A Unified Approach to Structural Change Tests Based on ML Scores, F Statistics, and OLS Residuals

Achim Zeileis
Wirtschaftsuniversität Wien

Abstract

Three classes of structural change tests (or tests for parameter instability) which have been receiving much attention in both the statistics and econometrics communities but have been developed in rather loosely connected lines of research are unified by embedding them into the framework of generalized M-fluctuation tests (Zeileis and Hornik 2003).

These classes are tests based on maximum likelihood scores (including the Nyblom-Hansen test), on F statistics (sup$F$, ave$F$, exp$F$ tests) and on OLS residuals (OLS-based CUSUM and MOSUM tests). We show that (representatives from) these classes are special cases of the generalized M-fluctuation tests, based on the same functional central limit theorem, but employing different functionals for capturing excessive fluctuations.

After embedding these tests into the same framework and thus understanding the relationship between these procedures for testing in historical samples, it is shown how the tests can also be extended to a monitoring situation. This is achieved by establishing a general M-fluctuation monitoring procedure and then applying the different functionals corresponding to monitoring with ML scores, F statistics and OLS residuals. In particular, an extension of the sup$F$ test to a monitoring scenario is suggested and illustrated on a real-world data set.

Keywords: structural change, parameter instability, functional central limit theorem, aggregation functional, fluctuation test, OLS-based CUSUM test, sup$F$ test, Nyblom-Hansen test, monitoring.

1. Introduction

Methods for detecting structural changes or parameter instabilities in parametric models, typically (linear) regression models, have been receiving much attention in both the econometrics and statistics communities. Various classes of tests emerged which have been developed focusing on different properties:

- *ML scores*
  Nyblom (1989) derived an LM test based on maximum likelihood (ML) scores for the alternative that the parameters follow a random walk, which was extended by Hansen (1992) to linear regression models. Recently, Hjort and Koning (2002) suggested a general class of ML score-based structural change tests (without mentioning explicitely that this generalizes the Nyblom-Hansen test).
- *F statistics*
  The class of tests based on F statistics (Wald, LR, and LM test statistics) has been developed for the alternative of a single shift at an unknown timing. The asymptotic theory was established for models estimated by generalized methods of moments (GMM) by Andrews (1993) focusing on the intuitive sup$F$ test and extended by Andrews and Ploberger (1994) who showed that the ave$F$ and exp$F$ tests enjoy certain optimality properties.
- *Fluctuation tests*
  Starting from the recursive CUSUM test of Brown, Durbin, and Evans (1975) a large variety

This is a preprint of an article published in *Econometric Reviews*, 24(4):445–466, 2005

of fluctuation tests for structural change in linear regression models estimated by ordinary least squares (OLS) have been suggested (see Kuan and Hornik 1995, for an overview). These tests are typically derived without having a particular pattern of deviation from parameter stability in mind, and have been emphasized to be also suitable as an explorative tool. In particular, fluctuation tests based on OLS residuals like the OLS-based CUSUM and MOSUM test (Ploberger and Krämer 1992; Chu, Hornik, and Kuan 1995a) are popular because they are easy to compute and to interpret.

Although developed for different alternatives (random walk / single shift / no particular) and for different estimation techniques (ML / GMM / OLS), these tests are more related to each other than obvious at first sight.

In the following, we provide a unifying view on all these structural change tests by embedding representatives from all three classes into the generalized M-fluctuation test framework (Zeileis and Hornik 2003). More precisely, those tests which are based on a single estimate of the parameters on the full sample (and not multiple estimates from recursively growing or rolling subsamples) can be shown to be special cases of the M-fluctuation framework. The M-fluctuation tests are always derived in the following steps: choose a model and an estimation technique (or equivalently its score or estimating function), compute the partial sum process of the scores for which a functional central limit theorem (FCLT) holds, and then compute a statistic by applying a scalar functional that captures the fluctuations in the process. Hence, the unified tests are based on the same FCLT and just use different functionals for computing a test statistic. This view also helps to separate the estimation technique from the functionals employed.

In terms of estimation techniques, we mainly focus on the linear regression model estimated by OLS—this is done only for simplicity and we would like to emphasize that the same types of test statistics can be derived for parameters estimated, e.g., by ML, instrumental variables (IV) or Quasi-ML, among others. GMM is also covered in the case where the number of parameters equals the number of moment restrictions. For the general case, some—but not all (as components of the parameter vector and components of the fluctuation process cannot be matched in general)—properties of the tests discussed can be obtained—see Sowell (1996) or also **?** for robust GMM.

As for the functionals employed, we focus on the most popular tests from the three frameworks discussed, namely the OLS-based CUSUM test, the supLM test and the Nyblom-Hansen test. By understanding the connections between these tests, it becomes more clear what they have in common and also what makes them (and their counterparts which are based on multiple parameter estimates) particularly suitable for certain alternatives. Furthermore, their common features can be exploited, e.g., for deriving new tests in a monitoring situation.

Monitoring structural changes is a topic that gained more attention recently (Chu, Stinchcombe, and White 1996; Leisch, Hornik, and Kuan 2000; Carsoule and Franses 2003; Zeileis, Leisch, Kleiber, and Hornik 2005; Horváth, Huškova, Kokoszka, and Steinebach 2004). It is concerned with detecting parameter instabilities online in a situation where new data is arriving steadily rather than detecting changes ex post in historical samples. Here, we establish an FCLT which yields a general class of M-fluctuation tests for monitoring which has similar unifying properties as for the historical tests. Subsequently, we apply functionals that correspond to monitoring with the OLS-based CUSUM, supLM and Nyblom-Hansen test, respectively. Whereas the OLS-based CUSUM test was considered previously for monitoring (Zeileis *et al.* 2005), new monitoring procedures are derived for the supLM and the Nyblom-Hansen test.

The remainder of this paper is organized as follows: Section 2 briefly reviews the class of generalized M-fluctuation tests into which the other classes of tests are embedded subsequently. Section 3 extends the M-fluctuation tests to the monitoring situation and discusses how the OLS-based CUSUM, supLM, and Nyblom-Hansen test can be employed for monitoring before illustrating the monitoring techniques by a Monte Carlo study and by applying them to a real-world data set. Conclusions are provided in Section 4 and proofs and tables of critical values are attached in an appendix.

2. Generalized M-fluctuation tests

We assume n observations of some dependent variable y_i and a regressor vector x_i, such that the y_i are

$$y_i \quad \sim \quad F(x_i, \theta_i) \qquad (i = 1, \dots, n). \tag{1}$$

following some distribution F with k-dimensional parameter θ_i, conditional on the regressors x_i.[1] The ordering of the observations usually corresponds to time. There are various sets of assumptions under which the results presented below hold, including Krämer, Ploberger, and Alt (1988), Bai (1997) or Andrews (1993).

The hypothesis of interest is "parameter stability", i.e.,

$$H_0: \; \theta_i \quad = \quad \theta_0 \qquad (i = 1, \dots, n) \tag{2}$$

against the alternative that the parameter θ_i changes over time.

