Tests for Constancy of Model Parameters Over Time

Nils Lid Hjort and Alexander Koning

University of Oslo and Erasmus Universiteit Rotterdam

Abstract. Suppose that a sequence of data points follows a distribution of a certain parametric form, but that one or more of the underlying parameters may change over time. This paper addresses various natural questions in such a framework. We construct canonical monitoring processes which under the hypothesis of no change converge in distribution to independent Brownian bridges, and use these to construct natural goodness-of-fit statistics. Weighted versions of these are also studied, and optimal weight functions are derived to give maximum local power against alternatives of interest. We also discuss how our results can be used to pinpoint where and what type of changes have occurred, in the event that initial screening tests indicate that such exist. Our unified large-sample methodology is quite general and applies to all regular parametric models, including regression, Markov chain and time series situations.

Key words: *Brownian bridges; change points; constancy of parameters; goodness of fit testing; parameter discontinuities*

1. Introduction and summary

Do the parameters of a statistical model stay constant, or do they experience changes over time? What are the best goodness-of-fit tests for the 'no change' hypothesis? What is necessary in order to claim that changes have occurred? If there are level shifts or other types of discontinuity, how can one spot them, or best describe their nature, or pinpoint their locations?

This paper is concerned with these general questions, and aims at devising generally applicable principles and methods. The basic theory is developed first in Sections 2–4, for the structurally and conceptually simplest general case, namely that of independent data with no extra covariate information. Here $Y_1, Y_2, \ldots$ are independent with densities of common form $f(y,\theta)$, but the parameter θ is not necessarily constant as time goes by. In Section 2 a certain p-dimensional monitoring process is constructed, p being the dimension of the θ parameter, behaving in the large-sample limit as p independent Brownian bridges. This makes it easy to construct various overall tests for the hypothesis of no change in the θs, having in mind as interesting alternatives those where the parameter changes over time. In Section 3 weighted versions of these processes are constructed. Section 4 provides results about local detection power, against various discontinuity alternatives of interest, and about optimal weight functions. To some extent we also learn about how to detect where changes have occurred, if indeed the screening tests indicate that such are present.

There would often be situations where covariate information is available for each Y_i, and where questions related to parameter constancy or change would be important. As an example, suppose Y_i is Poisson with mean parameter $\exp(a + b_1 x_{i,1} + b_2 x_{i,2})$, reflecting dependence on factors x_1 and x_2. Then perhaps b_2 changes over time, reflecting say increased dependence on factor x_2. The general regression framework is discussed in Section 5. Section 6 provides illustrations of our methods.

The scope of our methodology is broader than models with independence. Tests for parameter constancy, and results about these, may be derived also in more complex situations, like in Markov chain models, where the transition probabilities may have changed over time, and in time series regimes, where for example serial dependence parameters may not have been constant over time. This is explained in Section 7, along with other remarks and pointers to problems for further work.

There are several areas of applied statistics where questions and problems arise for which the methods of this paper would be applicable. One quite general such area is that of prediction. This is of central importance in econometrics, for example; see e.g. Ploberger and Krämer (1992). Ongoing debates and controversies concerning climate changes also involve prediction issues. Predictions rely heavily on the assumption that the future behaves in much the same way as in the past. Structural breaks may destroy the reliability of predictions. The investigation of the Dutch Ombudsman, described in section 6, and where our methods do discover a structural break in underlying parameters, was in fact initiated because of the poor prediction of capacity needed for so-called TBS-treatments.

Another general such area is statistical process control. Statistical process control methods aim at detecting non-constancy of parameters in the context of industrial statistics. Our methods are of relevance for analysing historical data (corresponding to what is sometimes referred to as 'stage I statistical process control' problems), see Sullivan and Woodall (1996) and Koning and Does (1997); and also to some extent for monitoring real-time data.

And a third general area of application would be that of stochastic simulation via processes that supposedly converge towards equilibrium distributions. Some simulation systems need a 'warming-up' period to reach stationarity. Some special cases of tests presented in this paper are in fact already used to investigate whether the system has warmed up sufficiently; see e.g. Schruben (1982, 1983) and Ripley (1987, Ch. 6).

It is worth remarking that the clear majority of articles dealing with goodness of fit problems for parametric models is concerned with a more 'static' problem formulation; one believes that a sample comes from a definite distribution and tests whether this distribution is of a specified type. The present formulation is 'dynamic' and focusses specifically on discontinuities over time. This helps explain why our large-sample theory leads to results that are both more unified and more simple than those obtained in the 'static' framework. Thus an infamous comment of Pollard (1984, p. 118), stating that "The interest aroused when Durbin (1973) applied weak convergence methods to get limit distributions for statistics analogous to those of Kolmogorov and Smirnov, but with estimated parameters, died down when the intractable limit processes asserted themselves," does not concern us.

