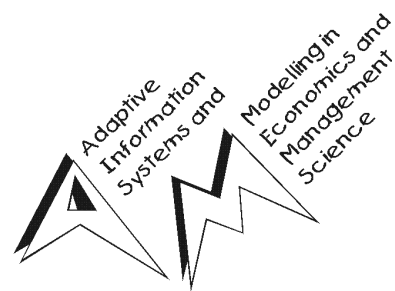Report Series

Generalized M-Fluctuation Tests for Parameter Instability

Achim Zeileis
Kurt Hornik

Report Series

June 2003

Papers published in this report series
are preliminary versions of journal articles.

Generalized M-Fluctuation Tests for Parameter Instability

Achim Zeileis

Institut für Statistik & Wahrscheinlichkeitstheorie,
Technische Universität Wien, Austria

Kurt Hornik

Institut für Statistik,
Wirtschaftsuniversität Wien, Austria

Abstract

A general class of fluctuation tests for parameter instability in an M-estimation framework is suggested. The tests are based on partial sum processes of M-estimation scores for which functional central limit theorems are derived under the null hypothesis of parameter stability and local alternatives. Special emphasis is given to parameter instability in (generalized) linear regression models and it is shown that the introduced M-fluctuation tests contain a large number of parameter instability or structural change tests known from the statistics and econometrics literature. The usefulness of the procedures is illustrated using artificial data and data for the German M1 money demand, historical demographic time series from Großarl, Austria, and youth homicides in Boston.

1 Introduction

Structural change is of central interest in many fields of research and data analysis: to learn if, when and how the structure of the mechanism underlying a set of observations changes. In parametric models structural change is typically described by parameter instability. If this instability is ignored, parameter estimates are generally not meaningful, inference is severely biased and predictions lose accuracy. Therefore, a large literature on tests for structural change or parameter instability emerged in particular in the econometrics community using the tests as tools for diagnostic checking against misspecification (see Hansen 2001, for a recent review). But more generally, such tests can also be used as explorative tools that can help to understand the structure in the data.

Usually, in structural change problems it is known with respect to which quantity the instability occurs: e.g., in time series regression it is natural to ask whether the relationship between dependent and explanatory variables changes over time. In clinical studies often changepoint problems arise where a regression relationship changes with respect to the size of one risk factor etc. The situations described have in common that the observations have some unique ordering with respect to which a structural change occurs but that the (potential) changepoint is unknown. Starting from the Recursive CUSUM test of Brown, Durbin, and Evans (1975) a large variety of tests has been suggested in both the econometrics and the statistics literature many of which can be broadly distinguished to belong to two different classes: generalized fluctuation tests (Kuan and Hornik 1995) that do not assume a particular pattern of deviation from the hypothesis of parameter constancy and F tests (Andrews 1993; Andrews and Ploberger 1994) that are built for a single shift alternative (of unknown timing).

The generalized fluctuation tests fit a parametric model to the data via ordinary least squares (OLS) —or equivalently via maximum likelihood (ML) using a normal approximation—and derive a process which captures the fluctuation of the recursive or OLS residuals (Brown et al. 1975; Ploberger and Krämer 1992; Chu, Hornik, and Kuan 1995a) or the recursive or rolling/moving estimates (Ploberger, Krämer, and Kontrus 1989; Chu, Hornik, and Kuan 1995b) and reject if this fluctuation is improbably large. In their seminal paper Brown et al. (1975) point out that this framework "... includes formal significance tests but its philosophy is basically that of data analysis as expounded by Tukey (1962). Essentially, the techniques are designed to *bring out departures*

1

from constancy in a graphic way instead of parametrizing particular types of departure in advance and then developing formal significance tests intended to have high power against these particular alternatives." (Brown et al. 1975, pp. 149–150)

Instead of capturing the fluctuation in residuals or parameter estimates empirical fluctuation processes can also be based on OLS first order conditions or ML scores respectively (Nyblom 1989; Hansen 1992; Hjort and Koning 2002). Although they have not been discussed in the generalized fluctuation test framework (Kuan and Hornik 1995) we show below that they can be seen as an extension of that framework which makes the class of empirical fluctuation processes richer. In this paper we show how under mild assumptions these ideas can be used in more general situations and employing different estimation techniques (not only OLS and ML). The resulting class of tests for parameter instability which is based on M-estimation scores contains many of the tests mentioned above as special cases and unifies the approaches to the construction of test statistics. Strategies are outlined for combining traditional significance testing with visualization methods for detecting the timing of a potential shift and which parameter is affected by it. More precisely, in Section 2 we introduce the parametric model and formulate the null hypothesis before we derive functional central limit theorems for partial sum processes of M-scores under the hypothesis of parameter stability and under local alternatives in Section 3. The construction of the generalized M-fluctuation tests—from the choice of the estimation technique to the test statistic that captures the fluctuation in the M-score processes—is described in Section 4. Section 5 discusses tests based on ML-scores in (generalized) linear models and the usefulness of the proposed tests is illustrated in Section 6 based on data for German M1 money demand, historical demographic time series of illegitimate births in Großarl, Austria, and youth homicides in Boston, USA, in a policy intervention framework.

2 The model

We assume n independent observations

$$Y_i \quad \sim \quad F(\theta_i) \qquad (i = 1, \ldots, n). \tag{1}$$

distributed according to some distribution F with k-dimensional parameter θ_i. We also assume that the observations are uniquely ordered by some external variable, usually time. The Y_i can possibly be vector valued, extensions to a regression situation where $Y_i = (y_i, x_i)^\top$ and the x_i are some additional covariates are presented later.

We are interested in testing the hypothesis

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

against the alternative that (at least one component of) θ_i varies over "time". For this alternative to be sensible the ordering assumption is necessary: If a parameter instability with a single breakpoint, say, occurs with respect to a certain ordering of the variables this single breakpoint interpretation would be lost by re-ordering in such a way that observations from the two regimes are mixed. The assumption of independence is assumed for convenience and will be weakened later.

3 Generalized M-fluctuation processes

In this section we suggest a general class of fluctuation processes that can capture instabilities in the coefficient θ. In the first two Sections 3.1 and 3.2 the fluctuation processes are introduced and their behaviour under the null hypothesis (2) is derived. The limiting process is first derived for known θ_0 and then for the case where it has to be estimated. In Section 3.3, the results are generalized to local alternatives and finally in Section 3.4 some further fluctuation processes are introduced.

3.1 Theoretical fluctuation processes

Consider some suitably smooth k-dimensional score function $\psi(\cdot)$, independent of n and i, with

$$\mathrm{E}[\psi(Y_i, \theta_i)] \quad = \quad 0 \tag{3}$$

and define the following matrices

$$A(\theta) \quad = \quad \mathrm{E}[-\psi'(Y, \theta)], \tag{4}$$
$$B(\theta) \quad = \quad \mathrm{COV}[\psi(Y, \theta)], \tag{5}$$
$$C(\theta) \quad = \quad \mathrm{E}[\psi(Y, \theta)u(Y, \theta)^\top] \tag{6}$$

where $Y \sim F(\theta_0)$, $\psi'(\cdot)$ is the first derivative of $\psi(\cdot)$ with respect to θ, and $u(\cdot, \theta)$ is

$$u(y, \theta) \quad = \quad \frac{\partial \log f(y, \theta)}{\partial \theta}, \tag{7}$$

and $f(\cdot, \theta)$ is the probability density function of F. Hence, u is the partial derivative of the log likelihood with respect to θ, also called Maximum Likelihood (ML) score. The first two matrices $A(\theta)$ and $B(\theta)$ are standard in M-estimation, $C(\theta)$ is only needed in Section 3.3. Note that given ψ the matrices $A(\theta)$ and $B(\theta)$ but not $C(\theta)$ can be estimated without further knowledge of F or f respectively.