To assess this hypothesis, the parameter θ is first estimated by M-estimation, which includes ML, OLS, IV, Quasi-ML, other robust estimation techniques, and is also related to GMM. The parameter estimate $\hat{\theta}$ is computed once for the full sample (assuming H_0 is true) along with a corresponding fluctuation process that captures departures from stability. Both, the estimate and the corresponding fluctuation process, depend on the choice of a suitable estimating function (or score function) $\psi(\cdot)$ which should have zero expectation at the true parameters $\mathsf{E}[\psi(y_i, x_i, \theta_i)] = 0$. Hence, under the null hypothesis the parameter estimate $\hat{\theta}$ can be computed from the first order conditions

$$\sum_{i=1}^{n} \psi(y_i, x_i, \hat{\theta}) \quad = \quad 0 \tag{3}$$

and the decorrelated partial sums of the expression on the left can be used as the fluctuation process capturing structural changes over time. The resulting cumulative score process is referred to as the empirical fluctuation process *efp*$(\cdot)$ and is formally defined as

$$W_n(t, \theta) \quad = \quad n^{-1/2} \sum_{i=1}^{\lfloor nt \rfloor} \psi(y_i, x_i, \theta) \tag{4}$$

$$\mathit{efp}(t) \quad = \quad \hat{J}^{-1/2} \, W_n(t, \hat{\theta}), \tag{5}$$

where $\hat{J}$ is some suitable consistent estimate of the covariance matrix of the scores $\psi(Y_i, \theta)$. The simplest estimator would be $\hat{J} = n^{-1} \sum_{i=1}^{n} \psi(y_i, x_i, \hat{\theta}) \psi(y_i, x_i, \hat{\theta})^\top$ which can be plugged into Equation 5 but also HC or HAC covariance matrix estimators could be used (see Zeileis and Hornik 2003, for more details).

Under the null hypothesis, an FCLT holds: on the interval $[0, 1]$, the empirical fluctuation process *efp*$(\cdot)$ converges to a k-dimensional Brownian bridge $W^0(\cdot)$, which can also be written as $W^0(t) = W(t) - tW(1)$, where $W(\cdot)$ is a standard k-dimensional Brownian motion. Under the alternative, the fluctuation should generally be increased and the process should typically exhibit peaks at the times changes in θ_i occur.

In some situations, it is helpful not to look at the cumulative score process itself but rather some transformation $\widetilde{\mathit{efp}} = \lambda_{\mathrm{trafo}}(\mathit{efp})$. For example, it has been shown in various situations that moving sums instead of cumulative sums are better suited to detect multiple changes. A moving score process can be obtained by transformation with the MOSUM transformation λ_{MOSUM} such that the limiting process is also transformed to $\lambda_{\mathrm{MOSUM}}(W^0(t)) = W^0(t + h) - W^0(t)$, the increments of a Brownian bridge with bandwidth h.

To define a test statistic based on the empirical fluctuation process, a scalar functional is required that captures the fluctuations in the process. The corresponding limiting distribution is then

[1] Instead of using the conditional approach, the distribution of the full vector of observations $(y_i, x_i)^\top$ could also be modelled.

determined by application of the functional to the limiting process. Closed form solutions exist for the distributions implied by certain functionals, but critical values can be obtained easily by simulations for any kind of functional. As the empirical process is essentially a matrix with n observations over time and k components (usually corresponding to parameters), this functional can typically be split up into a functional λ_{comp} which aggregates over the k components and a functional λ_{time} which aggregates over time. If λ_{comp} is applied first, a univariate process is obtained which can be inspected for changes over time. However, applying λ_{time} first results in k independent test statistics such that the component/parameter that causes the instability can be identified. Common choices for λ_{time} are the absolute maximum, the mean or the range and typical functionals λ_{comp} include the maximum norm (or L_∞ norm, denoted as $||\cdot||_\infty$) or the squared Euclidean norm (or L_2 norm, denoted as $||\cdot||_2^2$), see Hjort and Koning (2002) and Zeileis and Hornik (2003) for more examples.

The test statistics unified in this paper are all of the form

$$\lambda_{\mathrm{time}}\left(\frac{\lambda_{\mathrm{comp}}(\mathit{efp}(t))}{d(t)}\right), \tag{6}$$

where $d(\cdot)$ is a weighting function. Hence, statistics based on ML scores, F statistics and OLS residuals can all be shown to be based on the same empirical fluctuation process (and the same FCLT) and to only differ in the choice of the functionals λ_{time}, λ_{comp} and the function d.

By now, we did not specify a precise model to be estimated, i.e., in particular we did not yet specify the estimating functions $\psi(y, x, \theta)$ to be used. As discussed in Section 1, the tests unified in this paper were developed for rather different classes of models (ML / GMM / OLS), but all tests are directly applicable to the model with the greatest practical relevance, the linear regression model. Therefore, we will give some more details about this model, but we would like to emphasize that the results below do not only hold for the linear regression model. The model only determines the estimating functions that are used whereas our results are mainly about functionals for capturing parameter instabilities. However, if some specific estimating function is needed we use that of the linear regression model. In the linear model $y_i = x_i^\top \beta + u_i$ with error variance σ^2 we are faced with the question whether we want to regard $\theta = (\beta, \sigma^2)^\top$ as the parameter vector to be estimated or whether we treat σ^2 as a nuisance parameter and just assess the stability of β. For simplicity, we follow the latter approach and thus use the OLS estimating functions $\psi(y, x, \beta) = (y - x^\top \beta)x$. Furthermore, we assume (for this particular model) that an intercept is included, i.e., that the first component of x_i is equal to unity.

2.1. ML scores

Nyblom (1989) suggested an LM test based on ML scores for the hypothesis of parameter stability against a random walk alternative. Hansen (1992) extended this test to linear regression models where the ML scores and OLS first order conditions both give the estimating functions $\psi(y, x, \beta) = (y - x^\top \beta)x$ already introduced above. Based on these estimating functions (f_t in Hansen's notation, which additionally include a component for the variance σ^2), the cumulative score process $W_n(t, \hat\theta)$ (S_t in Hansen's notation) and the covariance matrix estimate $\hat J$ given above (V in Hansen's notation), Hansen (1992) derives a test statistic called L_C. It is defined in his Equation (9) and can be transformed as follows:

$$\begin{aligned} L_C &= n^{-1} \sum_{i=1}^{n} W_n(i/n, \hat\theta)^\top \hat J^{-1} W_n(i/n, \hat\theta) \\ &= n^{-1} \sum_{i=1}^{n} \mathit{efp}(i/n)^\top \mathit{efp}(i/n) \\ &= n^{-1} \sum_{i=1}^{n} ||\mathit{efp}(i/n)||_2^2 . \end{aligned}$$

Thus, it is a statistic of type (6) where the empirical fluctuation process is first aggregated over the components using the squared Euclidean norm and then over time using the mean. To be more precise, λ_{comp} is $||\cdot||_2^2$, the squared L_2 norm, λ_{time} is the mean and the weighting functions is $d(t) = 1$ for all t. Hence, the limiting distribution is $\int_0^1 ||W^0||_2^2$, the integral of the squared L_2 norm of a k-dimensional Brownian bridge. This functional is also called Cramér-von Mises functional (Anderson and Darling 1952).

Hansen (1992) suggests to compute this statistic for the full process $\mathit{efp}(t)$ to test all coefficients simultaneously and also for each component of the process $(\mathit{efp}(t))_j$ (denoting the j-th component of the process $\mathit{efp}(t)$, $j = 1, \dots, k$) individually to assess which parameter causes the instability. Note, that this approach leads to a violation of the significance level of the procedure if no multiple testing correction is applied. This can be avoided if a functional is applied to the empirical fluctuation process which aggregates over time first yielding k independent test statistics (see Zeileis and Hornik 2003, for more details).

2.2. F statistics

Andrews (1993) and Andrews and Ploberger (1994) suggested three types of test statistics—sup$F$, ave$F$ and exp$F$ statistics—that are based on different kinds of F statistics—Wald, LM or LR statistics—in a very general class of models fitted by GMM. As the statistics are not only easy to interpret but also possess certain optimality properties against single shift alternatives, these tests enjoy great popularity and are probably the most used in practice. The class of GMM estimators considered by Andrews (1993) is similar to the M-estimators considered here except that we only treat the case of pure and not partial structural changes.

Although the asymptotic behaviour for the tests based on Wald, LM and LR statistics is the same, only the test based on LM statistics can be embedded into the framework above because this is the only statistic which is only based on the full sample estimate $\hat{\theta}$. The other two require partial sample estimates before and after a hypothetical breakpoint which is moved over a subset of the sample Π, a closed subset of $(0, 1)$.

