Monitoring Structural Change in Dynamic Econometric Models

Achim Zeileis* Friedrich Leisch* Christian Kleiber+ Kurt Hornik†

* *Institut für Statistik & Wahrscheinlichkeitstheorie, Technische Universität Wien, Austria*

\+ *Institut für Wirtschafts- und Sozialstatistik, Universität Dortmund, Germany*

†*Institut für Statistik, Wirtschaftsuniversität Wien, Austria*

Abstract

The classical approach to testing for structural change employs retrospective tests using a historical data set of a given length. Here we consider a wide array of fluctuation-type tests in a monitoring situation – given a history period for which a regression relationship is known to be stable, we test whether incoming data are consistent with the previously established relationship. Procedures based on estimates of the regression coefficients are extended in three directions: we introduce (a) procedures based on OLS residuals, (b) rescaled statistics and (c) alternative asymptotic boundaries. Compared to the existing tests our extensions offer better power against certain alternatives, improved size in finite samples for dynamic models and ease of computation respectively. We apply our methods to two data sets, German M1 money demand and U.S. labor productivity.

Keywords: Online monitoring, CUSUM, MOSUM, moving estimates, recursive estimates.

JEL classification: C22, C52.

1 Introduction

Structural stability is of prime importance in applied time series econometrics. Estimates derived from unstable relationships erroneously considered as stable are not meaningful, inferences can be severely biased, and forecasts lose accuracy. In a comprehensive study using a sample of 76 representative US monthly time series and several thousand forecasting relations derived from them, Stock and Watson (1996) found evidence for parameter instability in a substantial fraction of their models. Not surprisingly, a recent special issue commemorating the twentieth anniversary of the *Journal of Business & Economic Statistics* in 2002, which reprints ten of the most frequently cited papers published in the Journal, includes two influential articles on structural change, Hansen (1992b) and Zivot and Andrews (1992).

The by now classical approach to the detection of structural changes attempts to detect breaks ex post, see Hansen (2001) for a state of the art survey. Starting with the pioneering work of Chu, Stinchcombe, and White (1996) a second line of research has emerged: given that in the real world new data arrive steadily it is frequently more natural to check whether incoming data are consistent with a previously established relationship, i.e., to employ a monitoring approach. Below we implement a variety of procedures for the monitoring situation.

Tests for structural change can be divided into two classes: F tests that are designed for a single-shift (of unknown timing) alternative (Hansen 1992b; Andrews 1993; Andrews and Ploberger 1994), and fluctuation tests that do not assume a particular pattern of structural change. Fluctuation tests can in turn either be based on estimates of the regression coefficients or on regression residuals (recursive or OLS), both from a widening data window or from a moving window of fixed size. The probably best-known test from the fluctuation test framework is the recursive (or standard) CUSUM test introduced by Brown, Durbin, and Evans (1975), later extended by Krämer, Ploberger, and Alt (1988) to dynamic models. A unifying view on fluctuation-type tests in historical samples is provided by Kuan and Hornik (1995).

These tests are commonly used to detect structural change *ex post* (historical tests). The class of fluctuation tests can be extended to the monitoring of structural changes, i.e. for detecting shifts *online*. Chu et al. (1996) introduced the first fluctuation test for monitoring by extending the recursive estimates test of Ploberger, Krämer, and Kontrus (1989). Leisch, Hornik, and Kuan (2000) generalized these results and established a class of estimates-based fluctuation tests for monitoring. We briefly summarize their results at the beginning of Section 3 and then extend the class of fluctuation tests for monitoring in three directions: we consider processes based on OLS residuals, rescale estimates-based processes in order to improve the empirical size, and consider alternative boundaries for the Brownian bridge in order to improve power against certain alternatives. In Section 4 we apply the methods introduced as well as some historical tests to two data sets: German M1 money demand (Lütkepohl, Teräsvirta, and Wolters 1999) – where one might suspect a structural shift following the German monetary union in 1990 – and U.S. labor productivity (Hansen 2001). The conclusions will be summarized in Section 5.

2 The model

Consider the standard linear regression model

$$y_i = x_i^\top \beta_i + u_i \qquad (i = 1, \dots, n, n+1, \dots), \tag{1}$$

where at time i, y_i is the observation of the dependent variable, $x_i = (1, x_{i2}, \dots, x_{ik})^\top$ is a $k \times 1$ vector of regressors, with the first component usually equal to unity, and β_i is the $k \times 1$ vector of regression coefficients.

We refer to the data from $i = 1, \dots, n$ as the history period, where the regression coefficients are assumed to be constant, i.e. $\beta_i \equiv \beta_0$, $i = 1, \dots, n$, and we want to monitor new data from time $n+1$ onwards to test whether any structural change occurs in this monitoring period. Thus, tests for monitoring are concerned with the hypothesis that

$$\beta_i \quad = \quad \beta_0 \qquad (i > n) \tag{2}$$

against the alternative that at some point in the future the coefficient vector β_i changes.

The results in this paper are based on two assumptions as in, e.g., Krämer et al. (1988), one about the disturbances and one about the regressors:

(A1) $\{u_i\}$ is a homoskedastic martingale difference sequence with respect to $\mathcal{A}_i$, the σ-field generated by $\{y_s, x_s, u_s | s < i\}$, with $\mathrm{E}[u_i^2|\mathcal{A}_i] = \sigma^2$.

(A2) $\{x_i\}$ is such that $\limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^n ||x_i||^{2+\delta} < \infty$ for some $\delta > 0$ and $||\cdot||$ the Euclidean norm; and furthermore that

$$\frac{1}{n}\sum_{i=1}^n x_i x_i^\top \xrightarrow{\mathrm{p}} Q$$

for some finite regular nonstochastic matrix Q.

Assumption (A2) allows for dynamic models, provided the regressors are (almost) stationary.

In what follows, $\hat\beta^{(i,j)}$ is the ordinary least squares (OLS) estimate of the regression coefficients based on the observations $i+1, \dots, i+j$, similarly matrices $Q_{(i,j)}$ indexed with (i,j) are composed using the observations from the same data window. Analogously, $\hat\beta^{(i)} \equiv \hat\beta^{(1,i)}$ denotes the OLS estimate based on all observations from 1 through i, and $Q_{(i)}$ is shorthand for $Q_{(1,i)}$. The OLS residuals are denoted as $\hat u_i = y_i - x_i^\top \hat\beta^{(n)}$ and $\hat\sigma^2$ is some suitable estimator of the disturbance variance, e.g., $\hat\sigma^2 = \frac{1}{n-k}\sum_{i=1}^n \hat u_i^2$.

