Delay times of sequential procedures for multiple time series regression models

Alexander Aue[1], Lajos Horváth[2], and Matthew Reimherr[3]

[1] Department of Statistics
University of California
One Shields Avenue
Davis, CA 95616, USA
alexaue@wald.ucdavis.edu

[2] Department of Mathematics
University of Utah
155 South 1440 East
Salt Lake City, UT 84112, USA
horvath@math.utah.edu

[3] Department of Statistics
University of Chicago
5734 S. University Avenue
Chicago, IL 60637, USA
mreimherr@galton.uchicago.edu

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.

Keywords. Change-Point detection; CUSUM statistic; Linear models; Sequential tests; Structural stability; Threshold function; Time series regressors.

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), and Stock and Watson (1996), among others. The interested reader is also referred to the monographs Brodsky and Darkhovsky (1993), and Csörgő and Horváth (1997) 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). We follow and extend here the approach taken in Chu et al. (1996). This setting will be described in Section 2, while the main result will be motivated, stated and discussed in Section 3.

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

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

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

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

Condition (1) is referred to as a training period in the literature. Because the model parameters remain stable until observation m, (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

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

and the break at an unknown time alternative hypothesis

$$\begin{aligned} H_A: \quad \text{There is a } \; k^* \geq 1 \; \text{ such that } \quad & \boldsymbol{\beta}_i = \boldsymbol{\beta}_0, \qquad i = m+1, \ldots, m+k^*-1, \\ \text{but } \quad & \boldsymbol{\beta}_i = \boldsymbol{\beta}_A, \qquad i = m+k^*, m+k^*+1, \ldots, \end{aligned} \tag{3}$$

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

$$\{\mathbf{x}_i\} \text{ be a stationary sequence;} \tag{4}$$

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

$$E\Big|\sum_{i=1}^{k}(x_{ij} - d_j)\Big|^{\mu} \leq Kk^{\mu/2}, \qquad j = 1, \ldots, p. \tag{5}$$

Condition (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 (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 (5) replaces the standard but more restrictive assumption which requires that $k^{-1}\sum_{i=1}^{k}\mathbf{x}_i\mathbf{x}_i^T$ 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). Conditions (4) and (5) are satisfied for large classes of time series relevant in theory and applications, see Aue et al. (2009) for details.

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

$$\left|\sum_{i=1}^{m}\varepsilon_i\right| = \mathcal{O}_P\left(\sqrt{m}\right) \qquad (m \to \infty), \tag{6}$$

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

$$\sup_{t \geq 1/m} \frac{1}{(mt)^{1/\nu}}\left|\sum_{i=m+1}^{m+mt}\varepsilon_i - \sigma W_m(mt)\right| = \mathcal{O}_P(1) \qquad (m \to \infty). \tag{7}$$

Assumption (6) 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 (7) 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 (6) and (7) 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

$$\{\mathbf{x}_i\} \text{ and } \{\varepsilon_i\} \text{ are independent.} \tag{8}$$

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., 2006, among others).

For $i \geq 1$, define the model residuals $\hat{\varepsilon}_i = y_i - \mathbf{x}_i^T\hat{\boldsymbol{\beta}}_m$, 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

$$\tau_m = \inf\{k \geq 1 \colon |\mathit{\Gamma}_m(k)| \geq g_m(k)\} \tag{9}$$

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

$$\mathit{\Gamma}_m(k) = \sum_{i=m+1}^{m+k}\hat{\varepsilon}_i \qquad \text{and} \qquad g_m(k) = q\sqrt{m}\left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^{\gamma} \tag{10}$$

with $q = q(\alpha) > 0$ and $0 \leq \gamma < 1/2$. The quantity $\mathit{\Gamma}_m(k)$ 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 (7) consequently implies that the detector $\mathit{\Gamma}_m(k)$ 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. One can in fact show that τ_m is asymptotically equivalent to the stopping time $\tau = \inf\{t \in [0,1] \colon W(t) \geq q^* t^{\gamma}\}$, where $\{W(t) \colon t \in [0,1]\}$ 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

$$P\left\{\sup_{0\leq t\leq 1}\frac{|W(t)|}{t^{\gamma}} > q^*\right\} = \alpha \qquad \text{under } H_0. \tag{11}$$

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 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). Alternative forms of boundary functions $g_m(k)$ may be entertained as well. The interested reader is in this regard referred to Andreou 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. (2006).