Andrews (1993) defines the ingredients for the supLM test in his Equation (4.4): he employs the process of cumulative estimating functions $W_n(t, \hat{\theta})$ ($\bar{m}_{1T}(\tilde{\theta}, \pi)$ in Andrews' notation), and a variance estimate of $\hat{J}^{-1}$ ($\hat{S}^{-1}\hat{M}(\hat{M}\hat{S}^{-1}\hat{M})^{-1}\hat{M}\hat{S}^{-1}$ in Andrews' notation) which is in linear models equivalent to the covariance matrix estimate used in the previous section. This supLM statistic can then be transformed as follows:

$$\begin{aligned}
\sup_{t \in \Pi} LM(t) &= \sup_{t \in \Pi} (t(1-t))^{-1} W_n(t, \hat{\theta}) \hat{J}^{-1} W_n(t, \hat{\theta}) \\
&= \sup_{t \in \Pi} (t(1-t))^{-1} \mathit{efp}(t)^\top \mathit{efp}(t) \\
&= \sup_{t \in \Pi} \frac{||\mathit{efp}(t)||_2^2}{t(1-t)}.
\end{aligned}$$

Therefore, this test statistic is also a special case of (6): the empirical fluctuation process is again first aggregated over the components using the squared L_2 norm, weighted by the variance of the Brownian bridge and then aggregated over time using the supremum over the interval Π. This can be intuitively interpreted as rejecting the null hypothesis when the L_2 aggregated process crosses the boundary $b(t) = c \cdot d(t)$ where c determines the significance level. More precisely, λ_{comp} is again $||\cdot||_2^2$, λ_{time} is $\sup_{t \in \Pi}$, and $d(t) = t(1-t)$. Hence, the limiting distribution is given by $\sup_{t \in \Pi}(t(1-t))^{-1}||W^0(t)||_2^2$.

The aveLM and expLM can be derived analogously, with the same λ_{comp} and d and replacing only λ_{time} by the average and the exp functional respectively.

Another view on the same statistic could be to not use the process efp but $\widetilde{\mathit{efp}} = \lambda_{\mathrm{LM}} \mathit{efp}$ where λ_{LM} is a transformation functional λ_{trafo} defined as $(t(1-t))^{-1}||\cdot||_2^2$. This yields the univariate process of LM statistics which just has to be aggregated over time using the supremum. This view corresponds to the argumentation of Andrews (1993) who establishes the FCLT not at the level of cumulative scores but at the level of F statistics.

For the Wald- and LR-based statistics, the same aggregation functionals are used and the limiting distribution is identical, but on the basis of a fluctuation process that requires estimation of the model on various sub-samples.

2.3. OLS residuals

The mother of all fluctuation tests is the CUSUM test of Brown *et al.* (1975) based on recursive residuals. Ploberger and Krämer (1992) showed how the CUSUM test can also be based on OLS residuals. Computing the test statistic is very simple—the corresponding formula is given in Equation (10) in Ploberger and Krämer (1992)—it is the absolute maximum of the cumulative sums of the OLS residuals scaled by an estimate $\hat{\sigma}^2$ of the error variance. To embed this statistic into the M-fluctuation test framework, the main trick is to exploit that the OLS residuals $\hat{u}_i = y_i - x_i^\top \hat{\beta}$ are the first component of the empirical estimating functions in linear regression models $(\psi(y, x, \beta))_1 = y - x^\top \beta$ when an intercept is included in the regression.

This allows for the following transformation:

$$
\begin{aligned}
\sup_{t \in [0,1]} \left| (\hat{\sigma}^2 n)^{-1/2} \sum_{i=1}^{\lfloor nt \rfloor} \hat{u}_i \right| &= \sup_{t \in [0,1]} \left| \hat{\sigma}^{-1} \, n^{-1/2} \sum_{i=1}^{\lfloor nt \rfloor} y_i - x_i^\top \hat{\beta} \right| \\
&= \sup_{t \in [0,1]} \left| \hat{\sigma}^{-1} \left(W_n(t, \hat{\theta}) \right)_1 \right| \\
&= \sup_{t \in [0,1]} \left| \hat{J}_{1,1}^{-1/2} \left(\hat{J}^{1/2} \mathit{efp}(t) \right)_1 \right|
\end{aligned}
$$

This functional looks rather complicated, but it just selects the first component of the fluctuation process *before* scaling with the full matrix $\hat{J}$ and scales it with the first diagonal element $\hat{J}_{1,1}$ instead which is an estimate of the error variance. As the process $W_n(t, \hat{\theta})$ is not decorrelated, the resulting test statistic captures changes in the conditional *mean* of y and not only in the *intercept* (to which the first component of the decorrelated process *efp* would correspond). More precisely, λ_{comp} is the absolute value of the first component of the scaled non-decorrelated process, λ_{time} is $\sup_{t \in [0,1]}$, and $d(t) = 1$. The corresponding limiting distribution is given by $\sup_{t \in [0,1]} |(W^0(t))_1|$, i.e., the supremum of a 1-dimensional Brownian bridge.

Instead of using the maximum absolute value, various other functionals for capturing the fluctuation in the CUSUM of the OLS residuals have been suggested: Krämer and Schotman (1992) use the range, Ploberger and Krämer (1996) employ the Cramér-von Mises functional (as used in the Nyblom-Hansen test) that provides a test that is trend-resistant, and Zeileis (2004) uses alternative boundaries proportional to the standard deviation of the Brownian bridge $d(t) = \sqrt{t(1-t)}$.

Another approach is to use moving sums instead of cumulative sums (Chu *et al.* 1995a). As pointed out above, the corresponding fluctuation process can be obtained by applying an appropriate transformation λ_{MOSUM} before aggregating the process to a test statistic.

In linear models that only have an intercept ($x_i = 1$, $i = 1, \dots, n$), the OLS-based CUSUM and MOSUM processes are equivalent to the recursive estimates (RE) process (Ploberger, Krämer, and Kontrus 1989) and the moving estimates (ME) process (Chu, Hornik, and Kuan 1995b) which fit regressions on growing or rolling windows of observations respectively. In models with more regressors, the RE and ME test are not special cases of the M-fluctuation test, but the underlying processes converge to the same limiting processes, i.e., a k-dimensional Brownian bridge and its increments respectively. Thus, the situation is similar as for the F statistics: when the model is estimated on multiple sub-samples a test can be obtained which is not strictly a special case but has very similar structural properties and in particular the same limiting distribution.

3. Monitoring with M-fluctuation tests

Monitoring of structural changes is concerned with detecting parameter instabilities online in incoming data, a topic that has been receiving much attention recently. Formally, this means that after the so-called history period of observations $1, \dots, n$ (corresponding to $t \in [0, 1]$) where the parameters are assumed to be stable $\theta_i = \theta_0$, it is tested whether they remain stable for further incoming observations $i > n$ (the monitoring period, corresponding to $t > 1$), The end of this monitoring period may in principle be infinity, but some power might be gained if it is limited to some finite $T > 1$ or $N = \lfloor nT \rfloor$, respectively.

The theory of monitoring structural changes in linear regression models was introduced by Chu *et al.* (1996), who used fluctuation processes based on recursive residuals and recursive estimates. Their test was extended by Leisch *et al.* (2000) to general estimates-based processes. Carsoule and Franses (2003) present an application to score-based processes in autoregressive models and Zeileis *et al.* (2005) discuss several extensions in the context of dynamic econometric models including processes based on OLS residuals and new boundary functions. In the statistical literature, Horváth *et al.* (2004) discuss various residual-based monitoring techniques using different boundary functions.