Theorem 1 *For the cumulative score process given by*

$$W_n(t, \theta) \quad = \quad \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} \psi(Y_i, \theta) \tag{8}$$

and under the assumptions stated above and under H_0 the following functional central limit theorem (FCLT) holds:

$$W_n(\cdot, \theta_0) \quad \xrightarrow{\mathrm{d}} \quad Z(\cdot),$$

where $Z(\cdot)$ is a gaussian process with continuous paths, mean function $\mathrm{E}[Z(t)] = 0$ and covariance function $\mathrm{COV}[Z(t), Z(s)] = \min(t, s) \cdot B(\theta_0)$.

Proof: The proof follows by direct application of Donsker's theorem (Billingsley 1999).

Corollary 1 *If $B(\theta_0)$ is regular, the following FCLT holds for the decorrelated fluctuation process*

$$B(\theta_0)^{-1/2} W_n(\cdot, \theta_0) \quad \xrightarrow{\mathrm{d}} \quad W(\cdot),$$

where $W(\cdot)$ is a k-dimensional Wiener process or standard Brownian motion.

3.2 Empirical fluctuation processes

Usually, in applications the parameter θ_0 under the null hypothesis is not known but has to be estimated. A suitable estimator can be based on the function $\psi(\cdot)$: the full sample M-estimator $\hat{\theta}_n$ is defined by the equation

$$\sum_{i=1}^{n} \psi(Y_i, \hat{\theta}_n) \quad = \quad 0. \tag{9}$$

Some properties of this M-estimator are well known (see e.g., Stefanski and Boos 2002). Taylor expansion of

$$S_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \psi(Y_i, \theta) \tag{10}$$

gives

$$0 \quad = \quad S_n(\hat{\theta}_n) \quad = \quad S_n(\theta_0) + S'_n(\theta_0)(\hat{\theta}_n - \theta_0) + R_n. \tag{11}$$

Under suitable regularity conditions

$$-S'_n(\theta_0) = \frac{1}{n}\sum_{i=1}^{n} -\psi'(Y_i, \theta_0) \quad \overset{\text{p}}{\longrightarrow} \quad A(\theta_0),$$

$$\sqrt{n}S_n(\theta_0) \quad \overset{\text{d}}{\longrightarrow} \quad \mathcal{N}(0, B(\theta_0)),$$

$$\sqrt{n}R_n \quad \overset{\text{p}}{\longrightarrow} \quad 0.$$

Therefore, the following holds

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \quad \overset{\text{d}}{\longrightarrow} \quad \mathcal{N}(0, V(\theta_0)), \tag{12}$$

where $V(\theta) = A(\theta)^{-1}B(\theta)\{A(\theta)^{-1}\}^\top$. See Stefanski and Boos (2002) or White (1994) for further details; White also gives a set of suitable regularity conditions (White 1994, Theorem 6.10, p. 104). Equivalently to (12) we can write

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \quad \overset{.}{=} \quad A(\theta_0)^{-1} \cdot W_n(1, \theta_0), \tag{13}$$

where $a_n \overset{.}{=} b_n$ means that $a_n - b_n$ tends to zero (in probability if a_n or b_n are stochastic).

Theorem 2 *Under H_0 the following FCLT holds for the empirical cumulative score process with M-estimated parameters*

$$W_n(\cdot, \hat{\theta}_n) \quad \overset{\text{d}}{\longrightarrow} \quad Z^0(\cdot),$$

where $Z^0(t) = Z(t) - tZ(1)$.

Proof:

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

Corollary 2 *If $B(\theta_0)$ is regular, the following FCLT holds for the decorrelated empirical fluctuation process with M-estimated parameters*

$$\hat{B}_n^{-1/2}W_n(\cdot, \hat{\theta}_n) \quad \overset{\text{d}}{\longrightarrow} \quad W^0(\cdot),$$

where $W^0(t)$ is a standard k-dimensional Brownian bridge with $W^0(t) = W(t) - tW(1)$ and $\hat{B}_n$ some consistent and regular covariance matrix estimate, e.g.,

$$\hat{B}_n \quad = \quad \frac{1}{n}\sum_{i=1}^{n}\psi(Y_i, \hat{\theta}_n)\psi(Y_i, \hat{\theta}_n)^\top. \tag{14}$$

3.3 Local alternatives

In parameter instability problems or structural change situations an alternative of interest is the local alternative

$$H_A: \theta_i \quad = \quad \theta_0 + \frac{1}{\sqrt{n}}\, g\left(\frac{i}{n}\right), \tag{15}$$

where $g(\cdot)$ is a function of bounded variation on $[0,1]$ which describes the pattern of departure from stability of the parameter θ_0 (Kuan and Hornik 1995; Hjort and Koning 2002).

Then Y_i has the probability density function

$$f(y,\theta_i) \doteq f(y,\theta_0) \left\{ 1 + u(y,\theta_0)^\top \frac{1}{\sqrt{n}} g\left(\frac{i}{n}\right) \right\}, \tag{16}$$

which can be easily derived from first order Taylor expansion of f.

Therefore, under a local alternative like (15) the components of the fluctuation process (8) no longer have zero mean in general but

$$\mathrm{E}[\psi(Y_i,\theta_0)] \quad \doteq \quad \int \psi(y,\theta_0) f(y,\theta_0) \, \mathrm{d}y + \int \psi(y,\theta_0) u(\theta_0)^\top f(y,\theta_0) \frac{1}{\sqrt{n}} g\left(\frac{i}{n}\right) \, \mathrm{d}y \tag{17}$$

$$= \quad 0 + \frac{1}{\sqrt{n}} C(\theta_0) g\left(\frac{i}{n}\right). \tag{18}$$

In fact, with the same arguments as in Section 3.1 and 3.2, the whole fluctuation process can be split into one part which is governed by the FCLT from Theorem 1 and a second part which is determined by the function g from (15):

$$W_n(\cdot,\theta_0) \quad \xrightarrow{\mathrm{d}} \quad Z^A(\cdot), \tag{19}$$

where $Z^A(t) = W(t) + C(\theta_0)G(t)$ and $G(\cdot)$ is the antiderivative of g with $G(t) = \int_0^t g(y) \, \mathrm{d}y$.

Finally, the following limiting process can be derived for the decorrelated empirical fluctuation process:

$$\hat{B}_n^{-1/2} W_n(t,\hat{\theta}_n) \quad \doteq \quad \hat{B}_n^{-1/2} \{W_n(t,\theta_0) - tW_n(1,\theta_0)\} \tag{20}$$

$$\doteq \quad W^0(t) + B(\hat{\theta}_n)^{-1/2} C(\hat{\theta}_n) G^0(t), \tag{21}$$

with $G^0(t) = G(t) - tG(1)$, provided $B(\cdot)$ is consistent under H_A.