The general idea behind all of the procedures we consider is to derive a process that captures the fluctuation either in estimates or in residuals of a regression model and to reject the null hypothesis of stability whenever there is excessive fluctuation in these processes, as assessed against asymptotic boundaries that the limiting processes are known to cross with a given probability. The following section presents a class of such fluctuation-type tests and extends them in several directions.

3 The generalized fluctuation test for monitoring

3.1 Estimates-based processes

Chu et al. (1996) were the first to extend a fluctuation test, namely the recursive estimates (RE) test, to the monitoring case. They suggested to employ the recursive estimates process

$$Y_n(t) = \frac{i}{\hat{\sigma}\sqrt{n}} \cdot Q_{(n)}{}^{\frac{1}{2}} \left(\hat{\beta}^{(i)} - \hat{\beta}^{(n)} \right), \tag{3}$$

where $Q_{(n)} = \cdot X_{(n)}{}^{\top} X_{(n)}/n$ and $i = \lfloor k + t(n-k) \rfloor$ with $t \geq 0$, and to reject the null hypothesis whenever (one component of) the process $Y_n(t)$ crosses the boundary $\pm b_1(t)$ where

$$b_1(t) = \sqrt{t(t-1)\left[\lambda^2 + \log\left(\frac{t}{t-1}\right)\right]} \tag{4}$$

in the monitoring period $1 < t < T$ and λ determines the significance level of this procedure, or equivalently when $\max_i |Y_{in}(t)|$, $i = 1, \ldots, k$, crosses $b_1(t)$.

Leisch et al. (2000) introduced the generalized fluctuation test for monitoring, which contains the test of Chu et al. (1996) as a special case. Specifically, they considered processes that reflect the fluctuation within estimates of the regression coefficients to detect structural changes. Another special case of this class of tests is the ME test which uses moving rather than recursive estimates, i.e.

$$Z_n(t|\,h) = \frac{\lfloor nh \rfloor}{\hat{\sigma}\sqrt{n}} \cdot Q_{(n)}{}^{\frac{1}{2}} \left(\hat{\beta}^{(\lfloor nt \rfloor - \lfloor nh \rfloor, \lfloor nh \rfloor)} - \hat{\beta}^{(n)} \right), \qquad (t \geq h) \tag{5}$$

and rejects the null hypothesis if (one component of) the process crosses the boundary $\pm c(t)$, where

$$c(t) = \lambda \cdot \sqrt{\log_+ t}, \tag{6}$$

in the monitoring period $1 < t < T$, where $\log_+ t$ is 1 for $t \leq e$ and $\log t$ otherwise. In theory the end of the monitoring period T may be infinity but in many applications using a finite T is more natural because the monitoring period, or at least a reasonable upper bound for it, is known in advance. In that way no size is lost for an infinite monitoring period on $[T, \infty)$. In addition, Leisch et al. (2000) consider tests based on the same processes but capturing the fluctuation with the range instead of the maximum of the deviation of the estimates.

3.2 Residual-based processes

As for tests for structural change in the history period, fluctuation tests for monitoring can not only be based on the differences of estimates of the regression coefficients but also on residuals; this was already considered by Chu et al. (1996) although they focused on the recursive estimates approach. Whereas they used a CUSUM procedure based on recursive residuals we will introduce monitoring processes based on the computationally much more convenient OLS residuals. The OLS residual- and estimates-based types of tests are equivalent in the case where there is only a constant regressor, a common situation in statistical quality control. The idea is as intuitive as for the estimates-based processes: the regression coefficients are just estimated once for the history period and based on these estimates the residuals of the observations in the monitoring period are computed. If there is a structural change in the monitoring period the residuals should deviate systematically from their zero mean. Thus, we introduce monitoring processes based on the OLS residuals

$$\hat{u}_i^{(n)} = y_i - x_i^{\top} \hat{\beta}^{(n)}. \tag{7}$$

The OLS-based CUSUM process for monitoring is then defined as:

$$B_n^0(t) = \frac{1}{\hat{\sigma}\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} \hat{u}_i^{(n)} \qquad (t \geq 0). \tag{8}$$

The following functional central limit (FCLT) holds for $B_n^0(t)$:

$$B_n^0(t) \quad \Rightarrow \quad W^0(t) \;=\; W(t) - t \cdot W(1), \tag{9}$$

where W and W^0 are the (1-dimensional) Brownian motion and Brownian bridge, respectively. The proof for (9) is essentially the same as in Ploberger and Krämer (1992) for the ordinary OLS-based CUSUM test except that t is from the compact interval $[0, T]$, with $T > 1$, rather than from $[0, 1]$: Rewrite (8) as

$$\hat{\sigma} B_n^0(t) \quad = \quad \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} u_i - \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} x_i^\top \left(\hat{\beta}^{(n)} - \beta \right). \tag{10}$$

As in Ploberger and Krämer (1992) the following relation holds uniformly in t on $[0, T]$:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} x_i^\top \left(\hat{\beta}^{(n)} - \beta \right) \quad = \quad \frac{t}{\sqrt{n}} \sum_{i=1}^{n} u_i + o_{\mathrm{p}}(1). \tag{11}$$

Hence (9) follows from the well-known fact that

$$\frac{1}{\sqrt{n}} \left(\sum_{i=1}^{\lfloor nt \rfloor} u_i - t \sum_{i=1}^{n} u_i \right) \quad \Rightarrow \quad \sigma \left(W(t) - t W(1) \right) \;=\; \sigma W^0(t). \tag{12}$$

The OLS-based MOSUM process for monitoring is defined analogously as:

$$M_n^0(t|h) \quad = \quad \frac{1}{\hat{\sigma}\sqrt{n}} \left(\sum_{i=\lfloor \eta t \rfloor - \lfloor nh \rfloor + 1}^{\lfloor \eta t \rfloor} \hat{u}_i \right) \qquad (t \geq h) \tag{13}$$

$$\quad = \quad B_n^0 \left(\frac{\lfloor \eta t \rfloor}{n} \right) - B_n^0 \left(\frac{\lfloor \eta t \rfloor - \lfloor nh \rfloor}{n} \right), \tag{14}$$

where $\eta = (n - \lfloor nh \rfloor)/(1 - h)$. From (14) together with (9) it follows directly that $M_n^0(t|h)$ satisfies the following FCLT:

$$M_n^0(t|h) \quad \Rightarrow \quad W^0(t) - W^0(t - h), \tag{15}$$

i.e., the OLS-based MOSUM process converges towards the process of the increments of the Brownian bridge. Therefore the limiting process for the OLS-based CUSUM and MOSUM process is the 1-dimensional special case of the k-dimensional recursive and moving estimates process. The respective empirical processes are in fact equivalent if $x_t = 1$ for all t. Thus, the boundaries given in the previous section can be used as well for the OLS-based processes.