2. Canonical monitoring processes

The framework for this section involves a sequence of independent observations Y_i, coming all from the same parametric family $f(y,\theta)$, where however the underlying parameters θ_i, all belonging to some open p-dimensional parameter region, may have changed over time. After having observed $Y_1,\ldots,Y_n$, we take particular interest in the hypothesis

$$H_0\colon \theta_1=\cdots=\theta_n, \tag{2.1}$$

which is to be tested against 'discontinuity alternatives'. We assume standard regularity conditions hold for the $f(y,\theta)$ family, sufficient to make the traditional maximum likelihood apparatus work.

2.1. Cumulative score processes. Let $u(y,\theta)$ and $i(y,\theta)$ be first and second derivatives of $\log f(y,\theta)$ w.r.t. θ. To learn about possible evidence against H_0, start out considering cumulative sums of $u(Y_i,\theta_0)$, where θ_0 is the common parameter value under H_0. These have mean zero and variance matrix $J=-\mathrm{E}i(Y,\theta_0)$, the information matrix of the model. By the Donsker theorem, see e.g. Billingsley (1968), combined with the Cramér–Wold device, it is not difficult to derive the result

$$\psi_n(t,\theta_0)=\frac{1}{\sqrt{n}}\sum_{i\le[nt]}u(Y_i,\theta_0)\to_d Z_0(t)\quad\text{in } D_p[0,1], \tag{2.2}$$

where Z_0 is zero-mean Gaußian with covariance function $\min(t_1,t_2)\,J$. The convergence takes place w.r.t. the Skorohod topology in the space $D_p[0,1]$ of right-continuous functions $x\colon[0,1]\to \mathbb{R}^p$ with left-hand limits. Note that Z_0 is a linear transformation of p independent Brownian motions.

Our main concern will be with the case of unknown parameters in the model. But it is worth pointing out that in the fully specified case, where H_0 states that all θ_is are equal to a specified θ_0, the component processes of $J^{-1/2}\psi_n(t,\theta_0)$ tend to p independent Brownian motions under H_0. This makes it particularly easy to construct and analyse test statistics.

2.2. Estimated cumulative score processes. When θ_0 is unknown, let $\widehat\theta$ be the maximum likelihood estimator, and consider the estimated cumulative score process:

$$\psi_n(t,\widehat\theta)=\frac{1}{\sqrt{n}}\sum_{i\le[nt]}u(Y_i,\widehat\theta)\quad\text{for } 0\le t\le 1.$$

Notice that this process both starts and ends at zero. We now use Taylor expansion in conjunction with well-known results about the sampling behaviour of $\widehat\theta$, e.g. $\sqrt{n}(\widehat\theta-\theta_0) \doteq J^{-1}\psi_n(1,\theta_0)$, where $A_n \doteq B_n$ means that $A_n - B_n$ tends to zero in probability. With $u(Y_i,\widehat\theta) \doteq u(Y_i,\theta_0) + i(Y_i,\theta_0)(\widehat\theta-\theta_0)$ this leads to

$$\begin{aligned}\psi_n(t,\widehat\theta) &\doteq \psi_n(t,\theta_0) + \frac{1}{n}\sum_{i\le[nt]} i(Y_i,\theta_0)\,\sqrt{n}(\widehat\theta-\theta_0)\\ &\doteq \psi_n(t,\theta_0) - tJ_{[nt]}J^{-1}\psi_n(1,\theta_0) \to_d Z(t) = Z_0(t) - tZ_0(1),\end{aligned}$$

where $J_n = -n^{-1}\sum_{i=1}^n i(Y_i,\theta_0)$ is consistent for J. The limit process Z is a p-dimensional process with covariance function $t_1(1-t_2)\,J$ for $t_1 \le t_2$, in other words a linear transformation of p independent Brownian bridges.

These results lead naturally to the construction of *the canonical monitoring process*,

$$M_n(t) = \widehat J^{-1/2}\psi_n(t,\widehat\theta) = \widehat J^{-1/2}\frac{1}{\sqrt{n}}\sum_{i\le[nt]} u(Y_i,\widehat\theta) \quad \text{for } 0 \le t \le 1, \tag{2.3}$$

where $\widehat J$ is any reasonable estimate of J (in this connection, see also Remark 7.1). The immediate and quite powerful result is then that

$$M_n \to_d M = J^{-1/2}Z = (W_1^0,\ldots,W_p^0)^{\rm t} \quad \text{under } H_0, \tag{2.4}$$

a vector with p independent Brownian bridges as component processes.