The results above include the results from Section 3.1 and 3.2 as special cases because under the null hypothesis of parameter stability (2) the function g is identical to zero $g \equiv 0$. But the results also imply that tests based on the empirical fluctuation processes will be consistent against suitable local alternatives of type (15).

3.4 Further fluctuation processes

Instead of capturing the fluctuation in a cumulative sum of scores, a moving or rolling sum of scores could be used equally well. More formally, we also consider processes of type

$$M_n(t,\theta) \quad = \quad \frac{1}{\sqrt{n}} \sum_{i=\lfloor N_n t \rfloor + 1}^{\lfloor N_n t \rfloor + \lfloor nh \rfloor} \psi(Y_i,\theta), \tag{22}$$

where $N_n = (n - \lfloor nh \rfloor)/(1-h)$ and h determines the bandwidth.

Under H_0 this process converges to the increments of a Brownian motion or bridge respectively, depending on whether the value θ_0 is known or has to be estimated. The latter case is stated formally in the following theorem.

Theorem 3 *Under H_0 the following FCLT holds for the empirical moving score fluctuation process:*

$$\hat{B}_n^{-1/2} M_n(\cdot,\hat{\theta}_n) \quad \xrightarrow{\mathrm{d}} \quad M^0(\cdot), \tag{23}$$

where $M^0(t) = W^0(t+h) - W^0(t)$ is the process of increments of a Brownian bridge.

Proof: The proof follows by application of Lemma A from Chu et al. (1995a) to the results of Section 3.2 and 3.1.

4 Generalized M-fluctuation tests

4.1 Choice of the scores ψ

We choose the M-estimation framework for estimation of the parameters θ as it contains many other estimation techniques as special cases by choosing a suitable score function ψ. Other classes of estimators are not strictly special cases but are strongly related to M-estimation and the principles introduced in Section 3 can be used to construct fluctuation processes with the same asymptotic properties in those situations a few of which are outlined in the following.

One of the most common choices for ψ is to use the partial derivative of some objective function Ψ

$$\psi(y, \theta) \quad = \quad \frac{\partial \Psi(y, \theta)}{\partial \theta}, \tag{24}$$

where Ψ could be the residual sum of squares or the Log-Likelihood, yielding the OLS or ML estimators $\hat{\theta}$ respectively (the dependence of $\hat{\theta}$ on the number of observations n is ignored in the following). In both cases the cumulative sums of the first order conditions ψ lead very naturally to fluctuation processes as described an the previous section.

Another estimating approach which is particularly popular in econometrics is to use a Quasi-Maximum Likelihood (White 1994) in a misspecification context. This again is similar to M-estimation in robust statistics (Huber 1964, 1972) which also accounts for violation of some of the standard model assumptions. Huber (1964) suggests to use the function

$$\psi_H(y, \theta) \quad = \quad \min(c, \max(y - \theta, -c)). \tag{25}$$

with some constant c for robust estimation of the mean of a symmetric distribution. Note that this function is not smooth but it is almost everywhere differentiable (except in $\pm c$) and in the definition of the matrix $A(\theta)$ in (4) integration and differentiation can be interchanged. Hence, almost identical results for estimation and construction of fluctuation processes can be derived using functions like Huber's ψ from (25).

Another approach is not to fully specify a model via its likelihood but to use some estimating equations (Godambe 1960, 1985) which are satisfied by the true model. Similarly, some moment or orthogonality conditions can be exploited to derive estimating functions which again yield parameter estimates. This approach is used in estimation techniques like instrumental variables in linear models (IV, Sargan 1958), the generalized method of moments (GMM, Hansen 1982) for the estimation of economic models or the generalized estimating equations (GEE, Liang and Zeger 1986) for models for longitudinal or time-series data in biostatistics. Further discussion of related estimation approaches can be found in Bera and Bilias (2002). Usually, these are regression models which are not yet covered by the methodology introduced above. However, with some modifications as described in Section 5 fluctuation processes with rather similar properties can be derived.

All these methods have in common that the estimation of θ is based on some score or estimating function ψ or a moment or orthogonality condition similar to (9) whose partial sums yield fluctuation processes satisfying some FCLT which again can be used to construct tests for parameter instability. The latter step will described in detail in the following section.

4.2 Test statistics

We derived the empirical fluctuation processes because they can capture departures from the null hypothesis (2) of parameter stability. Therefore, visual inspection alone conveys information about whether H_0 is violated or not. But this alone is, of course, not enough and we want to derive tests based on empirical fluctuation processes. One common strategy for this is to consider some scalar functional λ that can be applied to the fluctuation processes.

Given a finite sample as in model (1) an empirical fluctuation process is a $n \times k$ array $(efp_j(i/n))_{i,j}$ with $i = 1, \ldots, n$ and $j = 1, \ldots, k$ that converges to a k-dimensional limiting process which is

continuous in time. For the cumulative score process derived in Section 3

$$
efp_j\left(\frac{i}{n}\right) \quad = \quad \frac{1}{\sqrt{n}}\sum_{l=1}^{i}\psi_j(Y_l,\hat{\theta}), \tag{26}
$$

where $\psi_j(\cdot)$ is the jth component of $\psi(\cdot)$.

To aggregate this empirical process to a scalar test statistic several suitable functionals of the form

$$
\lambda\left(efp_j\left(\frac{i}{n}\right)\right). \tag{27}
$$

are conceivable. The limiting distribution for these test statistics can be determined fairly easily, it is just the corresponding (asymptotic) functional applied to the limiting process. But although closed form results for certain functionals of Brownian bridges exist, the critical values are typically easily derived by simulation so that this poses no constraint for the choise of λ.

λ can usually be split into two components: λ_{time} which aggregates over time and λ_{comp} which aggregates over the components of ψ. Common choices for λ_{time} are the absolute maximum, the mean or the range (Kuan and Hornik 1995; Hjort and Koning 2002; Zeileis, Leisch, Hornik, and Kleiber 2002a). Typical functionals λ_{comp} include the maximum norm (or L_∞ norm, denoted as $||\cdot||_\infty$) or the squared Euclidian norm (or L_2 norm, denoted as $||\cdot||_2^2$), see Hjort and Koning (2002) for more examples.

As the decorrelated processes are asymptotically independent it seems to be more intuitive to first aggregate over time and then have k independent univariate test statistics, each associated with one component of the process which can usually be matched with one component of the parameter vector θ. If the overall hypothesis is rejected the component(s) of θ which caused the instability can then be identified.

On the other hand, when there is evidence for a structural change a very natural question is *when* it occured. To focus on this question it is obviously better to first aggregate over j and then inspect the resulting univariate process for excessive fluctuation which can be also done visually, e.g. by checking whether this process crosses some boundary $b(t) = c \cdot d(t)$. In this case c determines the significance level and $d(\cdot)$ the shape of the boundary and the resulting test statistic is

$$
\max_{i=1,\ldots,n}\left|\frac{\lambda_{\text{comp}}\left(efp_j(i/n)\right)}{d(i/n)}\right|, \tag{28}
$$

i.e., a weighted maximum of the absolute values of the process aggregated by λ_{comp}. Natural choices are to weight all observations equally, i.e., $d(t) \equiv 1$, or by the (asymptotic) standard deviation of the fluctuation process, i.e., $d(t) = \sqrt{t(1-t)}$ for the cumulative score process. But other boundaries are also conceivable, see Zeileis (2004) or Zeileis, Leisch, Kleiber, and Hornik (2002c) for a closer discussion.