The advantage of estimates-based processes is that there is a process for each regression coefficient, hence it can be determined which coefficient(s) is (are) responsible for the rejection of the null hypothesis. The OLS-based processes on the other hand are much easier to compute because a linear model has to be fit only once for the whole process (and not in every single step) and then just residuals have to be computed. In many applications, like macroeconomics, computation time is not an issue because the time series considered are often relatively short. However, the situation may be different for real-time monitoring of high-frequency data in financial applications.

3.3 Rescaling of estimates-based processes

The estimates-based processes from (3) and (5) scale the estimates of the regression coefficients with the estimate $Q_{(n)}$ of their asymptotic covariance matrix Q that is based on the observations in the history period. Kuan and Chen (1994) showed by simulation of empirical sizes that the tests can be seriously distorted in dynamic models and suggested to rescale the processes to repair this defect. Instead of estimating Q always on the basis of the full history period, each estimate $\hat{\beta}^{(i,j)}$ is scaled with the corresponding estimate of the covariance matrix $Q_{(i,j)}$, i.e., the modified processes can be written as

$$Y_n^*(t) = \frac{i}{\hat{\sigma}\sqrt{n}} \cdot Q_{(i)}{}^{\frac{1}{2}} \left(\hat{\beta}^{(i)} - \hat{\beta}^{(n)}\right), \tag{16}$$

$$Z_n^*(t|\,h) = \frac{\lfloor nh \rfloor}{\hat{\sigma}\sqrt{n}} \cdot Q_{(\lfloor nt \rfloor - \lfloor nh \rfloor, \lfloor nh \rfloor)}{}^{\frac{1}{2}} \left(\hat{\beta}^{(\lfloor nt \rfloor - \lfloor nh \rfloor, \lfloor nh \rfloor)} - \hat{\beta}^{(n)}\right). \tag{17}$$

The respective limiting processes remain of course the same, because both estimates of the covariance matrices also converge to Q as $n \to \infty$, but Kuan and Chen (1994) show that they converge faster in dynamic models. Using the same idea it is quite intuitive that the rescaling of the estimates-based processes might also provide benefits for monitoring in dynamic models.

Following Kuan and Chen (1994) we consider three data generating processes (DGPs):

$$y_i = \varrho \cdot y_{i-1} + u_i, \qquad y_0 = 0, \tag{18}$$

$$y_i = 2 + \varrho \cdot y_{i-1} + u_i, \qquad y_0 = 0, \tag{19}$$

$$y_i = 2 + \varrho \cdot x_i + u_i, \qquad x_i = \varrho \cdot x_{i-1} + \varepsilon_i, \tag{20}$$

with u_i and ε_i n.i.d.(0,1), and simulate the size of the corresponding tests for a range of sample sizes and of ϱ with $\alpha = 0.1$ and $h = 0.5$. We use four different values for the sample size ($n = 10, 25, 50, 100$) and two different values for the monitoring period ($T = 2, 10$) and compute the empirical size based on 1000 replications. Our simulations show that the problem is the same in the monitoring case, especially for short history and long monitoring periods: For large values of ϱ the empirical size of the ME test is seriously distorted, see Table 1.

DGP	T	n	Autocorrelation coefficient ϱ									
			0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
(18)	2	10	12.7	14.2	17.1	18.6	21.5	22.0	26.5	32.6	39.7	51.4
		25	10.6	8.3	12.6	13.2	15.1	17.0	21.3	26.1	34.1	48.5
		50	14.1	12.6	12.2	11.5	14.6	14.1	16.7	20.1	24.2	36.4
		100	10.5	8.8	11.2	10.3	12.3	12.8	13.3	17.1	19.9	28.3
	10	10	44.8	54.1	62.6	68.9	74.0	85.5	90.7	94.2	98.6	99.9
		25	21.0	23.7	30.2	39.2	49.9	63.4	75.6	89.5	98.8	100.0
		50	15.8	17.4	21.4	29.3	35.0	46.1	61.7	77.7	93.9	99.7
		100	10.7	12.2	11.9	14.4	14.4	18.8	22.2	26.8	39.2	67.1
(19)	2	10	16.5	24.1	24.4	32.1	39.2	49.9	62.5	81.6	93.5	93.8
		25	13.7	13.0	15.6	17.1	24.9	30.5	42.8	56.2	78.7	98.8
		50	18.3	10.5	11.1	14.7	16.8	18.3	22.7	36.8	62.9	93.7
		100	10.1	11.4	11.5	13.6	14.0	12.9	17.1	21.9	40.0	78.6
	10	10	61.5	69.4	81.2	85.4	92.1	97.0	99.7	100.0	100.0	100.0
		25	23.5	33.6	42.0	50.2	66.0	79.3	93.3	99.2	99.9	100.0
		50	18.6	24.5	30.5	36.0	48.6	62.5	79.1	90.7	99.8	100.0
		100	11.3	11.7	14.0	14.4	19.0	21.3	30.8	38.1	65.1	99.2
(20)	2	10	18.5	17.1	18.2	20.6	23.1	22.4	25.8	28.0	28.7	37.3
		25	12.1	12.1	12.0	13.5	11.5	15.7	15.3	17.9	23.1	26.4
		50	13.2	11.9	9.2	12.3	10.1	13.2	11.7	15.0	14.8	22.5
		100	11.2	9.6	10.1	10.0	8.4	10.9	11.0	11.2	12.4	17.2
	10	10	61.3	67.8	80.0	84.3	92.6	98.0	99.9	100.0	100.0	100.0
		25	24.8	33.8	44.6	55.8	66.6	81.7	93.0	99.8	100.0	100.0
		50	20.2	22.9	29.0	37.5	49.8	63.3	78.3	92.7	99.7	100.0
		100	10.8	10.3	11.9	10.6	11.0	12.0	13.0	13.8	17.3	30.1

Table 1: Empirical size of the ME test

Kuan and Chen (1994) illustrate this phenomenon with the following equation for the DGP (18):

$$\mathrm{E}\left(\frac{Q_{(n)}}{Q_{(\lfloor nt \rfloor)}}\right) = 1 - \frac{2(1+\varrho^2)(n - \lfloor nt \rfloor)}{n \lfloor nt \rfloor (1-\varrho^2)} + O(n^{-2}). \quad (21)$$

The second term on the right hand side is a bias term tending to 0 for $n \to \infty$ for fixed ϱ, but for a fixed sample size n it approaches infinity for $\varrho \to 1$. However this bias is even enhanced if $t > 1$, so that rescaling will even increase the distortion of the empirical size of this test. Therefore rescaling makes no sense in the case of the recursive estimates test for monitoring, but it does for the moving estimates test, because the parameter that determines the window size is not t but h and $h \leq 1$. This is confirmed by our simulations: Table 2 shows that the bias is much smaller for the rescaled processes, especially when the history size n is reasonably large.

