoring changes in autoregressive models

Other than Horváth et al. (2004), we do not assume any longer in this section that the regressors are independent of the errors. This, for example, allows for lag-dependent variables as regressors. Consider, e.g., an application of the monitoring procedure to an AR(1) model $y_i = \beta y_{i-1} + u_i$, where $0 \le \beta < 1$. Note that the first assumption of Theorem 3 is not satisfied, since the mean regressor is zero. But for detecting a change in the parameter of the AR(1) model it suffices to monitor the linear model

$$y_i y_{i-1} = \begin{cases} \beta y_{i-1}^2 + u_i y_{i-1}, & m+1 \le i < m+k^*; \\ \beta_* y_{i-1}^2 + u_i y_{i-1}, & i = m+k^*, m+k^*+1, \ldots, \end{cases} \tag{9}$$

where $0 \le \beta < 1$ and $\beta^* \ge 1$.

In the sequel we assume that

$$\{u_i,\ -\infty < i < \infty\} \text{ is a centered i.i.d. sequence} \tag{10}$$

such that

$$\begin{gathered}\{y_i,\ -\infty < i < \infty\} \text{ is strictly stationary, strongly mixing and} \\ \text{that condition (3) holds with the } \varepsilon_i\text{'s being replaced by } y_i\text{'s.}\end{gathered} \tag{11}$$

If $\gamma(j) = Ey_1 y_{1+j}$ denotes the autocovariance function, we also assume that there is a positive constant $\tau > 0$ such that

$$\Big|\frac{1}{n}\sum_{i=1}^{n} y_i^2 - \gamma(0)\Big| = O\left(n^{-\tau}\right) \quad \text{a.s.} \quad (n \to \infty). \tag{12}$$

We note that $\left\{y_i^2 - \gamma(0),\ 0 \le i < \infty\right\}$ is also strongly mixing. Thus, it is sufficient to replace (12) by a moment condition guaranteeing that the squares satisfy an invariance principle (cf., e.g., Kuelbs and Philipp, 1980, Theorem 4). Then, via the approximating Wiener process, a Marcinkiewicz-Zygmund type law of large numbers yields the desired assertion. We set

$$\hat{\beta}_m = \Big\{\sum_{i=0}^{m-1} y_i^2\Big\}^{-1}\sum_{i=0}^{m-1} y_i y_{i+1} \qquad \text{and} \qquad \hat{R}_m(k) = \sum_{i=m+1}^{m+k}(y_i y_{i-1} - \hat{\beta}_m y_{i-1}^2).$$

Theorem 4. *Assume that the conditions* (10) – (12) *hold. Then, under* H_0, *we have*

$$\lim_{m\to\infty} P\Big(\frac{1}{\Gamma}\sup_{1\le k<\infty}\frac{|\hat{R}_m(k)|}{g_m^*(k)} > c\Big) = P\Big(\sup_{0<t\le 1}\frac{|W(t)|}{t^\gamma} > c\Big),$$

where $\Gamma^2 = E\left(u_1 y_0\right)^2$ *and* $\{W(t),\ 0 \le t < \infty\}$ *is a standard Wiener process.*

The statement of Theorem 4 remains true if we plug in a consistent estimator $\hat{\Gamma}_m$ based on the stable training period. According to (10), condition (11) holds with y_i being replaced by $u_i y_{i-1}$, i.e.

$$\sum_{i=1}^{m} u_i y_{i-1} - \Gamma W(m) = O_P\left(m^{1/(2+\delta)}\right).$$

From the autocovariances of the AR(1) process we compute that $\sigma_u^2 = Eu_1^2 = \gamma(0) - \beta\gamma(1)$ and $\Gamma^2 = \gamma^2(0) - \beta\gamma(1)\gamma(0)$. Therefore $\hat{\Gamma}_m^2 = \hat{\gamma}_m^2(0) - \hat{\beta}_m\hat{\gamma}_m(1)\hat{\gamma}_m(0)$ is a natural estimator for Γ^2, as a combination of covariance estimators and the least squares estimator $\hat{\beta}_m$, and hence is consistent.

Despite the fact that a change in the parameter also causes a change in the distribution of the errors $u_i y_{i-1}$ of the linear model (9), we can establish asymptotic power one.

Theorem 5. *Assume that the conditions* (10) – (12) *hold. If, in addition,* $P(y_1 = 0) = 0$, *then, under* H_A, *we have*

$$\lim_{m\to\infty} P\Big(\frac{1}{\hat{\Gamma}_m}\sup_{1\le k<\infty}\frac{|\hat{R}_m(k)|}{g_m^*(k)} > M\Big) = 1 \quad \textit{for all} \quad M > 0.$$

Remark 3. We point out that the monitoring of (9) overcomes the restriction in Theorem 3 concerning detectable parameter shifts. For another approach we refer to Hušková and Koubková (2005) and Koubková (2006) who introduced CUSUM type test statistics based on weighted residuals which are able to detect any change in the slope parameter of a linear model with i.i.d. errors. Moreover, they extended these results to AR(p) time series (cf. Hušková and Koubková, 2006).