The only class of test statistics which allows for both identification of the component j as well as the time point i/n of a potential structural instability is when the maximum is used for aggregating over both, time and components, i.e.,

$$
\max_{i=1,\ldots,n}\max_{j=1,\ldots,k}\left|\frac{efp_j(i/n)}{d(i/n)}\right| \tag{29}
$$

where the $efp_j(i/n)$ which cross some absolute critical value c can be regarded as violating the hypothesis of stability (Mazanec and Strasser 2000).

4.3 Special cases

The rich class of generalized M-fluctuation tests introduced in this paper contains various tests for parameter instability or structural change known from the statistics and econometrics literature.

ML scores: Most importantly, the generalized M-fluctuation tests contain the tests of Hjort and Koning (2002) who develop a general class of fluctuation processes based on the ML scores from (7). These yield the ML estimate of θ_0 and $A(\theta_0) = B(\theta_0) = C(\theta_0) = I(\theta_0)$ is the usual Fisher information matrix. Hjort and Koning (2002) illustrate how to construct three types of tests which are all included in the more general framework above. In particular they construct a Cramér-von Mises type test which is the average of the Euclidian norm of the fluctuation process at time t. However, they do not point out that this is the test of Nyblom (1989) which Hansen (1992) generalized to linear regression models. Nyblom (1989) showed that this test is the locally most powerful test against the alternative that the parameters follow a random walk.

Changes in the mean: In the case that the Y_i are (not necessarily normally) distributed with mean μ_0 and variance σ^2 various tests for the constancy of the mean can be shown to be special cases of the approach presented above: if μ_0 is estimated by means of OLS the OLS-based CUSUM test (Ploberger and Krämer 1992) and the recursive estimates test (Ploberger et al. 1989) are both equivalent to the natural test resulting from the ideas described above: derive the partial sum process of the scores from Corollary 2 and reject if the maximum absolute value of the process—which is just the scaled cumultative sum of the OLS residuals—is too large. With boundaries proportional to the standard deviation the resulting test is the alternative OLS-based CUSUM test of Zeileis (2004) and with moving instead of cumulative sums as in (23) the resulting process is equivalent to the OLS-based MOSUM process (Chu et al. 1995a) and the moving estimates process (Chu et al. 1995b). Kuan and Hornik (1995) also consider tests based on the range instead of maximum absolute value for the same processes. If robust M-estimation instead of OLS is used for estimating μ_0 tests like the (non-recursive) robust CUSUM tests of Sen (1984) or Sibbertsen (2000) can be constructed.

The connection between the generalized M-fluctuation tests and the OLS-based CUSUM test and the Nyblom-Hansen test (Cramér-von Mises type test) respectively in a linear regression framework will be described in more detail in the following section.

5 Testing for parameter instability in (generalized) linear regression models

The general framework for constructing fluctuation processes and tests based on M-scores presented in the first sections of this paper is already extremely useful for testing the constancy of model parameters over time; but to explain the generality of the approach without too much technical overhead two assumptions which do not really restrict the generality of the results have been made for convenience. These will be weakened in this Section. First, the observations were assumed to be independent—an assumption which is likely to be violated in time series applications but which can usually be overcome easily as commented on at the end of this section. Second, the probability density function $f(y_i, \theta_i)$ was assumed to describe the full distribution of the Y_i—but if the observations can be split into $Y_i = (y_i, x_i)^\top$ with a response or dependent variable y_i and additional regressors or covariates x_i the usual approach is to model the conditional distribution $f(y_i \mid x_i, \theta_i)$ given the x_i. The common assumption is that the x_i form a weakly dependent process without deterministic or stochastic trends, see Andrews (1993) for technical details or also Hansen (1992) or Hjort and Koning (2002). Under such suitable assumptions the same asymptotic distribution can be derived for the processes based on estimates of the regression coefficients which can be obtained by various procedures as discussed in Section 4.1.

To make the dependence on the covariates obvious the score function ψ from (3) is now written as

$$\psi(Y_i, \theta_i) \quad = \quad \psi(y_i, x_i, \theta_i). \tag{30}$$

Of course, it is still required to have zero expectation (with respect to $f(y_i \mid x_i, \theta_i)$) which is not

difficult to fulfill, more crucial is the assumption that the variances stabilize:

$$\frac{1}{n}\sum_{i=1}^{n}\text{COV}[\psi(y_i, x_i, \theta_0)] = J_n \quad \overset{\text{p}}{\longrightarrow} \quad J, \tag{31}$$

where the matrix J in a regression context corresponds to $B(\theta)$ from (5) in the no-covariate context. This follows for example from the weak dependence assumption stated above.

It is easy to show that functional central limit theorems similar to Theorem 1 and 2 hold for the resulting theoretical and empirical fluctuation processes based on $\psi(y_i, x_i, \theta_i)$, in particular the limiting processes are the same with J instead of $B(\theta_0)$. To derive explicit test statistics for two important classes of regression models with popular estimation techniques the general linear model (LM) und the generalized linear model (GLM) will be treated in more detail in the following.

5.1   The general linear regression model

Consider the general linear regression model

$$y_i = x_i^\top \beta + u_i \qquad (i = 1, \ldots, n), \tag{32}$$

where the disturbances have zero mean and variance σ^2. The precise formulation of these assumptions is not important as long as they imply the same FCLT—for different sets of assumptions see, e.g., Hansen (1992), Andrews (1993), Ploberger and Krämer (1992) or Bai (1997). To simplify notation and to emphasize common properties of the scores in the LM and the GLM the mean of the y_i is sometimes denoted $\mu_i = x_i^\top \beta$.

The model parameters $\theta = (\beta, \sigma^2)^\top$ are usually estimated by OLS or ML (based on a normal model) which leads to the same first order conditions (or scores in a ML framework):

$$\psi(y_i, x_i, \theta) = \left(\psi_\beta(y_i, x_i, \beta), \psi_{\sigma^2}(y_i, x_i, \beta, \sigma^2)\right)^\top, \tag{33}$$

$$\psi_\beta(y_i, x_i, \beta) = x_i(y_i - x_i^\top \beta) = x_i(y_i - \mu_i), \tag{34}$$

$$\psi_{\sigma^2}(y_i, x_i, \beta, \sigma^2) = (y_i - x_i^\top \beta)^2 - \sigma^2. \tag{35}$$

These give the common estimates $\hat\beta$ and $\hat\sigma^2$. Due to the independence of the two estimates the covariance matrix J_n is block diagonal

$$J_n = \sigma^2 \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} x_i x_i^\top & 0 \\ 0 & 2\sigma^2 \end{pmatrix}.$$

Therefore, independent test statistics for the constancy of β and σ respectively can be computed and the stability of the parameters can be assessed independently. Three test statistics will be derived in the following: the Nyblom-Hansen test, the double max test and the OLS-based CUSUM test.