DGP	T	n	Autocorrelation coefficient ϱ									
			0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
(18)	2	10	12.0	15.3	16.2	19.3	18.4	23.8	24.1	27.9	32.5	39.4
		25	9.9	12.0	11.7	14.0	15.7	12.9	17.1	21.2	27.3	35.3
		50	11.9	8.7	10.1	10.5	12.3	11.9	12.4	16.0	20.3	28.4
		100	7.7	10.1	10.3	10.3	9.6	10.4	10.8	11.6	15.7	22.7
	10	10	20.0	18.2	23.8	23.5	27.0	29.9	37.2	41.1	51.0	60.7
		25	13.6	12.5	13.0	15.1	15.7	17.4	22.4	26.9	33.6	50.5
		50	10.2	11.2	13.3	13.1	13.4	17.5	18.3	19.8	31.1	40.7
		100	10.9	12.4	10.5	9.9	11.4	13.9	13.9	16.5	19.7	31.1
(19)	2	10	13.2	12.0	15.1	13.9	15.8	15.8	18.9	18.6	22.8	33.5
		25	9.3	9.2	10.2	11.3	12.3	11.6	10.5	12.4	15.2	21.0
		50	11.8	8.4	8.0	10.3	12.3	11.8	13.1	13.1	10.9	15.4
		100	9.6	9.8	9.2	8.5	9.9	9.4	10.7	10.4	10.5	13.6
	10	10	17.4	16.3	19.1	22.0	20.7	24.0	27.9	27.7	31.5	43.4
		25	11.0	10.6	11.2	10.7	14.4	15.0	17.3	17.8	18.1	23.4
		50	10.0	10.4	10.6	11.0	11.6	13.8	14.5	17.4	18.1	17.9
		100	10.8	10.0	9.2	10.9	11.4	10.8	12.3	12.5	13.3	11.4
(20)	2	10	15.2	17.0	17.0	14.6	15.6	18.2	18.5	21.1	21.8	25.4
		25	9.4	10.4	10.6	10.3	10.3	11.7	11.5	13.2	13.3	16.8
		50	10.5	9.8	11.0	7.2	10.6	10.5	11.1	11.7	12.8	16.0
		100	8.4	11.6	9.4	10.5	11.2	9.2	8.0	8.3	11.3	13.2
	10	10	16.8	16.2	19.4	20.3	22.1	22.8	25.1	27.1	32.7	45.5
		25	8.4	9.2	9.9	13.7	12.9	16.1	16.9	21.1	19.7	21.3
		50	9.8	10.4	10.7	10.8	11.8	14.5	17.6	13.8	17.4	18.9
		100	10.1	9.7	9.5	11.4	10.1	11.1	7.6	11.4	12.1	14.5

Table 2: Empirical size of the rescaled ME test

3.4 Boundaries

The shape of the boundaries for empirical fluctuation processes does not make a big difference under the null hypothesis, because they are always chosen to be crossed with the (asymptotic) probability α; but then again under the alternative they can affect very much the chance to detect certain patterns of structural changes. For example, the CUSUM tests (in historical samples) perform poorly if a change occurs late in the sample period. Zeileis (2000) suggests alternative boundaries which are able to increase the detection chances of the OLS-based CUSUM test for early and late changes. Also in the case of monitoring the detection properties for structural changes in the monitoring period strongly depend on the shape of the boundaries—a topic which has not yet been studied in detail.

Chu et al. (1996) already state that the RE test for monitoring has good chances to detect changes early in the monitoring period, but gets increasingly insensitive to late structural changes. This is due to the fact that

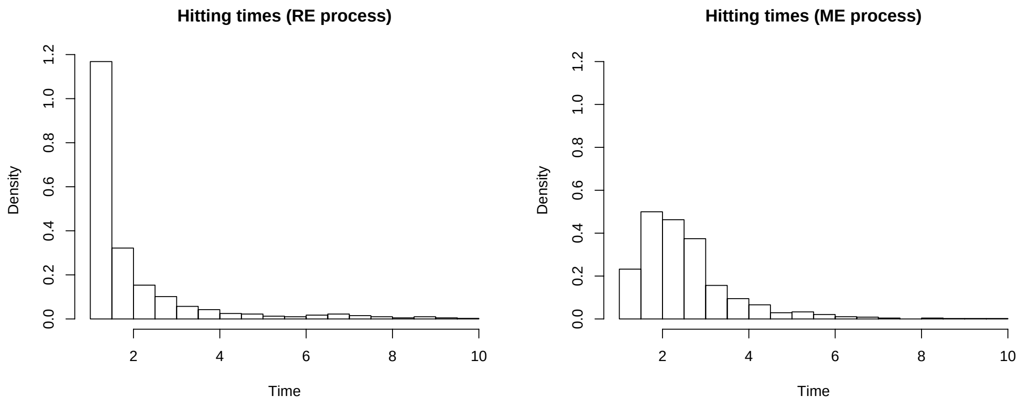

Figure 1: Comparison of hitting times for RE and ME process with standard boundaries

most of the size of the test is used at the very beginning of the monitoring period as Figure 1 shows. It shows the distribution of hitting times (from 10,000 runs) for the Brownian bridge, the asymptotic approximation to the RE process, with the standard boundary (4) at level $\alpha = 0.1$ and for the increments of a Brownian bridge (with $h = 0.5$). It can be seen that the size of the ME test is spread much more evenly; in fact 25 % of the size of the RE test is used on the interval $[1, 1.09]$. This is caused by the shape of the boundary $b_1(t)$, which can be seen in Figure 3: it starts together with the Brownian bridge in 0 at $t = 1$ and so most random crossings will occur very early. We will introduce boundaries for the RE process that distribute the size more evenly. The reason that we control size rather than power is that it is easier on the one hand and also reasonable on the other due to the close connection between the two: typically, the empirical processes start to fluctuate and deviate from their zero mean at the time of the structural change.