The inverse square root matrix is calculated as usual, via eigen-analysis; there is an orthonormal P and a diagonal D such that $P\widehat JP^{\rm t} = D$, containing eigenvalues of $\widehat J$ in decreasing order, and one puts $\widehat J^{-1/2} = P^{\rm t}D^{-1/2}P$. Another option is the so-called LU-root.

2.3. *Omnibus tests for constancy of parameters over time.* Result (2.4) can easily be utilised for testing purposes, as we demonstrate below. It is also quite useful to monitor the component processes $M_{n,j}(t)$ graphically, particularly in cases where constancy has been rejected; such plots would help in trying to pinpoint in which way or ways H_0 does not hold. See Section 4.3 and the examples of Section 6.

TEST 1: Classes of chi squared type tests can be developed as follows. Divide $[0,1]$ into m windows $I_1,\ldots,I_m$. For component j, consider increments

$$\Delta M_{n,j}(I_k) = \frac{1}{\sqrt{n}}\sum_{i/n\in I_k} (\widehat J^{-1/2})_{(j)}\,u(Y_i,\widehat\theta).$$

These tend to $(\Delta W_j^0(I_1),\ldots,\Delta W_j^0(I_m))^{\rm t}$. Inverting the covariance matrix one finds that

$$A_{n,j}^2 = \sum_{k=1}^m \frac{\{\Delta M_{n,j}(I_k)\}^2}{|I_k|} \to_d \chi^2_{m-1} \quad \text{under } H_0,$$

where $|I_k|$ is the length of interval I_k. These are component test statistics of separate use and interest. They may also be combined to form one overall test, via

$$A_n^2 = \sum_{j=1}^{p} A_{n,j}^2 \to_d \chi^2_{p(m-1)} \quad \text{under } H_0.$$

Test 2: A p-dimensional Kolmogorov–Smirnov type test would be $U_n = \max_{0\le t\le 1} \|M_n(t)\|^2$, which can be written

$$\max_{0\le t\le 1} \psi_n(t,\widehat\theta)^{\rm t}\widehat J^{-1}\psi_n(t,\widehat\theta) = \frac{1}{n}\max_{1\le j\le n-1}\Bigl(\sum_{i\le j} u(Y_i,\widehat\theta)\Bigr)^{\rm t}\widehat J^{-1}\Bigl(\sum_{i\le j} u(Y_i,\widehat\theta)\Bigr).$$

Its limit distribution under H_0 is that of

$$\max_{0\le t\le 1}\|W^0(t)\|^2 = \max_{0\le t\le 1}\Bigl\{\sum_{i=1}^{p} W_i^0(t)^2\Bigr\}^{1/2}.$$

Let us also point to a sum-of-Kolmogorov–Smirnov type tests option:

$$\begin{aligned} U_n' &= \max_{0\le t\le 1}|M_{n,1}(t)| + \cdots + \max_{0\le t\le 1}|M_{n,p}(t)| \\ &= \sum_{j=1}^{p}\frac{1}{\sqrt{n}}\max_{1\le l\le n-1}\Bigl|\sum_{i\le l}(\widehat J^{-1/2})^{\rm t}_{(j)}u(y_i,\widehat\theta)\Bigr|, \end{aligned}$$

with limiting null distribution $\sum_{j=1}^{p}\max_{0\le t\le 1}|W_j^0(t)|$. The upper 0.05 quantile of the distribution of any one these components is 1.358, for example; the distribution of a sum of two or more such components can be found via simulation.

We also mention the natural option of weighing by inverse standard deviation. The point is that

$$T_{n,j} = \max_{\varepsilon\le t\le 1-\varepsilon}\frac{|M_{n,j}(t)|}{\{t(1-t)\}^{1/2}} \to_d \max_{\varepsilon\le t\le 1-\varepsilon}\frac{|W_j^0(t)|}{\{t(1-t)\}^{1/2}} \quad \text{under } H_0,$$

for each component j. This distribution can be simulated or approximated, cf. Miller and Siegmund (1982). Upper 0.10 and 0.05 quantiles are approximately 2.89 and 3.15, for instance, for the case of $\varepsilon = 0.05$. The $T_{n,j}$ would be calculated as the maximum over all right- and left-hand limits at points $t = k/n$ for which $\varepsilon \le k/n \le 1-\varepsilon$.