Nyblom-Hansen test: We follow the approach of Hansen (1992) and test both parameters $\theta = (\beta, \sigma^2)^\top$ simultanously. In his equation (9) he gives the formula for the test statistic L_C based on the following empirical fluctuation process and covariance estimate

$$W_n(t, \hat\theta) = \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt \rfloor} \psi(y_i, x_i, \hat\theta), \tag{36}$$

$$\hat{J} = \frac{1}{n}\sum_{i=1}^{n} \psi(y_i, x_i, \hat\theta)\psi(y_i, x_i, \hat\theta)^\top. \tag{37}$$

The test statistic L_C is then given as

$$L_C = \frac{1}{n}\sum_{i=1}^{n} W_n\left(\frac{i}{n}, \hat\theta\right)^\top \hat{J}^{-1} W_n\left(\frac{i}{n}, \hat\theta\right)$$

9

$$= \frac{1}{n} \sum_{i=1}^{n} efp \left(\frac{i}{n} \right)^{\top} efp \left(\frac{i}{n} \right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left\| efp \left(\frac{i}{n} \right) \right\|_2^2,$$

which can be interpreted easily in terms of the considerations in Section 4.2. The asymptotic distribution is $\int_0^1 ||W^0||_2^2$, where W^0 is a k-dimensional Brownian bridge. Nyblom (1989) first suggested this test in a structural change context and showed that it is locally most powerful for the alternative that the parameters follow a random walk. Without relating to Nyblom's earlier work, Hjort and Koning (2002) refer to it as a Cramér-von Mises type test, but they also point out that visual inspection of the process $efp(t)$ might convey information about the timing of a potential structural change. Although enhancing classical significance testing by visual means is a good idea, this approach has the problem that what is tested and what is visualized differ. From the transformation above it becomes clear that it is much more natural to use a plot of $||efp(t)||_2^2$ with two horizontal lines, one for the empirical mean and one for the critical value which visualizes both the significance test and excessive fluctuation (i.e., information about the timing of the shift). An example for this fluctuation process and its visualization is given in Section 6.1.

Double max *test:* As mentioned in Section 4.2 the only test statistic which allows the identification of both the timing of a structural change and the component of the parameter vector θ which changed is

$$\max_{j=1,\ldots,k} \max_{0 \le t \le 1} |efp_j(t)|, \tag{38}$$

which is the same functional for measuring excessive fluctuation as in the recursive estimates test (Ploberger et al. 1989). The limiting distribution is $\max_{j=1,\ldots,k} ||W_j^0(t)||_\infty$. Again, this test can also be performed graphically by plotting each individual process with a horizontal boundary for $\pm$ the critical value.

OLS-based CUSUM test: As already indicated in the previous section, the OLS-based CUSUM test for a change in the mean (without covariates) can be shown to be a special case of the the generalized M-fluctuation test framework. In the regression setup this means that there is just a constant regressor $x_i \equiv 1$ and the variance σ^2 is treated as a nuisance parameter. If we consider the OLS-based CUSUM test for a setup with covariates another interesting interpretation emerges from the M-fluctuation view. If an intercept is included in the regression, the OLS-based CUSUM process is equivalent to the first component of the non-decorrelated processes $W_n(t, \hat{\theta})$ from (36) standardized by $\hat{\sigma}^2$ which is element $(1,1)$ of the estimated covariance matrix $\hat{J}$. Thus, whereas the first component of the decorrelated fluctuation process $\hat{J}^{-1/2}W_n(t, \hat{\theta})$ captures instabilities of the intercept the first component of the non-decorrelated fluctuation process $W_n(t, \hat{\theta})$ captures instabilities in the (expected) mean $E[x]^{\top}\beta$. The latter is proved formally in Ploberger and Krämer (1992). This result implies that shifts orthogonal to the mean regressor $E[x]$ cannot be detected using fluctuation processes based on residuals as such shifts can be interpreted to change the variance rather than the (expected) mean. If the mean L_2 norm rather than the L_∞ norm is used (as in the Nyblom-Hansen test) to measure excessive fluctuation in the OLS-based CUSUM process the resulting test is also trend-resistant, i.e., can deal with trending regressors (Ploberger and Krämer 1996).

5.2 The generalized linear regression model

Now consider the generalized linear model (GLM) like in McCullagh and Nelder (1989). To fix notations, y_i is a response variable distributed independently according to a distribution $F(\theta, \phi)$ where θ is the canonical parameter and ϕ is the dispersion parameter common to all y_i. The probability density has the form

$$f(y_i \mid \theta, \phi) = \exp \left\{ \frac{y_i\theta - q(\theta)}{w(\phi)} + p(y_i, \phi) \right\}, \tag{39}$$

for some known functions $p(\cdot)$, $q(\cdot)$ and $w(\cdot)$, so that $\mathrm{E}[y_i] = \mu_i = q'(\theta)$ and $\mathrm{VAR}[y_i] = w(\phi)q''(\theta) = w(\phi)V(\mu_i)$.

The following relationship is assumed for the relation between covariates and responses:

$$\mu_i = h(\eta_i) = h(x_i^\top \beta_i) \qquad (i = 1, \ldots, n), \tag{40}$$

where $h^{-1}(\cdot)$ is a known link function, β is again the vector of regression coefficients and η_i is the linear predictor.

The regression coefficients β are usually estimated by ML and ϕ is treated as a nuisance parameter (or is known anyway). The resulting score function for β is

$$\psi(y_i, x_i, \beta) = x_i \, h'(x_i^\top \beta) \, V(\mu_i)^{-1}(y_i - \mu_i), \tag{41}$$

where $h'(\cdot)$ is the derivative of the inverse link funtion. The corresponding covariance matrix J_n is given by

$$J_n = \frac{1}{n} \sum_{i=1}^{n} h'(x_i^\top \beta)^2 w(\phi) \, V(\mu_i)^{-1} x_i x_i^\top. \tag{42}$$

In the following, we give explicit formulae for the empirical fluctuation processes in two important special cases of the GLM: the binomial (logistic) regression model and the log-linear poisson model.

Binomial model: Let y_i be the proportion of successes from m trials such that $m y_i$ is binomially distributed $\mathrm{Bin}(\mu_i, m)$. Then the variance is determined by $w(\phi) = 1/m$ and $V(\mu) = \mu(1 - \mu)$ and if the canonical logit link is used then $h'(x) = \exp(x)/(1 + \exp(x))^2$. Given the ML estimates $\hat{\beta}$ and the corresponding fitted values $\hat{\mu}_i$, the covariance matrix for the M-fluctuation process can be estimated by

$$\hat{J} = \frac{1}{nm} \sum_{i=1}^{n} \frac{h'(x_i^\top \hat{\beta})^2}{\hat{\mu}_i(1 - \hat{\mu}_i)} x_i x_i^\top. \tag{43}$$

The empirical fluctuation process is then given by

$$efp(t) = \hat{J}^{-1/2} \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} h'(x_i^\top \hat{\beta}) \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i(1 - \hat{\mu}_i)} x_i. \tag{44}$$

Corresponding test statistics could be derived, e.g., by taking again the average Euclidian norm or the double maximum etc. Note that this methodology can also be applied if $m = 1$ where at each time i there is only one observation of success ($y_i = 1$) or failure ($y_i = 0$). An application of this process in a binomial model can be found in Section 6.2 and 6.3.