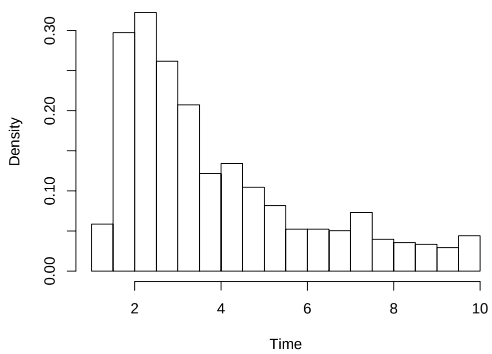

Figure 2: Hitting times for RE process with alternative boundaries

For obtaining a boundary that does not use up the size of the corresponding test at the beginning of the monitoring period it seems natural to choose a boundary with an offset in $t = 1$, but with the correct asymptotical growth rate t. The simplest boundary that fulfills these requirements is

$$b_2(t) \quad = \quad \lambda \cdot t. \tag{22}$$

One might want to consider a boundary which is constant at the beginning of the monitoring period like the boundary (6), but this is inappropriate for a process with growing variance such as the Brownian bridge, because simulations show that most of the size will then be used at the point where the boundary changes from being constant to growing. Because there is no (known) closed-form result for the crossing probability of a Brownian bridge for the boundary (22), we simulate the appropriate critical values of λ for different values of T as for the ME test.

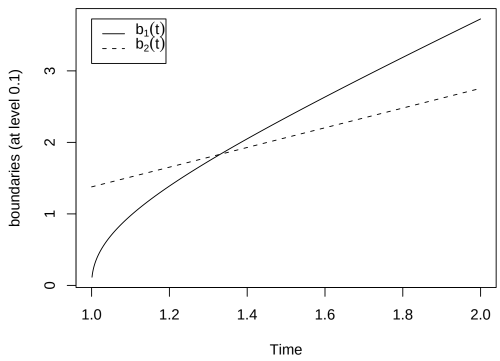

Figure 3: Boundaries for the Brownian bridge

The size of the corresponding test is also distributed more evenly (see Figure 2). Figure 3 shows the resulting alternative boundary $b_2(t)$ (at level 0.1) in comparison to the standard boundary. It can be seen that the boundaries cross at about $t = 1.3$ which means that the detection chances decrease only for very early changes, but increase for all other changes. This is emphasized by simulations under a single shift alternative like in Chu et al. (1996) and Leisch et al. (2000) with the following setup: the data generating process is n.i.d.(2,1) with a history size of $n = 100$ and a monitoring period $T = 10$. Under the alternative the mean switches from 2 to 2.8 in the monitoring period either at $t = 1.1$ or at $t = 3$. Leisch et al. (2000) showed (by simulation with 1,000 replications) that the mean detection delay in this scenario for the RE test (with standard boundaries) at the 10% level is 27, i.e., the structural break at observation $1.1n = 110$ is discovered on average at observation 137. With the new boundaries, the RE test performs a bit (but not dramatically) worse for such an early shift: the mean of the detection delay is 35. However if the the change occurs late in the sample period at $3n = 300$ the mean detection delay for the standard boundaries is 128 but only 100 for the linear boundary $b_2(t)$, the standard deviation is also reduced. The results can also be seen in Table 3 in comparison to the ME test.

shift date	ME	RE (with b_1)	RE (with b_2)
110	32(12)	27(16)	35(14)
300	36(21)	128(74)	100(55)

Table 3: Mean (and standard deviation) of detection delay

It is desirable, of course, to have a more flexible and less heuristic instrument to select the boundaries for fluctuation tests: one might want to choose the boundaries according to a specified prior distribution for the timing of the shift under the alternative. This issue is currently under investigation.

4 Applications

We illustrate the methods introduced in the previous section by applying them to two dynamic models: German M1 money demand (Lütkepohl et al. 1999) and U.S. labor productivity (Hansen 2001).

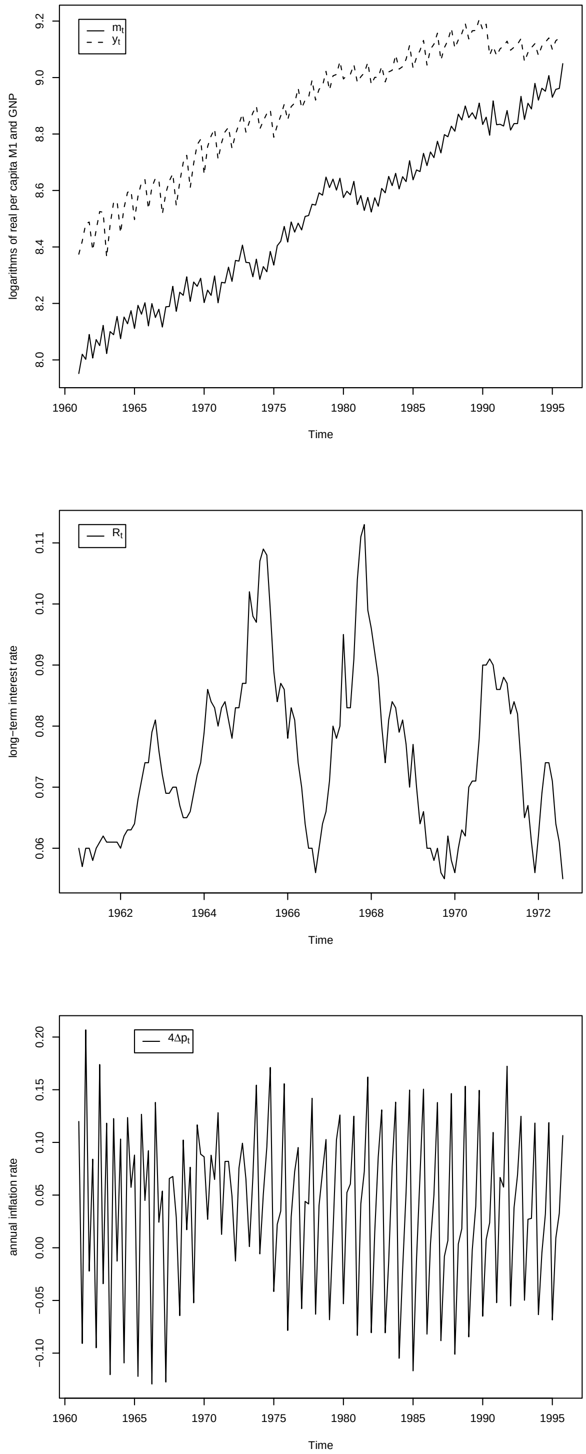

Figure 4: Time series used

4.1 German M1 money demand

Lütkepohl et al. (1999) investigated the stability and linearity of a German M1 money demand function based on data from the German central bank using seasonally unadjusted quarterly data from 1961(1) to 1995(4). The data are available online in the data archive of the Journal of Applied Econometrics (`http://qed.econ.queensu.ca/jae/1999-v14.5/lutkepohl-terasvirta-wolters/`).