As illustrated by Carsoule and Franses (2003) and Zeileis *et al.* (2005), there are various different approaches to the application of monitoring for data analysis. The most intuitive is probably in a policy intervention setting where it should be assessed if and when a known intervention becomes effective. In such a situation, it is plausible to establish a fitted model once before the intervention and then compare the incoming data with this fitted model. Another application might be diagnostic checking of a model which is actively used for data analysis during the monitoring period. Here, the practitioner typically wants to update the model with every incoming observation which leads naturally to the recursive/moving estimates monitoring tests that can be carried out with virtually no additional computations. Monitoring is also useful for exploratory analysis of time series, especially when there is a large number of high-frequent series. Tests based on OLS residuals are particularly attractive in such a situation because they are very easy to compute and interpret. For more details see Zeileis *et al.* (2005).

Here we extend these monitoring techniques in two directions: (1) we establish a general class of M-monitoring processes and (2) apply functionals to them corresponding to the Nyblom-Hansen, $\sup LM$, and OLS-based CUSUM test. As for (1), an FCLT has to be established for the extended empirical M-fluctuation processes that makes them applicable to much more general models than only linear regression. The resulting M-monitoring class has unifying properties that are completely analogous to the historical tests. As for (2), appropriate boundary functions have to be chosen. This is different from testing in historical samples where only a single statistic has to be computed whereas monitoring is a sequential testing problem in which some rule is needed how to spread type I errors over the monitoring period.

3.1. Extending the historical tests

Establishing the FCLT is rather straightforward: The parameter $\hat{\theta}$ is still estimated only once on the history period where the parameters are known to be stable, and the empirical fluctuation process $\mathit{efp}(t)$ from Equation (5) is extended by evaluating the estimating functions on new incoming observations (i.e., for $1 < t \leq T$). The resulting process $\mathit{efp}(t) = \hat{J}^{-1/2}\, W_n(t, \hat{\theta})$ still converges to a Brownian bridge $W^0(t) = W(t) - tW(1)$ on the interval $[0, T]$. A formal proof is given in the appendix. The covariance matrix estimate $\hat{J}$ might or might not be the same as for the historical tests, for the FCLT to hold it is only important that it is consistent. In the simplest case, the covariance matrix estimator is also evaluated on the history sample, but in some cases rescaling might be beneficial (Zeileis *et al.* 2005). Based on this FCLT, it is easy to provide the probabilistic ingredients for a monitoring procedure: As for the historical tests, we capture the fluctuation using some scalar functional $\lambda(\mathit{efp}(t))$. But in contrast to the historical setup, this is not only evaluated once, but re-evaluated sequentially for each incoming observation. Thus, we do not need a single critical value but a boundary function $b(t)$ and the hypothesis of parameter stability throughout the monitoring period is rejected if the process $\lambda(efp(t))$ crosses the boundary $b(t)$ for any $t \in [1, T]$. To obtain a sequential testing procedure with asymptotic significance level α, this needs to fulfill $1-\alpha = \mathsf{P}(\lambda(W^0(t)) \le b(t) \mid t \in [1, T])$. For boundaries of type $b(t) = c \cdot d(t)$ in which $d(t)$ determines the shape of the boundary and c the significance level, it is easy to obtain appropriate values of c for any given $d(t)$ by simulation. However, the challenge is to choose a shape $d(t)$ that spreads the power (or size) of the procedure rather evenly (if no further knowledge about the location of potential shifts is available) or directs it at the (potential) timing of the shift (see Zeileis *et al.* 2005; Horváth *et al.* 2004, for a more detailed discussion of boundaries for monitoring).

OLS-based CUSUM test

Applying the functionals corresponding to the historical tests is easiest for the OLS-based CUSUM process. In the linear regression model, the first component of the empirical fluctuation process $\hat{J}_{1,1}^{-1/2} \left(\hat{J}^{1/2} efp(t)\right)_1$ is of course still equivalent to the cumulative sums of the OLS residuals for which appropriate boundaries are discussed in Zeileis *et al.* (2005). They recommend using $d(t) = t$.

supLM test

The basic idea for extending the supLM test to the monitoring setup is also straightforward: in the historical test, the hypothesis of parameter stability is rejected if the process $||efp(t)||_2^2$ crosses a boundary which is proportional to the variance of the Brownian bridge $t(1-t)$. For monitoring, the same idea can be used; the boundary should then be proportional to $t(t-1)$, the variance of the Brownian bridge for $t > 1$. However, this poses the same problem as in the historical test, because at $t = 1$ both the process and the boundary are 0 and it has to be bounded away for the asymptotic theory to be valid. In the historical test, this is done by bounding it away on the time scale, i.e., taking the supremum only over the compact interval Π. For monitoring, this is rather unintuitive because one could not start to monitor directly from the beginning. An alternative approach is to bound it away from zero in the direction of $b(t)$ using some offset. Two conceivable approaches are to add some constant π and thus use $d(t) = t^2 - t + \pi$ or to simply use $d(t) = t^2$ instead of $t^2 - t$. The former is probably more similar in spirit to the historical test, the latter leads to a procedure which can be seen as an extension of the monitoring procedure based on OLS residuals given above. Let us assume for a moment that we have a linear regression model with just one constant regressor $x_i = 1$. Then, $efp(t)$ is the process of cumulative OLS residuals and the OLS-based monitoring procedure rejects the null hypothesis if

$$\begin{aligned} |\,efp(t)\,| \;>\; c \cdot t \quad &\Leftrightarrow \quad (efp(t))^2 \;>\; c^2 \cdot t^2 \\ &\Leftrightarrow \quad ||efp(t)||_2^2 \;>\; c^2 \cdot t^2. \end{aligned}$$

Therefore, the general k-dimensional case using the boundary $b_1(t) = c \cdot t^2$ can be seen as an extension of this 1-dimensional case. For $k = 1$ the squared critical values from Zeileis *et al.* (2005) can be used and are given in Table 2 in the appendix along with new critical values for $k > 1$. Table 3 reports critical values for boundary $b_2(t) = c \cdot (t^2 - t + \pi)$ with $\pi = 0.1$. The boundary b_1 spreads its power rather evenly over the monitoring period while b_2 directs most of its power against changes at the beginning of the monitoring period. This is emphasized by Figure 1 that shows both boundaries for $T = 2$ and $\alpha = 0.1$. It can be seen that the boundaries are crossing at about $t = 1.55$ such that b_1 will perform better for earlier changes and b_2 better for changes that occur later. This is confirmed by simulated hitting times which are depicted in the appendix.

In summary, both boundaries are suitable for capturing fluctuations in the $||efp(t)||_2^2$ process: b_1 can be seen as an extension of the procedure suggested in Zeileis *et al.* (2005) and spreads its power rather evenly while b_1 uses a trimming parameter similar to the historical procedure and is especially suitable for detecting changes early in the monitoring period.

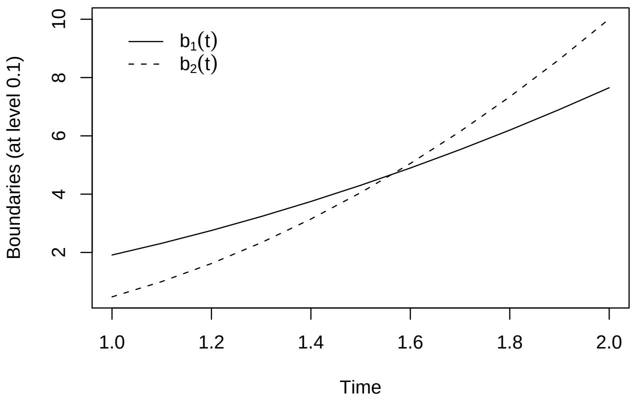

Figure 1: Comparison of boundaries for $||W^0(t)||_2^2$

Nyblom-Hansen test

To extend the Nyblom-Hansen test statistic—the mean of $||efp(t)||_2^2$—to the monitoring situation, a natural idea would be to consider the cumulative mean process $\lfloor nt \rfloor^{-1} \sum_{i=1}^{\lfloor nt \rfloor} ||efp(t)||_2^2$. Suitable boundaries can be found in Borodin and Salminen (2002, p. 378). However, the cumulative mean is varying very slowly and it will become increasingly difficult to detect fluctuations in $efp(t)$. As a low detection delay is crucial in monitoring, this functional does not seem to be very suitable for this task. A way to overcome this problem, at least partially, would be to use a running mean process $n^{-1} \sum_{i=\lfloor nt \rfloor - n+1}^{\lfloor nt \rfloor} ||efp(t)||_2^2$ with bandwidth n instead of the cumulative mean process. Both have in common that the process gives the historical test statistic for $t = 1$. Of course, other bandwidths than n would also be feasible even if they would not yield an immediate extension of the historical statistic. However, none of these processes seems to be promising for monitoring with a low detection delay. Hence, monitoring based on cumulative or running means of squared Euclidian norms is not pursued further here.