The main aim of this exposition 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 $\delta_m = \mathbf{d}^T \boldsymbol{\Delta}_m$. It is assumed that there are positive constants C_1, C_2, and $C_3 \leq C_4$ such that, as $m \to \infty$,

$$\frac{C_1}{\log m} \leq |\boldsymbol{\Delta}_m| \leq C_2, \qquad \frac{C_1}{\log m} \leq |\delta_m| \leq C_2 \qquad \text{and} \qquad C_3 \leq \frac{|\delta_m|}{|\boldsymbol{\Delta}_m|} \leq C_4, \tag{12}$$

where $|\cdot|$ denotes the maximum norm of vectors. In (12), 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 (12) is satisfied. The lower bounds for $|\boldsymbol{\Delta}_m|$ and $|\delta_m|$ in (12) include the term $\log m$ in the denominator to control for the fluctuations of the time series regressors. 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 $m \to \infty$,

$$k^* = \mathcal{O}\left(m^\theta\right) \qquad \text{for some} \quad 0 \leq \theta < \frac{1-2\gamma}{4(1-\gamma)}. \tag{13}$$

Condition (13) is not motivated by the fact that CUSUM-type detectors perform better than other available procedures in the case of early changes but is rather a technical necessity imposed to ensure that the limit result of Theorem 1 holds in fact true. Given that (13) 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 1. *If conditions (1), (3)–(5), (6)–(8), (12) and (13) hold, and if $\mu > 8(1-\gamma)/(1-2\gamma)$ then as $m \to \infty$,*

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

for all real z, where

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

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

and $c_m = (qm^{1/2-\gamma}/|\delta_m|)^{1/(1-\gamma)}$ with $q = \sigma q^*(\alpha)$ determined by (11).

Theorem 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 (13) 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

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

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 $\delta_m = \mathbf{d}^T \boldsymbol{\Delta}_m$ 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 one can verify that τ_m satisfies the weak law of large numbers with normalizing sequence c_m, that is,

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

The standard approach in sequential analysis is to determine the average run length (ARL) of the monitoring procedures under consideration only. Theorem 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 which is a special case of Theorem 1.

The idea behind the proof of Theorem 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 $k \leq N$. Consequently, the probabilities of these two events must be the same, that is

$$P\{\tau_m > N\} = P\left\{ \max_{1 \leq k \leq N} \frac{|\Gamma_m(k)|}{g_m(k)} \leq 1 \right\}. \tag{14}$$

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 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.

Equation (14) 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 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 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 Timmerman (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 so-called meta distributions, the derivation of a delay time as in our main result becomes obsolete and makes the two approaches incompatible.

References

Andreou, E. and 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., and Horváth, L. (2004). Delay time in sequential detection of change. *Statistics & Probability Letters*, **67**, 221–231.

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*, forthcoming.

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

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

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

Csörgő, M., and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis.* John Wiley & Sons, Chichester.

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 U.S. labor productivity. *Journal of Economic Perspectives*, **15**, 17–128.

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

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

Leisch, F., Hornik, K., and Kuan, C.M. (2000). Monitoring structural changes with the generalized fluctuation test. *Econometric Theory*, **16**, 835-854.

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.

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

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

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.

Stock, J.H., and Watson, M.W. (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business and Economic Statistics*, **14**, 11–30.

Zeileis, A., Leisch, F., Kleiber, C., and Hornik, K. (2005). Monitoring structural change in dynamic econometric models. *Journal of Applied Econometrics*, **20**, 99-121.