Lütkepohl et al. (1999) found a stable relationship for the M1 money demand for the time before the German monetary unification on 1990-06-01 but a clear structural instability for the extended sample period up to 1995(4), which they modelled by smooth transition regression techniques. Specifically, Lütkepohl et al. (1999) established a stable and linear regression relationship for the German M1 money demand using an error correction model (ECM) based on data for the logarithm of real M1 per capita m_t, the logarithm of a price index p_t, the logarithm of the real per capita gross national product y_t and the long-run interest rate R_t. The time series can be seen in Figure 4.

OLS estimation of their model gives the following result for the phase from 1961(1) to 1990(2) before the German monetary unification:

$$\begin{aligned}\Delta m_t &= -0.30\Delta y_{t-2} - 0.67\Delta R_t - 1.00\Delta R_{t-1} - 0.53\Delta p_t \\ &\quad -0.12 m_{t-1} + 0.13 y_{t-1} - 0.62 R_{t-1} \\ &\quad -0.05 - 0.13Q1 - 0.016Q2 - 0.11Q3 + \hat{u}_t,\end{aligned} \tag{23}$$

where $Q1 - Q3$ are seasonal dummies and all coefficients (except the intercept) are highly significant; the fitted model gives an adjusted $R^2 = 0.943$. In a cointegration relationship, the estimators of coefficients on I(1) variables converge at a faster rate than the coefficients on I(0) variables, hence they may be treated as known when monitoring the error correction model. We therefore aggregate the cointegrated variables m_{t-1}, y_{t-1} and R_{t-1} to a single variable $e_{t-1} = -0.12m_{t-1} + 0.13y_{t-1} - 0.62R_{t-1}$ to assure stationarity of this regressor in the ECM. This way of testing for structural change in ECMs is similar to the procedure suggested by Hansen (1992a).

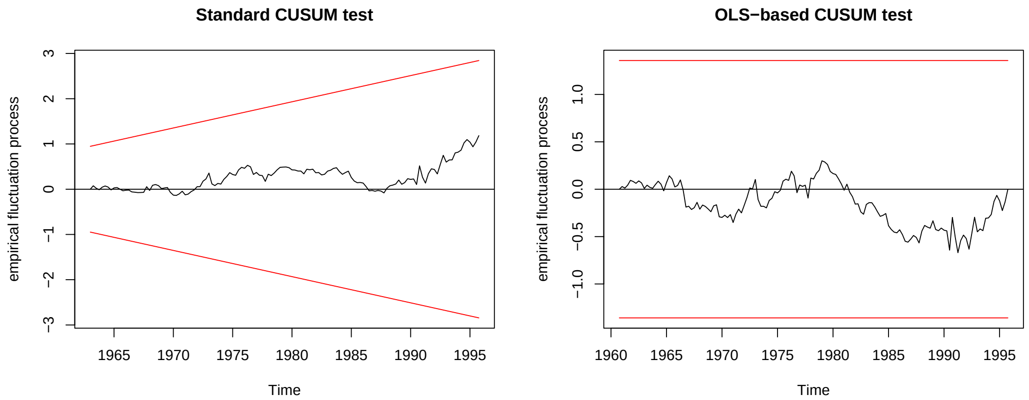

Figure 5: Historic residual-based fluctuation tests

The structural change modelled by Lütkepohl et al. (1999) can be detected with fluctuation tests in two ways: either with historical tests finding a structural change *ex post* or using the monitoring methods introduced in the previous section detecting the structural change *online*. Although we focus on the latter approach in this paper we will first carry out historical tests as well. All computations and simulations have been performed in the statistical software package R (`http://www.R-project.org/`) and in particular the package `strucchange` (Zeileis, Leisch, Hornik, and Kleiber 2002).

Firstly two residual-based fluctuation tests—the recursive (or standard) and the OLS-based CUSUM test—are applied to the model (23). Figure 5 shows that both CUSUM processes lack any significant fluctuation; thus, both tests fail to detect structural change in the data. However, two estimates-based tests—the RE and the ME test—both show a clear peak or shift respectively in the beginning of the 90s (cf. Fig. 5), i.e., both tests detect the structural change after the monetary unification that is also described by Lütkepohl et al. (1999). The reason that the residual-based tests are insensitive to this change while the estimates-based tests are not is the well-known fact that the power of both CUSUM tests depends on the angle between the shift and the mean regressor; in particular they do not have power against shifts orthogonal to the mean regressor (Krämer et al. 1988; Ploberger and Krämer 1992). The suspicion that this might be the case in the given data is confirmed: the estimate for the angle is $90.27°$ (assuming that there is just one structural shift, which is a reasonable hypothesis for the present data).

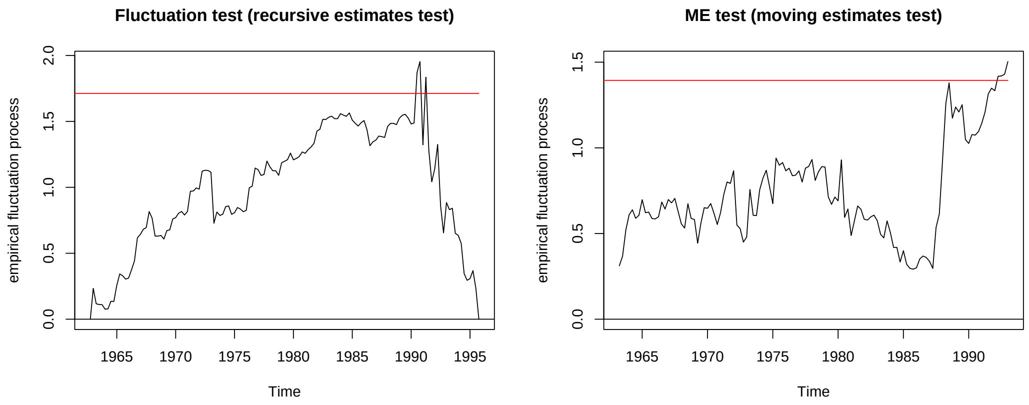

Figure 6: Historic estimates-based fluctuation tests

Now we will confirm the instability of the coefficients in the regression relationship for the money demand function using the tools introduced in Section 3: the OLS-based CUSUM process with the alternative boundary $b_2(t)$ from (22) and the rescaled moving estimates process. We consider the observations from 1961(1)-1990(2) as the history period of the monitoring process and the observations after the monetary unification from 1990(3)-1995(4) as the monitoring period. Thus, we put ourselves in the position of a researcher in 1990 who wants to find out whether the model established for the pre-1990 money demand becomes unstable following the monetary unification.