3.2. Simulation of size and power

Before applying these monitoring procedures to real-word data, a Monte Carlo study is conducted to study size and power properties in a scenario where the data generating process can be controlled. Following Carsoule and Franses (2003), an AR(2) model is considered

$$y_i \quad = \quad \beta_1 + \beta_2 y_{i-1} + \beta_3 y_{i-2} + u_i \tag{7}$$

where β_1 is the mean, β_2 and β_3 are the autocorrelations at lag 1 and 2 and u_i are standard normal innovations. In the history period ($t \in [0, 1]$, first n observations), the mean is zero and the autocorrelations are 1.2 and -0.4, respectively. In the monitoring period up to $T = 2$, the new incoming observations are tested using the OLS-based CUSUM test and the supLM test with boundaries b_1 and b_2 as defined in the previous section. At time t_0 there is a structural break and the coefficients change to $\beta = (0.5, 1.2, -0.7)^\top$ for $t > t_0$. This is essentially the setup of Carsoule and Franses (2003), but in addition to the autocorrelations we monitor the intercept instead of the variance. Monitoring the variance is also covered by the M-fluctuation framework, but as we have treated the variance as a nuisance parameter for the previous examples, we continue to do so here. As before, the parameters are estimated by OLS and critical values for $\alpha = 0.1$ are used.

All of the critical values can be obtained from the tables in the appendix.[2] In the simulation, the size of the history sample n and the timing of the shift t_0 are varied: n is taken to be either 25, 50, 100 or 500 and t_0 is one of 1.0, 1.25, 1.5 or 2 where the latter corresponds to 'no break'.

Table 1 reports the empirical boundary crossing probabilities from 5,000 replications in each cell. For the first scenario ($t_0 = 2$, no break), this corresponds to the size of the test and for the second ($t_0 = 1$) to power only. For the remaining two scenarios ($t_0 = 1.25$ and 1.5), the empirical boundary crossing probability has to be split up into type I error (crossing for $t \leq t_0$) and power (crossing for $t > t_0$). Confirming the findings of Carsoule and Franses (2003) and Zeileis *et al.* (2005), the tests are somewhat oversized in small samples with pronounced autocorrelations. The power for history samples as small as $n = 25$ has therefore be taken with a grain of salt. However, both size and power improve significantly with the sample size showing a small advantage for the supLM-based tests. This is not surprising as the OLS-based test is only sensitive to changes in the conditional mean. As for the comparison between the boundary b_1 and b_2, it seems that the boundary $b_1(t) = c \cdot t^2$ is more robust to random crossings early in the monitoring period because it is better bounded away from zero while having similar power properties.

In summary, this shows that the tests perform quite well. However, they should be treated carefully when applying them in autoregressive models with strong autocorrelations and/or few observations. Zeileis *et al.* (2005) show that estimates-based tests exhibit similar size distortions in autoregressive models that can be tackled by rescaling the fluctuation processes with different covariance matrix estimates. This is also a potential route of enhancement for score-based processes, but lies beyond the scope of this paper.

		$t_0 = 2$	$t_0 = 1.0$	$t_0 = 1.25$		$t_0 = 1.5$	
n	type	size	power	type I	power	type I	power
25	OLS	27.9	30.8	4.7	22.0	13.9	11.9
	supLM (b_1)	51.8	66.0	18.0	43.4	34.5	21.7
	supLM (b_2)	56.0	69.5	30.8	32.5	44.7	14.9
50	OLS	18.6	29.1	2.9	19.0	9.9	10.2
	supLM (b_1)	36.0	67.2	10.9	46.0	22.9	23.1
	supLM (b_2)	41.1	68.6	22.4	36.4	33.8	15.3
100	OLS	14.9	37.3	1.6	22.5	7.0	12.1
	supLM (b_1)	27.4	86.0	6.0	60.0	15.6	33.1
	supLM (b_2)	32.4	84.2	16.5	48.6	25.6	23.2
500	OLS	11.1	95.4	0.6	72.7	3.7	36.0
	supLM (b_1)	15.1	100.0	1.4	98.5	5.2	88.4
	supLM (b_2)	18.9	100.0	7.6	92.4	13.0	78.0

Table 1: Finite sample size and power (in %) for simulated AR(2) model

3.3. Application to seatbelt data

Although the main purpose of this paper is to give a unifying view on testing and monitoring changes with various functionals and not to suggest new testing/monitoring techniques, we want to illustrate the OLS-based CUSUM test and supLM test for monitoring using a real-world data set. The well-known seatbelt data (Harvey and Durbin 1986) provides a monthly time series from 1969(1) to 1984(12) of the number of car drivers in Great Britain killed or seriously injured in traffic accidents. The series exhibits several breaks, in particular one in 1983(1) associated with the seatbelt law introduction in the UK on 1983-01-31. Harvey and Durbin (1986) analyzed this data set with historical tests, but a monitoring approach would probably have been more natural for evaluating the impact of this policy intervention (had the methodology been available at that

[2]For the OLS-based CUSUM test it is the square root of the value for $k = 1$ in Table 2: 1.383. For the supLM tests the values for $k = 3$ parameters have to be taken from Table 2 and 3: 3.823 and 8.787, respectively.

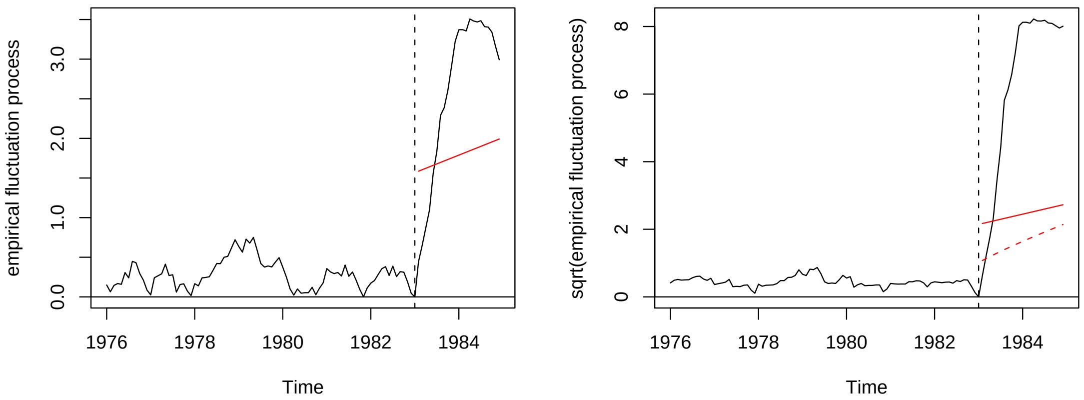

Figure 2: Monitoring with OLS-based CUSUM test (left) and supLM test (right)

time). Here, we monitor the impact of the seatbelt law introduction using the observations from 1976(1) to 1983(1) as the history period—excluding all previous breaks—based on a multiplicative SARIMA$(1,0,0)(1,0,0)_{12}$ model for the log frequencies fitted by OLS as in Zeileis, Kleiber, Krämer, and Hornik (2003).