Poisson model: If the y_i are poisson distributed $\mathrm{Poi}(\mu_i)$ then $V(\mu) = \mu$ and $w(\phi) = 1$. Using the canonical log link yields $h'(x) = \exp(x)$ so that the covariance can be estimated by

$$\hat{J} = \frac{1}{n} \sum_{i=1}^{n} \hat{\mu}_i x_i x_i^\top. \tag{45}$$

The empirical fluctuation process which is also given in Hjort and Koning (2002) simplifies to

$$efp(t) = \hat{J}^{-1/2} \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} (y_i - \hat{\mu}_i) x_i \tag{46}$$

and test statistics can be computed like above. As pointed out for the OLS-based CUSUM test in Section 5.1 a test for changes in the mean the first component of the non-decorrelated process can also be used alone which has to be standardized by element (1,1) of the covariance matrix alone. For the poisson model this gives

$$efp(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}}}, \tag{47}$$

11

where $\bar{\mu}$ is the arithmetic mean of the fitted values $\hat{\mu}_i$. If the variances are constant, i.e., if there is only a constant regressor $x_i \equiv 1$ this is a CUSUM process based on the poisson residuals. If not, it is almost a Pearson residuals CUSUM process only that the variance is estimated by $\bar{\mu}$ rather than $\hat{\mu}_i$. The latter is not possible if $\hat{\mu}_i$ is not consistent for the asymptotic variance, that is, the variance has to be estimated from a set of observations and the size of this set has to go to infinity with n. This Pearson residual-based CUSUM process is applied to a Poisson model in Section 6.4.

With a simple modification both processes can not only be used in poisson but also in quasi poisson models where overdispersion is allowed. Then, the variance is not required to be equal to the mean but can be $\text{VAR}[y_i] = \phi\mu_i$. Note that in this case the density function is not given by (39) with $w(\phi) = \phi$. The dispersion parameter is a nuisance parameter and can be consistently estimated by $X^2/(n-k)$, where X^2 is the usual Pearson χ^2 statistic. To obtain properly standardized fluctuation processes $efp(t)$ from (46) or (47) respectively has to be multiplied with $1/\sqrt{\hat{\phi}}$ which can then be used as usual for testing the constancy of the regression coefficients β.

5.3 Dependent data

As stated above, the assumption of independent observations is often (but not necessarily) violated, in particular when dealing with time series data. Several approaches are conceivable when the methodology introduced above should be applied to dependent data.

When using ML estimation techniques the parameters can be estimated from a fully specified likelihood or from a conditional likelihood and the fluctuation processes can be derived accordingly. But in many situations this is not necessary as consistent estimates $\hat{\theta}$ (or $\hat{\beta}$ in regression frameworks) can be obtained from the usual estimating equations (Godambe 1985; Liang and Zeger 1986). But as Lumley and Heagerty (1999) point out it is crucial for inference in such models to compute consistent estimates for the covariance matrix $\hat{B}$ (or $\hat{J}$ respectively in regression models). Lumley and Heagerty (1999) suggest a class of weighted empirical adaptive variance estimators which are consistent in the presence of correlation in the data. These can be plugged into the fluctuation processes described above which renders the asymptotic theory valid again.

6 Applications

We illustrate a few of the tests discussed above by applying them to the following four models: an error correction model (ECM) for German M1 money demand, a binomial GLM for the fraction of illegitimate births in Großarl and for simulated binary data and a Poisson model for the number of youth homicides in Boston. The three "real world" data sets are included in the package strucchange (Zeileis, Leisch, Hornik, and Kleiber 2002b) implemented in the R system for statistical computing (http://www.R-project.org/) and available from the Comprehensive R Archive Network http://CRAN.R-project.org/.

6.1 German M1 money demand

Lütkepohl, Teräsvirta, and Wolters (1999) investigated the stability and linearity of a German M1 money demand function and found a stable relationship for the time before the German monetary unification on 1990-06-01 but a clear structural change afterwards. They used seasonally unadjusted quarterly data from 1961(1) to 1995(4) 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 data were originally provided by the German central bank and are now available on the World Wide Web 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) used smooth transition regression to model the parameter instability; Zeileis et al. (2002c) discuss this model in a structural change framework, but only based on OLS

residuals and estimates not based on M-scores. We use the adapted model of Zeileis et al. (2002c) for the German M1 money demand to test the stability of the full sample estimates. Figure 1 shows the L_2 norm of the score-based fluctuation process defined by (36) and (37) as discussed in Section 5.1. The dashed horizontal line represents the mean L_2 norm $||efp(t)||_2^2$, i.e., the test statistic of the Nyblom-Hansen test, which exceeds its 5% critical value (solid line). Additionally to the information that the test finds evidence for structural change in the data, the clear peak in the fluctuation process conveys the information that the break seems to have occured in about 1990, corresponding to the German monetary unification (highlighted by the dotted vertical line). The corresponding p value is 0.022.
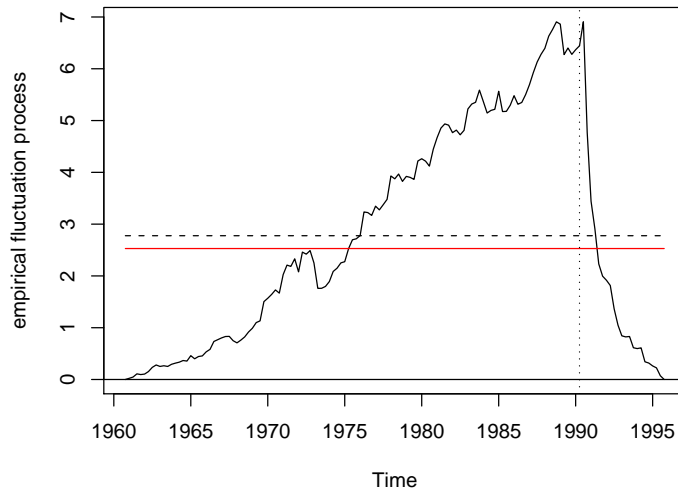
Figure 1: Score-based fluctuation process (mean L_2 norm) for German M1 data

6.2 Illegitimate births in Großarl

In 18th century Salzburg, Austria, the reproductive behaviour was confined to marital unions due to sanctions by the catholic church and the legal system. Nevertheless, illegitimate births conceived outside wedlock happened although the catholic church tried to prevent those by moral regulations of increasing severity. Veichtlbauer, Hanser, Zeileis, and Leisch (2002) discuss the impact of these and other policy interventions on the population system in Großarl, a small village in the Austrian Alps in the region of the archbishopric Salzburg. Zeileis and Veichtlbauer (2002) model the structural breaks in the annual fraction of illegitimate births (see Figure 2) by means of OLS.