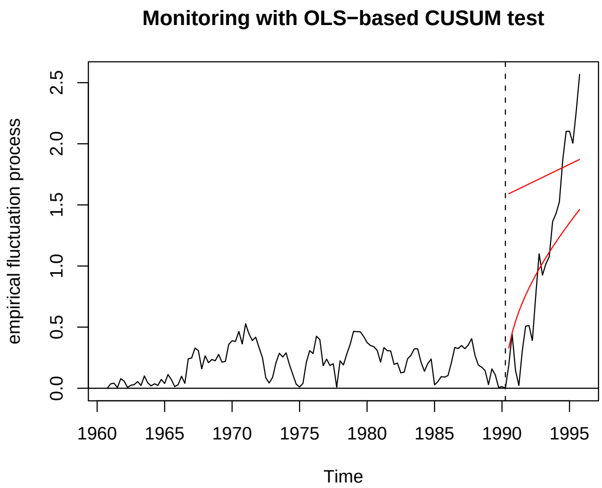

Figure 7: OLS-based CUSUM process

Figure 7 shows the OLS-CUSUM process with the history period left of the vertical dashed line and the monitoring period on the right. Whereas the process does not exhibit much fluctuation before 1990 it does so after the start of the monitoring period and crosses the standard boundary after ten observations in 1992(4) and the alternative boundary another seven observations later. As the break occurs immediately after the end of the monitoring period the standard boundaries perfom a bit better, but note that there is almost a crossing in 1990(4) after just two observations. In a "real" monitoring situation it would be hard to decide if such a crossing was just a type I error or caused by a structural change. The moving estimates process also has a clear shift (see Figure 8), but crosses its boundary a little bit later: in the third quarter of 1994. Hence we can find overwhelming evidence that there has been a structural change in the money demand relationship after the monetary unification.

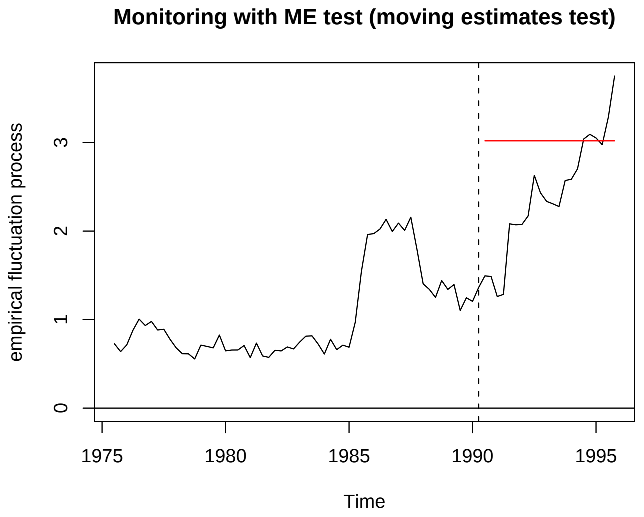

Figure 8: Moving estimates process

4.2 U.S. labor productivity

In his recent overview of "The new econometrics of structural change" Hansen (2001), examines U.S. labor productivity in the manufacturing/durables sector, a monthly time series with observations from 1947(2) through 2001(4) which is available from Bruce Hansen's homepage (`http://www.ssc.wisc.edu/~bhansen/`). He uses a first order autoregressive model for the U.S. labor productivity in the manufacturing/durables sector which is measured by x_t, the growth rate of the Industrial Production Index to average weekly labor hours. The time series for the period considered is depicted in Figure 9.

Hansen (2001) finds a clear structural change in about 1994 and two weaker changes in 1963 and 1982. For illustration, we choose the time from 1964(1) until 1979(12) as the history period, because we are interested in monitoring the two later changes. We exclude the time before 1964, because there must not be a break in the history period. OLS estimation of the coefficients in the historical AR(1) model gives

$$x_t = 0.0025 - 0.186 x_{t-1} + \hat{u}_t, \tag{24}$$

with both coefficients being highly significant. As for the money demand data we monitor the data using the OLS-based CUSUM process with boundaries $b_2(t)$ from (22) and the rescaled moving estimates process.

This approach is slightly different from the one in our first example: there is no known event that might cause an instability in the model considered. We rather assume that we are in a position where we have established a model equation we want to work with, and we want to learn whether we have to update it or not.

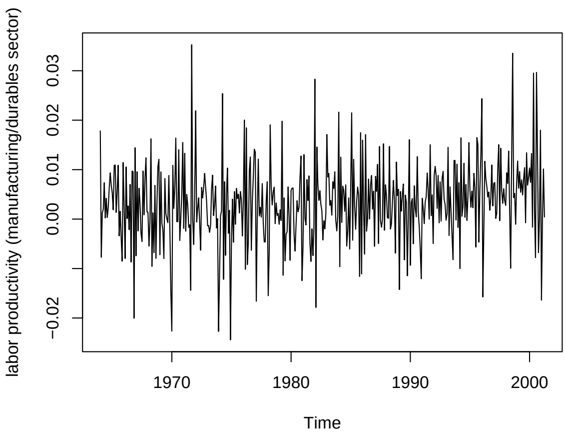

Figure 9: Labor productivity in the manufacturing/durables sector

The OLS-based CUSUM process in Figure 10 reveals two clear structural changes: the first in about 1983, where the path starts to depart from 0, and the second in about 1991. Neither the standard nor the alternative boundaries detect the first shift at the 5% level, but the process crosses both boundaries after the second break: the new boundary already in 1998(2) the standard one only in 2000(5). The moving estimates process (see Figure 11) also exhibits two breaks. However, it drifts off somewhat later compared to the CUSUM process. It also does not detect the first change at the 5% level but crosses its boundary after the second change in 1998(8).

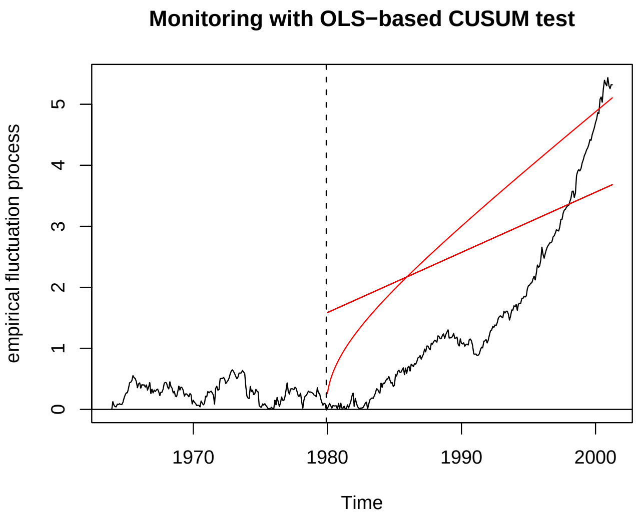

Figure 10: OLS-based CUSUM process

The reason that the OLS-based CUSUM test performs better than the ME test on this particular data set is the usage of the new boundaries. An RE test with standard boundaries would even fail to detect a significant change at the 5% level.