Figure 2 depicts both monitoring processes—for the OLS-based CUSUM test and the supLM test—along with their boundaries (in red) and a dashed vertical line for the beginning of the monitoring period. Both are based on the same empirical fluctuation process $efp(t)$ computed by using the OLS estimating functions. The OLS-based CUSUM process is computed, just as in the historical case, as the first component of the re-correlated process $\left|\hat{J}_{1,1}^{-1/2}\left(\hat{J}^{1/2}efp(t)\right)_1\right|$ using the usual OLS estimate for the variance as $\hat{J}_{1,1}$. The process shows only small fluctuations in the history period but starts to deviate from 0 immediately after the start of the monitoring period and crosses its boundary $b(t) = 1.568 \cdot t$ (employing the 5% critical value for $T = 2$) in 1983(7), signalling that the seatbelt law intervention was effective. The clear deviation from zero which continues after the boundary crossing emphasizes that this is not a random crossing but is caused by a structural change in the data.

Monitoring with the supLM test leads to very similar results: the right panel of Figure 2 shows the result of monitoring with the process of squared Euclidian norms $||efp(t)||_2^2$ together with the boundaries $b_1(t) = 4.603 \cdot t^2$ (solid line) and $b_2(t) = 10.334 \cdot (t^2 - t + 0.1)$ (dashed line). To make the graph more intellegible, the square root of the process and its boundaries is plotted. It also clearly deviates from zero with the beginning of the monitoring period, crosses both boundaries and thus also clearly signals a structural change. The boundary b_1 is crossed in 1983(5) and b_2 (not surprisingly) a bit earlier in 1983(3). In summary, all three methods perform very similar on this data set and are all able to detect the effect of the policy intervention quickly after only a few observations in the monitoring period.

4. Conclusions

In this paper, we provide a unifying view on three classes of structural change tests by embedding them into the framework of generalized M-fluctuation tests. The three classes are tests based on ML scores, F statistics and OLS residuals which have been developed in rather loosely connected lines of research. Special emphasis is given to the most prominent representatives from these classes, namely the Nyblom-Hansen test, the supLM test and the OLS-based CUSUM test, which can be shown to be based on the same empirical fluctuation process, only employing different functionals for capturing excessive fluctuations within the process.

The knowledge about the connections between these historical tests is subsequently used to extend the tests to online monitoring of structural changes. To accomplish this, a general FCLT for empirical M-fluctuation processes in a monitoring situation is established and several strategies for extending the supLM and Nyblom-Hansen test are discussed. Finally, the methods are illustrated in a policy intervention context for the UK seatbelt data.

Computational details

The results in this paper were obtained using R 2.1.1 (R Development Core Team 2005, `http://www.R-project.org/`) and the package **strucchange** 1.2-11 (Zeileis, Leisch, Hornik, and Kleiber 2002) which are both freely available at no cost under the terms of the GNU General Public Licence (GPL) from the Comprehensive R Archive Network at `http://CRAN.R-project.org/`.

Acknowledgements

We are thankful to Christian Kleiber, Friedrich Leisch and Pal Révész for helpful comments and discussions.

References

Anderson TW, Darling DA (1952). "Asymptotic Theory of Certain Goodness-of-fit Criteria Based on Stochastic Processes." *Annals of Mathematical Statistics*, **23**, 193–212.

Andrews DWK (1993). "Tests for Parameter Instability and Structural Change With Unknown Change Point." *Econometrica*, **61**, 821–856.

Andrews DWK, Ploberger W (1994). "Optimal Tests When a Nuisance Parameter is Present Only Under the Alternative." *Econometrica*, **62**, 1383–1414.

Bai J (1997). "Estimation of a Change Point in Multiple Regression Models." *Review of Economics and Statistics*, **79**, 551–563.

Borodin AN, Salminen P (2002). *Handbook of Brownian Motion – Facts and Formulae*. Birkhäuser, Basel, 2nd edition.

Brown RL, Durbin J, Evans JM (1975). "Techniques for Testing the Constancy of Regression Relationships over Time." *Journal of the Royal Statistical Society B*, **37**, 149–163.

Carsoule F, Franses PH (2003). "A Note on Monitoring Time-Varying Parameters in an Autoregression." *Metrika*, **57**, 51–62.

Chu CSJ, Hornik K, Kuan CM (1995a). "MOSUM Tests for Parameter Constancy." *Biometrika*, **82**, 603–617.

Chu CSJ, Hornik K, Kuan CM (1995b). "The Moving-Estimates Test for Parameter Stability." *Econometric Theory*, **11**, 669–720.

Chu CSJ, Stinchcombe M, White H (1996). "Monitoring Structural Change." *Econometrica*, **64**(5), 1045–1065.

Hansen BE (1992). "Testing for Parameter Instability in Linear Models." *Journal of Policy Modeling*, **14**, 517–533.

Harvey AC, Durbin J (1986). "The Effects of Seat Belt Legislation on British Road Casualties: A Case Study in Structural Time Series Modelling (with Discussion)." *Journal of the Royal Statistical Society, Series A*, **149**, 187–227.

Hjort NL, Koning A (2002). "Tests for Constancy of Model Parameters Over Time." *Nonparametric Statistics*, **14**, 113–132.

Horváth L, Huškova M, Kokoszka P, Steinebach J (2004). "Monitoring Changes in Linear Models." *Journal of Statistical Planning and Inference*, **126**, 225–251.

Krämer W, Ploberger W, Alt R (1988). "Testing for Structural Change in Dynamic Models." *Econometrica*, **56**(6), 1355–1369.

Krämer W, Schotman P (1992). "Range vs. Maximum in the OLS-based Version of the CUSUM Test." *Economics Letters*, **40**, 379–381.

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

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

Nyblom J (1989). "Testing for the Constancy of Parameters Over Time." *Journal of the American Statistical Association*, **84**, 223–230.

Ploberger W, Krämer W (1992). "The CUSUM Test With OLS Residuals." *Econometrica*, **60**(2), 271–285.

Ploberger W, Krämer W (1996). "A Trend-Resistant Test for Structural Change Based on OLS Residuals." *Journal of Econometrics*, **70**, 175–185.

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

R Development Core Team (2005). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3, URL http://www.R-project.org/.

Sowell F (1996). "Optimal Tests for Parameter Instability in the Generalized Methods of Moments Framework." *Econometrica*, **64**, 1085–1107.

Zeileis A (2004). "Alternative Boundaries for CUSUM Tests." *Statistical Papers*, **45**, 123–131.

Zeileis A, Hornik K (2003). "Generalized M-Fluctuation Tests for Parameter Instability." *Report 80*, SFB "Adaptive Information Systems and Modelling in Economics and Management Science". URL http://www.wu-wien.ac.at/am/reports.htm#80.

Zeileis A, Kleiber C, Krämer W, Hornik K (2003). "Testing and Dating of Structural Changes in Practice." *Computational Statistics & Data Analysis*, **44**(1–2), 109–123.

Zeileis A, Leisch F, Hornik K, Kleiber C (2002). "strucchange: An R Package for Testing for Structural Change in Linear Regression Models." *Journal of Statistical Software*, **7**(2), 1–38. URL http://www.jstatsoft.org/v07/i02/.

Zeileis A, Leisch F, Kleiber C, Hornik K (2005). "Monitoring Structural Change in Dynamic Econometric Models." *Journal of Applied Econometrics*, **20**, 99–121.