Here, we discuss the data on the number of illegitimate and legitimate births between 1700 and 1800 in a binomial regression framework which is more appropriate for this kind of data (although the fitted values are equivalent for a regression on a constant). There were about 55 births per year in Großarl during the 18th century—about seven of which were illegitimate—with an IQR of (48, 63). During this time the close linkage between religiosity and morality and between church and state led to a policy of moral suasion and social disciplining, especially concerning unwanted forms of sexuality. Moral regulations aimed explicitly at avoiding such unwanted forms of sexuality, e.g., by punishing fornication by stigmatising corrections, corporal punishment, compulsory labour or workhouse-prison. Women sometimes even had to leave the court district afterwards due to repetition danger. After secularisation such regulations were abolished in the 19th century. To

assess whether such interventions have any effect on the mean fraction of illegitimate births we employ the CUSUM test based on ML-scores from a binomial model as defined in Equation (44). The resulting empirical fluctuation process in Figure 2 clearly exceeds its boundary and therefore provides evidence for a decrease in the fraction of illegitimate births suggesting that the moral regulations have been efficient. The peak in the process conveys that there has been at least some structural break at about 1750, but two minor peaks on the left and the right can also be seen in the process. These match very well with the three major moral interventions in 1736, 1753 and 1771 respectively (indicated by dotted lines). The corresponding p value is < 0.0001.
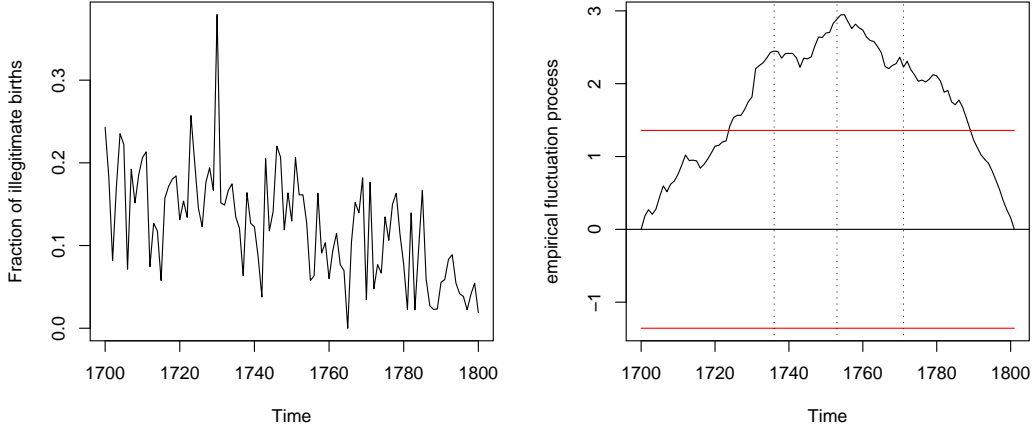
Figure 2: Illegitimate births in Großarl and the binomial CUSUM process

6.3 Artificial binary data

To show that the approach of M-fluctuation processes is not only applicable in situations like above where also OLS estimation techniques could be used despite a binomial GLM being more appriopriate, we analyse an artifical data set of binary observations with covariates. We simulate $n = 200$ observations from a binomial GLM as described in Section 5.2 with the canonical logit link. At each time i the response variable is only $m = 1$ observation of success ($y_i = 1$) or failure ($y_i = 0$) and the vector of covariates is $x_i = \left(1, (-1)^i\right)^\top$. A single shift model with changepoint $t = 0.5$ is used, i.e., 100 observations in each segment, and the vector of regression coefficients in segment 1 is $\beta^{(1)} = (1, 1)^\top$ which changes to $\beta^{(2)} = (0.2, 1)^\top$ in segment 2. Thus, the model corresponds to alternating success probability $\mu_i = 0.5$ and 0.881 in the first segment which drop to alternating success probabilities of 0.31 and 0.769 in the second segment. As only the first regression coefficient but not the remaining one changes this type of alternative is also called partial structural change. Unlike the previous example, the inspection of the raw time series data in Figure 3 does not shed much light on whether or not there has been a change in the parameters of the underlying model.

However, if the empricial fluctuation process from Equation (44) is derived the structural instability can clearly be seen as in Figure 4. It depicts the 2-dimensional fluctuation process with the boundary and process for the intercept in the upper panel and the process for the covariate $(-1)^i$ in the lower panel. This corresponds to using the double max statistic from (38) which allows for both identification of the instable parameter and the timing of the shift as discussed in Section 4.2. As only the first process crosses its boundary the test is able to pick up that the partial break that is only associated with the intercept, while the moderate fluctuation of the second process reflects
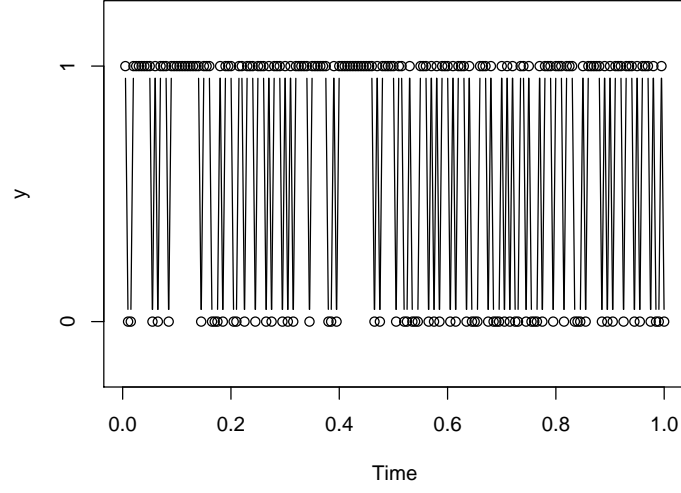
14

Figure 3: Artificial binary data

that the corresponding regression coefficient remains constant. The peak in the middle of sample period matches the true breakpoint of $t = 0.5$ (dotted line) very well. In addition, the fluctuation processes in Figure 4 illustrate that although the response variable is just binary the functional limit theorem works very well. The p value corresponding to the double max test is 0.029.
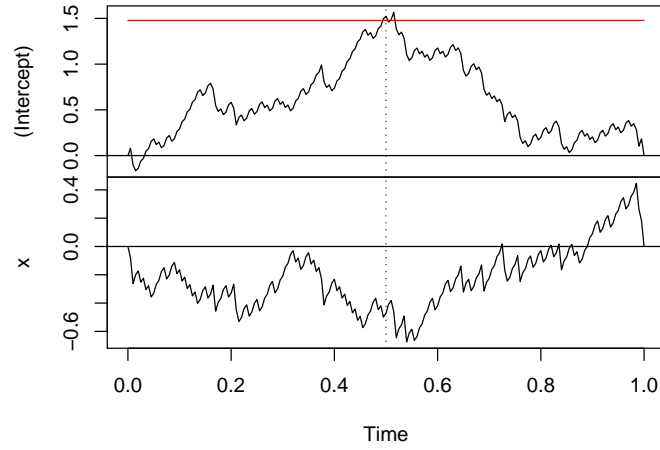

Figure 4: 2-dimensional binomial CUSUM process

6.4 Boston homicide data

To address the problem of continuing high homicide rates in Boston, in particular among young people, a policing initiative called the "Boston Gun Project" was launched in early 1995. This project implemented what became known as the "Operation Ceasefire" intervention in the late spring of 1996 which aimed at lowering homicide rates by a deterrance strategy. More information about youth homicide in Boston can be found in Kennedy, Piehl, and Braga (1996). As a single shift alternative seems reasonable but the precise start of the intervention cannot be determined Cooper, Piehl, Braga, and Kennedy (2001) chose to model the number of youth homicides per month in Boston (see Figure 5) using modifications of the F tests for structural change of Andrews (1993) and Andrews and Ploberger (1994) assessing the significance via Monte Carlo results. In their regression model they include control variables like the population or a factor coding the month, but both have no significant influence at a 10% level. Hence, we use a much simpler model with a straightforward corresponding test: as is natural for count data we model the mean of the number of homicides by a Poisson model and assess the stability of the mean using the CUSUM process of the Pearson residuals as defined in Equation (47).