However, the moving estimates have the additional benefit that they allow to determine the nature of the break. The 2-dimensional process for the estimates of the intercept and the coefficient on x_{t-1} can be

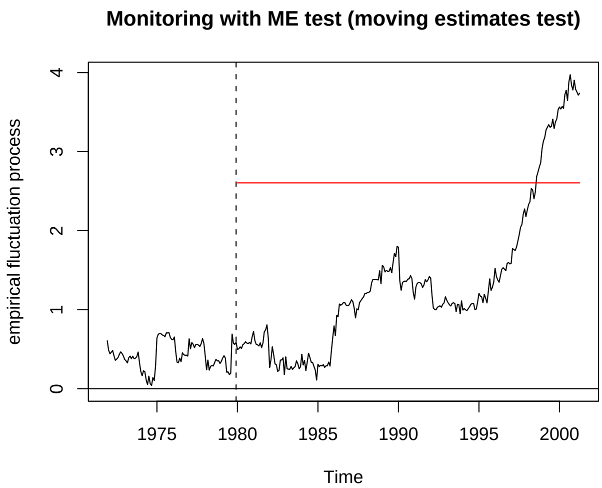

Figure 11: Moving estimates process

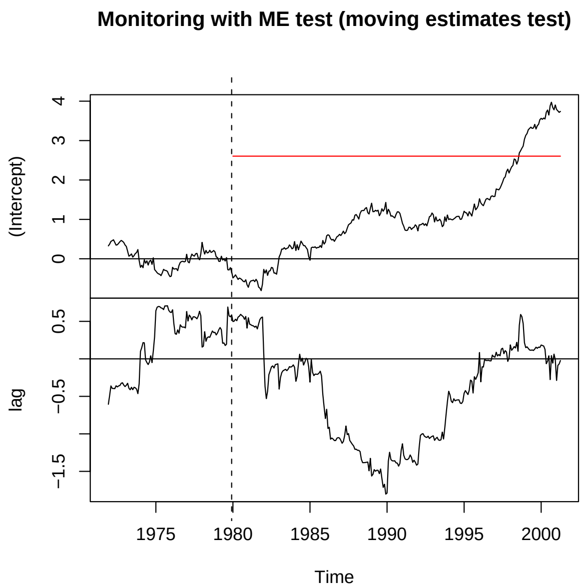

Figure 12: 2-dimensional moving estimates process

plotted separately as in Figure 12. This shows that the break in the 1980s affects both parameters, because both processes have a shift, but not significantly so. The second break in the 1990s just affects the intercept but not the AR coefficient, thus we are able to conclude that the type of the detected structural shift is instability in the intercept term.

5 Conclusions

Online monitoring of regression relationships that are known to be stable for a history period is frequently more natural and more practical than the commonly employed retrospective tests. In this paper, we have presented a unified approach to the online monitoring of econometric models which includes three new extensions to tests based on regression estimates: processes based on OLS residuals, rescaled processes and alternative boundaries. These offer advantages concerning power against certain alternatives, finite sample properties in dynamic models and ease of computation. We have illustrated the feasibility of the methods introduced using fluctuation-type tests on two standard dynamic models, a stationary autoregression and an error correction model.

The determination of optimal asymptotic boundaries, in the sense of minimal detection delay, deserves further study.

Acknowledgements

The research of Achim Zeileis, Friedrich Leisch and Kurt Hornik was supported by the Austrian Science Foundation (FWF) under grant SFB#010 (‘Adaptive Information Systems and Modeling in Economics and Management Science’).

The work of Christian Kleiber was supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 475.

References

Andrews, D. W. K. (1993), “Tests for Parameter Instability and Structural Change With Unknown Change Point,” *Econometrica*, 61, 821–856.

Andrews, D. W. K. and Ploberger, W. (1994), “Optimal Tests When a Nuisance Parameter is Present Only Under the Alternative,” *Econometrica*, 62, 1383–1414.

Brown, R. L., Durbin, J., and Evans, J. M. (1975), “Techniques for Testing the Constancy of Regression Relationships over Time,” *Journal of the Royal Statistical Society*, B 37, 149–163.

Chu, C.-S. J., Stinchcombe, M., and White, H. (1996), “Monitoring Structural Change,” *Econometrica*, 64, 1045–1065.

Hansen, B. E. (1992a), “Testing for Parameter Instability in Linear Models,” *Journal of Policy Modeling*, 14, 517–533.

— (1992b), “Tests for Parameter Instability in Regressions With I(1) Processes,” *Journal of Business & Economic Statistics*, 10, 321–335.

— (2001), “The New Econometrics of Structural Change: Dating Breaks in U.S. Labor Productivity,” *Journal of Economic Perspectives*, 15, 117–128.

Krämer, W., Ploberger, W., and Alt, R. (1988), “Testing for Structural Change in Dynamic Models,” *Econometrica*, 56, 1355–1369.

Kuan, C.-M. and Chen, M.-Y. (1994), “Implementing the Fluctuation and Moving-Estimates Tests in Dynamic Econometric Models,” *Economics Letters*, 44, 235–239.

Kuan, C.-M. and Hornik, K. (1995), "The Generalized Fluctuation Test: A Unifying View," *Econometric Reviews*, 14, 135–161.

Leisch, F., Hornik, K., and Kuan, C.-M. (2000), "Monitoring Structural Changes With the Generalized Fluctuation Test," *Econometric Theory*, 16, 835–854.

Lütkepohl, H., Teräsvirta, T., and Wolters, J. (1999), "Investigating Stability and Linearity of a German M1 Money Demand Function," *Journal of Applied Econometrics*, 14, 511–525.

Ploberger, W. and Krämer, W. (1992), "The CUSUM Test With OLS Residuals," *Econometrica*, 60, 271–285.

Ploberger, W., Krämer, W., and Kontrus, K. (1989), "A New Test for Structural Stability in the Linear Regression Model," *Journal of Econometrics*, 40, 307–318.

Stock, J. H. and Watson, M. W. (1996), "Evidence on Structural Instability in Macroeconomic Time Series Relations," *Journal of Business & Economic Statistics*, 14, 11–30.

Zeileis, A. (2000), "p Values and Alternative Boundaries for CUSUM tests," Working Paper 78, SFB "Adaptive Information Systems and Modelling in Economics and Management Science".

Zeileis, A., Leisch, F., Hornik, K., and Kleiber, C. (2002), "`strucchange`: An R Package for Testing for Structural Change in Linear Regression Models," *Journal of Statistical Software*, 7, 1–38.

Zivot, E. and Andrews, D. W. K. (1992), "Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis," *Journal of Business & Economic Statistics*, 10, 251–270.