A. Appendix

A.1. Proofs

In Zeileis and Hornik (2003), it is shown that the empirical fluctuation process from Equation (5) converges to a Brownian bridge on the unit interval $[0, 1]$. Here, the results are extended to any compact interval $[0, T]$ with $T \geq 1$. As in the proofs of Zeileis and Hornik (2003) the fact that $t \in [0, 1]$ is never needed, the same argumentation can be used. Therefore, we just sketch the most important steps using the same notation.

$$A(\theta) = \mathsf{E}[-\psi'(y_i, x_i, \theta)], \tag{8}$$
$$J(\theta) = \mathsf{VAR}[\psi(y_i, x_i, \theta)], \tag{9}$$

where $y_i \sim F(x_i, \theta_0)$, $\psi'(\cdot)$ is the partial derivative of $\psi(\cdot)$ with respect to θ.

Under suitable regularity conditions, $\hat{\theta}$ is consistent for θ_0 under the null hypothesis and $\sqrt{n}(\hat{\theta} - \theta_0)$ is asymptotically normal with zero mean and covariance matrix $A(\theta)^{-1} J(\theta) \{A(\theta)^{-1}\}^\top$. Equivalently, we can write

$$\sqrt{n}(\hat{\theta} - \theta_0) \quad \dot{=} \quad A(\theta_0)^{-1} \cdot W_n(1, \theta_0), \tag{10}$$

where $a_n \dot{=} b_n$ means that $a_n - b_n$ tends to zero in probability.

Applying a first order Taylor expansion then yields the FCLT:

$$\begin{aligned} W_n(t, \hat{\theta}_n) &\dot{=} \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} \psi(y_i, x_i, \theta_0) + \frac{1}{n} \sum_{i=1}^{\lfloor nt \rfloor} \psi'(y_i, x_i, \theta_0) \cdot \sqrt{n}(\hat{\theta} - \theta_0) \\ &\dot{=} W_n(t, \theta_0) - \frac{\lfloor nt \rfloor}{n} A(\theta_0) \cdot A(\theta_0)^{-1} W_n(1, \theta_0) \\ &\xrightarrow{\text{d}} Z(t) - t \cdot Z(1), \end{aligned}$$

where $Z(\cdot)$ is a Gaussian process with continuous paths, mean function $\mathsf{E}[Z(t)] = 0$ and covariance function $\mathsf{COV}[Z(t), Z(s)] = \min(t, s) \cdot J(\theta_0)$. Therefore, with a consistent non-singular estimate $\hat{J}$ of $J(\theta_0)$ $\mathit{efp}(t) = \hat{J}^{-1/2} W_n(t, \hat{\theta})$ converges to a Brownian bridge $W^0(t) = W(t) - tW(1)$.

A.2. Monitoring with supLM test

For monitoring with the supLM test, the process $||\mathit{efp}(t)||_2^2$ is used and the hypothesis of parameter stability is rejected if this process crosses a boundary of type $b(t) = c \cdot d(t)$ in the monitoring period $[1, T]$. The function $d(t)$ determines the shape of the boundary and above we have suggested using $d(t) = t \cdot (t - 1) +$ trimming and in particular $d(t) = t^2$ (in b_1) or $d(t) = t^2 - t + 0.1$ (in b_2). Under the null hypothesis, the process $||\mathit{efp}(t)||_2^2$ converges to the Euclidean norm process of a k-dimensional Brownian bridge $||W^0(t)||_2^2$ on $[0, T]$ and hence the critical value c has to be chosen such that the following equation holds:

$$\mathsf{P}\left(||W^0(t)||_2^2 \; < \; c \cdot d(t) \quad | \quad t \in [1, T]\right) \quad = \quad 1 - \alpha.$$

Suitable simulated values of c for selected values of α, k and T are provided in Tables 2 and 3 for the boundaries b_1 and b_2. Each of these is based on 10,000 replications, where each Brownian bridges is simulated from 10,000 normal pseudo-random numbers per unit time interval.

To compare the properties of different monitoring procedures, Zeileis *et al.* (2005) employ histograms of hitting times for the limiting process (under the null hypothesis). Using this approach, insight is gained how the test spreads its size (and typically also power) over the monitoring interval without having to focus on a small set of alternatives from the infinite set of conceivable patterns of deviation from parameter stability. Figures 3 and 4 depict the hitting times derived from 1-dimensional and 5-dimensional Brownian bridges with boundaries b_1 and b_2 at 10% significance level. Both show that b_2 directs most of its size to the beginning of the monitoring period whereas b_1 spreads it a bit more evenly such that the corresponding monitoring procedure will have more power against changes that occur very late in the monitoring period. Comparing the hitting time distributions for $k = 1$ and $k = 5$, the pictures are very similar but somewhat shifted to the right in the latter case.

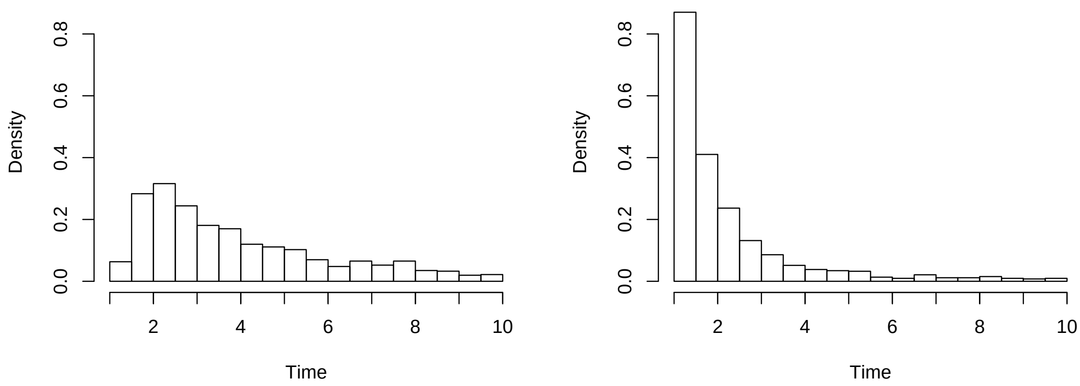

Figure 3: Hitting times for $||W^0(t)||_2^2$ process with $k = 1$ and boundary b_1 (left) and b_2 (right)

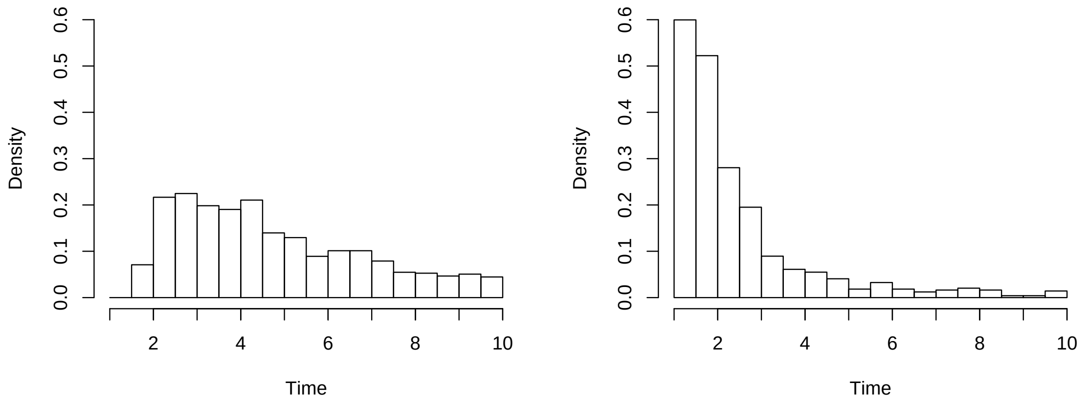

Figure 4: Hitting times for $||W^0(t)||_2^2$ process with $k = 5$ and boundary b_1 (left) and b_2 (right)