2.2 Monitoring changes in linear models with NED regressors

In this section it is shown that our approach is also applicable for near-epoch dependent (NED) regressors. The NED concept covers widely used nonlinear time series like, e.g., the GARCH models (see Davidson, 2002). In the particular case of an NED sequence on an independent process, Ling (2007) established a strong law of large numbers (SLLN) and a strong invariance principle. For convenience we state the definition adapted from Ling (2007).

Let $\{\varepsilon_t,\ -\infty < t < \infty\}$ be a sequence of independent random variables and assume that x_t is $\mathcal{F}_{-\infty}^t$-measurable, where $\mathcal{F}_{-\infty}^t = \sigma\left(\ldots, \varepsilon_{t-1}, \varepsilon_t\right)$.

Definition 1. The process $\{x_t,\ 1 \le t < \infty\}$ is called $L_p(\nu)$-near-epoch dependent ($L_p(\nu)$-NED) on $\{\varepsilon_t,\ -\infty < t < \infty\}$ if

$$\sup_{1 \le t < \infty} \left\| x_t \right\|_p < \infty \quad \text{and} \quad \sup_{1 \le t < \infty} \left\| x_t - E(x_t | \mathcal{F}_{t-k}^t) \right\|_2 = O\left(k^{-\nu}\right) \quad (k \to \infty),$$

where $p \ge 1$ and $\nu > 0$.

If we assume that $\{\varepsilon_t,\ -\infty < t < \infty\}$ is i.i.d. and centered with $\|\varepsilon_1\|_4 < \infty$ and $\{x_t,\ 1 \le t < \infty\}$ is $L_4(\nu)$-NED with $\nu > 1/2$ and constant variance (say) $\sigma_x^2 > 0$, then the regression model

$$y_t = \beta_1 + \beta_2 x_{t-1} + \varepsilon_t$$

allows for an application of the monitoring procedure, provided conditions (3) – (6) hold.

Here, we can establish (3) and (4) via the well-known "Hungarian construction", (cf., e.g., Csörgő and Révész, 1981, Theorem 2.6.3).

Moreover, it can be proved that $\{x_{t-1}\varepsilon_t,\ 1 \le t < \infty\}$ is an $L_2(\nu)$-NED sequence on $\{\varepsilon_t,\ -\infty < t < \infty\}$, which also satisfies an appropriate mixingale property, so that an application of McLeish's (1975) maximal inequality for mixingales yields condition (6).

Finally, since the regressors are assumed to be $L_4(\nu)$-NED with $\nu > 1/2$ and constant variance $\sigma_x^2 > 0$, some further estimations show that $\left\{x_t^2 - \sigma_x^2,\ 1 \le t < \infty\right\}$ is a centered $L_2(\nu)$-NED sequence. Hence, the required rate in (5) follows from an application of the SLLN for NED sequences obtained by Ling (2007).

For details of the arguments above confer Schmitz and Steinebach (2008).

References

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

Aue, A., Horváth, L., and Reimherr, M. L. (2009). Delay times of sequential procedures for multiple time series regression models. *Journal of Econometrics*, to appear.

Chu, C.-S. J., Stinchcombe, M., and White, H. (1996). Monitoring structural change. *Econometrica*, **64**, 1045–1065.

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

Davidson, J. (2002). Establishing conditions for the functional central limit theorem in nonlinear time series processes. *Journal of Econometrics*, **106**, 243–269.

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

Horváth, L., Kokoszka, P., and Steinebach, J. (2007). On sequential detection of parameter changes in linear regression. *Statistics and Probability Letters*, **77**, 885–895.

Hušková, M. and Koubková, A. (2005). Monitoring jump changes in linear models. *Journal of Statistical Research*, **39**, 51–70.

Hušková, M. and Koubková, A. (2006). Sequential procedures for detection of changes in autoregressive sequences. In: *Proceedings of Prague Stochastics 2006* (M. Hušková and M. Janžura, Eds.), MATFYZPRESS, Prague, 437–447.

Koubková, A. (2006). Sequential change-point analysis. PhD thesis, Charles University Prague.

Kuelbs, J. and Philipp, W. (1980). Almost sure invariance principles for partial sums of mixing B-valued random variables. *Annals of Probability*, **8**, 1003–1036.

Ling, S. (2007). Testing for change points in time series models and limiting theorems for NED sequences. *Annals of Statistics*, **35**, 1213–1237.

McLeish, D.L. (1975). A maximal inequality and dependent strong laws. *Annals of Probability*, **5**, 829–839.

Peligrad, M. and Shao, Q.-M. (1995). Estimation of the variance of partial sums for ρ-mixing random variables. *Journal of Multivariate Analysis*, **52**, 140–157.

Perron, P. and Qu, Z. (2007). Estimating and testing multiple structural changes in multivariate regressions. *Econometrica*, **75**, 459–502.

Philipp, W. (1986). Invariance principles for independent and weakly dependent random variables. In: *Dependence in Probability and Statistics.* (E. Eberlein and M. S. Taqqu, Eds). *Progress in Probability and Statistics*, **11**, 225–268, Birkhäuser, Boston.

Schmitz, A. and Steinebach, J. G. (2008). A note on the monitoring of changes in linear models with dependent errors. Preprint, University of Cologne, 14 p.