The corresponding empirical fluctuation process can be seen on the right in Figure 5. As the process crosses its boundary (i.e., the L_∞ norm is used) there is evidence for a decrease of the number of homicides. Furthermore, the peak in the process indicates that the change seems to have occured around early 1996 when the Operation Ceasefire was implemented (dotted line). The corresponding p value is < 0.0001.
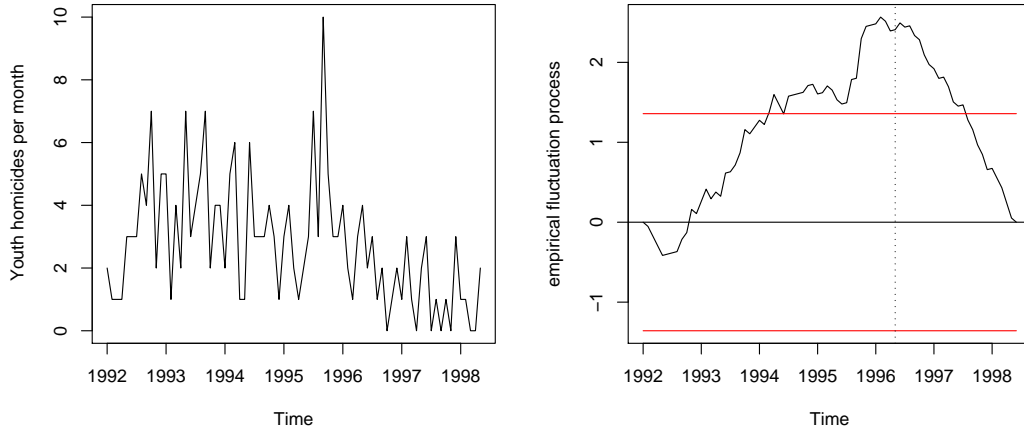

Figure 5: The Boston homicide data and the Poisson CUSUM process

7 Conclusions

In this paper, we propose a general class of tests for parameter instability based on partial sum flucutation processes of M-estimation scores. Based on functional central limit theorems for these flucutation processes both under the hypothesis and local alternatives it is shown how structural changes in parametric models, with a special emphasis to regression models, can be discovered by test statistics that capture the fluctuation in the M-score processes.

The strategy derived can be summarized as follows: Given a data set, choose a model which should be tested for parameter stability or in which structural changes should be revealed, and choose

an estimation technique sensible for the data under consideration. From the estimation technique the choice of the scores ψ follows naturally yielding an empirical fluctuation process. To check for excessive fluctuation in this process a functional has to be chosen which brings out either the timing of a structural change or the component incorporating the instability or both. For many such functionals the resulting significance test can be enhanced by a visualization method that not only displays the result of the test procedure but also conveys information about the type of structural instability and thus allows for better understanding of the structure of the data.

Acknowledgements

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.

Bai, J. (1997), "Estimating Multiple Breaks One at a Time," *Econometric Theory*, 13, 315–352.

Bera, A. K. and Bilias, Y. (2002), "The MM, ME, ML, EL, EF and GMM Approaches to Estimation: a Synthesis," *Journal of Econometrics*, 107, 51–86.

Billingsley, P. (1999), *Convergence of Probability Measures*, New York: Wiley, 2nd ed.

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., Hornik, K., and Kuan, C.-M. (1995a), "MOSUM Tests for Parameter Constancy," *Biometrika*, 82, 603–617.

— (1995b), "The Moving-Estimates Test for Parameter Stability," *Econometric Theory*, 11, 669–720.

Cooper, S. J., Piehl, A. M., Braga, A. A., and Kennedy, D. M. (2001), "Testing for Structural Breaks in the Evaluation of Programs," Unpublished Paper, John F. Kennedy School of Government.

Godambe, V. P. (1960), "An Optimum Property of Regular Maximum Likelihood Estimation," *Annals of Mathematical Statistics*, 31, 1208–1211.

— (1985), "The Foundations of Finite-Sample Inference in Stochastic Processes," *Biometrika*, 48, 419–428.

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

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

Hansen, L. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

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

Huber, P. J. (1964), "Robust Estimation of a Location Parameter," *Annals of Mathematical Statistics*, 35, 73–101.

— (1972), "Robust Statistics: A Review," *Annals of Mathematical Statistics*, 43, 1041–1067.

Kennedy, D. M., Piehl, A. M., and Braga, A. A. (1996), "Youth Violence in Boston: Gun Markets, Serious Youth Offenders, and a Use-Reduction Strategy," *Law and Contemporary Problems*, 59, 147–183.

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

Liang, K.-Y. and Zeger, S. L. (1986), "Longitudinal Data Analysis Using Generalized Linear Models," *Biometrika*, 73, 13–22.

Lumley, T. and Heagerty, P. (1999), "Weighted Empirical Adaptive Variance Estimators for Correlated Data Regression," *Journal of the Royal Statistical Society*, B 61, 459–477.

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.

Mazanec, J. A. and Strasser, H. (2000), *A Nonparametric Approach to Perceptions-Based Market Segmentation: Foundations*, Berlin: Springer.

McCullagh, P. and Nelder, J. A. (1989), *Generalized Linear Models*, London: Chapman & Hall, 2nd ed.

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

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

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

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.

Sargan, J. D. (1958), "The Estimation of Economic Relationships Using Instrumental Variables," *Econometrica*, 26, 393–415.

Sen, P. K. (1984), "Recursive M-Tests for the Constancy of Multivariate Regression Relationships Over Time," *Sequential Analysis*, 3, 191–211.

Sibbertsen, P. (2000), "Robust CUSUM-M Test in the Presence of Long-Memory Disturbances," Technical Report 28, SFB 475 "Reduction of Complexity for Multivariate Data Structures".

Stefanski, L. A. and Boos, D. D. (2002), "The Calculus of M-Estimation," *American Statistician*, 56, 29–38.

Tukey, J. W. (1962), "The Future of Data Analysis," *Annals of Mathematical Statistics*, 33, 1–67.

Veichtlbauer, O., Hanser, E., Zeileis, A., and Leisch, F. (2002), "The Impact of Policy Interventions on a Pre-industrial Population System in the Austrian Alps," Forthcoming.

White, H. (1994), *Estimation, Inference and Specification Analysis*, Cambridge University Press.

Zeileis, A. (2004), "Alternative Boundaries for CUSUM Tests," *Statistical Papers*, forthcoming.

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

— (2002b), "`strucchange`: An R Package for Testing for Structural Change in Linear Regression Models," *Journal of Statistical Software*, 7, 1–38.

Zeileis, A., Leisch, F., Kleiber, C., and Hornik, K. (2002c), "Monitoring Structural Change in Dynamic Econometric Models," Report 64, SFB "Adaptive Information Systems and Modelling in Economics and Management Science".

Zeileis, A. and Veichtlbauer, O. (2002), "Policy Interventions Affecting Illegitimacy in Preindustrial Austria: A Structural Change Analysis," in *Festschrift 50 Jahre Österreichische Statistische Gesellschaft*, ed. Dutter, R., Österreichische Statistische Gesellschaft, pp. 133–146, ISSN 1026-597X.