k	α (in %)	T 1.25	1.5	2	3	4	6	8	10
1	20.0	0.541	0.917	1.343	1.766	2.045	2.256	2.375	2.455
	15.0	0.628	1.064	1.570	2.088	2.384	2.621	2.782	2.849
	10.0	0.754	1.291	1.913	2.528	2.873	3.201	3.378	3.460
	5.0	0.979	1.690	2.459	3.291	3.760	4.186	4.368	4.528
	1.0	1.570	2.669	3.905	5.290	5.871	6.620	6.744	7.022
	0.1	2.353	3.827	5.929	7.779	8.839	10.407	10.433	11.567
2	20.0	0.876	1.481	2.182	2.932	3.316	3.702	3.950	4.064
	15.0	0.989	1.676	2.477	3.302	3.753	4.195	4.475	4.612
	10.0	1.161	1.948	2.875	3.849	4.394	4.932	5.201	5.316
	5.0	1.440	2.460	3.525	4.846	5.407	6.010	6.535	6.612
	1.0	2.055	3.494	5.058	7.051	7.721	8.789	9.255	9.245
	0.1	3.164	4.622	7.054	9.648	10.438	12.939	14.190	13.764
3	20.0	1.183	1.953	2.974	3.915	4.469	5.036	5.136	5.316
	15.0	1.322	2.174	3.315	4.359	4.946	5.563	5.744	5.963
	10.0	1.523	2.503	3.823	4.959	5.632	6.307	6.598	6.855
	5.0	1.817	3.030	4.603	6.016	6.816	7.621	8.006	8.329
	1.0	2.532	4.195	6.392	8.350	9.381	10.360	11.212	11.534
	0.1	3.548	5.630	9.047	11.422	12.697	14.876	15.491	15.671
4	20.0	1.454	2.417	3.631	4.896	5.532	6.085	6.380	6.524
	15.0	1.607	2.680	3.979	5.407	6.079	6.704	7.079	7.228
	10.0	1.814	3.064	4.539	6.123	6.872	7.608	7.948	8.083
	5.0	2.151	3.661	5.375	7.266	8.125	9.043	9.489	9.741
	1.0	2.861	4.955	7.240	9.682	11.012	12.280	12.457	13.044
	0.1	3.932	6.598	10.092	12.876	14.164	16.875	16.653	17.439
5	20.0	1.714	2.897	4.327	5.803	6.461	7.217	7.479	7.790
	15.0	1.875	3.197	4.692	6.365	7.105	7.896	8.214	8.541
	10.0	2.090	3.598	5.256	7.162	7.917	8.873	9.216	9.604
	5.0	2.463	4.232	6.135	8.372	9.320	10.388	10.838	11.172
	1.0	3.224	5.519	8.178	11.022	12.082	13.811	14.356	14.858
	0.1	4.284	7.078	11.076	14.259	16.324	19.442	18.021	20.323
10	20.0	2.967	5.010	7.445	9.885	11.281	12.498	13.213	13.383
	15.0	3.176	5.378	7.952	10.569	12.115	13.405	14.145	14.392
	10.0	3.458	5.884	8.658	11.545	13.138	14.631	15.453	15.716
	5.0	3.897	6.691	9.753	13.094	14.824	16.446	17.581	17.834
	1.0	4.897	8.386	12.516	16.318	18.317	20.212	21.394	22.346
	0.1	6.079	10.407	15.855	20.095	22.292	25.794	26.585	28.056
15	20.0	4.176	7.001	10.361	13.999	15.818	17.404	18.191	18.690
	15.0	4.424	7.428	10.947	14.867	16.782	18.346	19.371	19.842
	10.0	4.747	7.950	11.862	15.931	18.107	19.769	20.787	21.441
	5.0	5.298	8.778	13.202	17.704	20.028	21.862	23.135	23.920
	1.0	6.449	10.660	16.025	21.607	24.251	26.250	28.078	29.326
	0.1	7.626	12.553	19.416	26.807	30.234	31.283	32.880	35.525

Table 2: Simulated critical values for supLM test with boundary b_1

k	α (in %)	T 1.25	1.5	2	4	5	6	8	10
1	20.0	2.465	3.069	3.554	3.878	3.969	4.128	4.146	4.180
	15.0	2.846	3.516	4.053	4.452	4.536	4.724	4.760	4.716
	10.0	3.389	4.153	4.766	5.191	5.268	5.439	5.522	5.434
	5.0	4.330	5.233	6.043	6.373	6.605	6.904	6.762	6.750
	1.0	6.603	7.867	9.064	9.402	9.831	10.185	9.796	10.350
	0.1	9.638	11.148	13.748	13.571	14.382	15.250	15.727	14.254
2	20.0	3.869	4.800	5.506	5.877	6.125	6.231	6.291	6.317
	15.0	4.290	5.365	6.071	6.473	6.822	6.907	6.945	7.030
	10.0	4.968	6.180	6.941	7.358	7.759	7.862	7.830	7.915
	5.0	6.037	7.442	8.411	8.973	9.255	9.361	9.249	9.387
	1.0	8.388	10.231	11.705	12.477	12.858	12.816	12.760	12.940
	0.1	11.749	13.687	16.285	17.155	17.809	16.242	18.481	17.510
3	20.0	5.018	6.319	7.170	7.578	7.825	8.027	8.043	8.180
	15.0	5.570	6.943	7.864	8.284	8.537	8.815	8.770	8.964
	10.0	6.316	7.758	8.787	9.233	9.543	9.802	9.810	10.012
	5.0	7.515	9.056	10.334	10.825	11.097	11.599	11.559	11.772
	1.0	10.454	12.486	13.799	14.624	14.745	15.682	15.297	15.462
	0.1	15.296	16.253	18.294	19.459	19.822	20.143	19.861	20.332
4	20.0	6.229	7.620	8.579	9.190	9.516	9.658	9.671	9.797
	15.0	6.787	8.308	9.296	10.011	10.338	10.499	10.442	10.631
	10.0	7.545	9.268	10.293	11.053	11.534	11.631	11.597	11.785
	5.0	8.814	10.884	11.920	12.696	13.312	13.483	13.452	13.595
	1.0	11.719	14.253	15.352	16.566	17.422	17.871	17.574	17.660
	0.1	15.599	17.773	20.459	22.374	22.532	22.710	22.012	23.279
5	20.0	7.294	8.841	9.890	10.787	10.951	11.165	11.301	11.348
	15.0	7.938	9.673	10.682	11.623	11.828	12.054	12.143	12.271
	10.0	8.756	10.786	11.770	12.736	12.985	13.293	13.385	13.421
	5.0	10.163	12.286	13.611	14.686	15.078	15.157	15.334	15.447
	1.0	13.147	15.439	17.088	18.176	18.985	19.285	19.563	19.613
	0.1	16.797	19.179	22.383	23.175	23.759	24.916	25.602	25.863
10	20.0	12.157	14.718	16.534	17.475	17.971	18.128	18.294	18.429
	15.0	13.002	15.746	17.543	18.591	19.083	19.215	19.460	19.515
	10.0	14.048	16.911	18.903	20.047	20.462	20.774	20.992	20.943
	5.0	15.796	18.842	21.169	22.207	22.778	23.149	23.397	23.477
	1.0	19.450	22.810	25.853	27.226	27.880	28.133	28.678	28.393
	0.1	25.148	27.875	33.228	32.496	33.011	32.983	35.636	36.626
15	20.0	16.787	20.542	22.493	23.796	24.305	24.572	24.821	24.972
	15.0	17.783	21.588	23.643	25.066	25.552	25.823	26.067	26.279
	10.0	18.988	23.190	25.187	26.768	27.216	27.403	27.638	28.063
	5.0	21.144	25.548	27.723	29.384	29.684	30.140	30.377	30.536
	1.0	25.641	30.410	33.347	35.561	35.850	35.640	35.652	36.259
	0.1	31.345	36.453	40.389	42.309	43.366	41.742	42.356	43.278

Table 3: Simulated critical values for supLM test with boundary b_2