Test 3: As a final example of a general construction, consider this p-dimensional Cramér–von Mises type test:

$$C_n^2 = \int_0^1 \|M_n(t)\|^2\,{\rm d}t = \frac{1}{n^2}\sum_{j=1}^{n-1}\Bigl(\sum_{i\le j} u(Y_i,\widehat\theta)\Bigr)^{\rm t}\widehat J^{-1}\Bigl(\sum_{i\le j} u(Y_i,\widehat\theta)\Bigr).$$

Under H_0,

$$C_n^2 \to_d \sum_{j=1}^{p} \int_0^1 W_j^0(t)^2 \,\mathrm{d}t = \sum_{k=1}^{\infty} \frac{1}{\pi^2 k^2} \chi^2_{k,p}.$$

Similarly, an Anderson–Darling type weighted version of this can easily be put up.

2.4. Examples. The apparatus above, with monitoring processes and test criteria built on these, can be routinely applied to any regular parametric model.

EXAMPLE 1: Assume the Y_is come from a normal (μ, σ^2) distribution, where one at the outset could be interested in monitoring both parameters for possible changes. Here $u(y,\theta) = \sigma^{-1}(z, z^2 - 1)$, where $z = (y-\mu)/\sigma$, and one quickly finds $J = \sigma^{-2}\mathrm{diag}(1,2)$. This leads to

$$M_n(t) = \frac{1}{\sqrt{n}} \sum_{i \le [nt]} \begin{pmatrix} Z_i \\ 2^{-1/2}(Z_i^2 - 1) \end{pmatrix}, \quad \text{where } Z_i = (Y_i - \widehat{\mu})/\widehat{\sigma}.$$

Tests for the constancy of μ or of σ or both can be constructed based on the two component processes, as per the methods above. See also Remark 7.1.

The first component process coincides with the standardised time series used in Schruben (1982, 1983), and is also a special case of the least squares cumulative sum method in Ploberger and Krämer (1992).

EXAMPLE 2: Let the Y_is come from a Gamma distribution with parameters (a,b), i.e. with density $\{b^a/\Gamma(a)\}\, y^{a-1} \exp(-by)$. Here one finds

$$M_n(t) = \begin{pmatrix} \psi'(\widehat{a}) & -1/\widehat{b} \\ -1/\widehat{b} & \widehat{a}/\widehat{b}^2 \end{pmatrix}^{-1/2} \frac{1}{\sqrt{n}} \sum_{i \le [nt]} \begin{pmatrix} \log Y_i - \psi(\widehat{a}) + \log \widehat{b} \\ \widehat{a}/\widehat{b} - Y_i \end{pmatrix},$$

in terms of maximum likelihood estimators $(\widehat{a}, \widehat{b})$. The two component processes are again approximately independent, and contribute combined information about the mean level of $\log Y_i$ and the mean level of Y_i. One may also construct monitoring processes focussing on a separately or b separately, see the Remark below.

EXAMPLE 3: Now take the Y_is to be Poisson with parameters μ_i. The natural process to monitor these becomes $M_n(t) = n^{-1/2} \sum_{i \le [nt]} (Y_i - \bar{Y})/\bar{Y}^{1/2}$. In the limit this is a Brownian bridge.

EXAMPLE 4: Suppose a die is thrown many times. Assume that its face probabilities $(p_1, \ldots, p_6)$ are unknown, and imagine that they somehow may have changed over time. To monitor this, let the data be registered via $Y_i = (Y_{i,1}, \ldots, Y_{i,6})$, with a 1 for the face showing and 0 for the others. The recipe above gives

$$M_n(t) = \widehat{J}^{-1/2} \frac{1}{\sqrt{n}} \sum_{i \le [nt]} \begin{pmatrix} Y_{i,1}/\widehat{p}_1 - Y_{i,6}/\widehat{p}_6 \\ \vdots \\ Y_{i,5}/\widehat{p}_5 - Y_{i,6}/\widehat{p}_6 \end{pmatrix},$$

where $\widehat{p}_j = n^{-1}\sum_{i=1}^n Y_{i,j}$ and where the 5×5 matrix J^{-1} has $p_j(1-p_j)$ along its diagonal and $-p_jp_k$ outside. The five component processes are approximately independent Brownian bridges, in the case that the probabilities have been constant. Again various test statistics can be written down as per Section 2.3.

Note that this example is relevant for the problem of checking whether a probability density has changed over time, via the monitoring of histograms. More sophisticated methods for nonparametric monitoring for changes of probability densities are given in Hjort and Koning (1999a).

Example 5: Let pairs (X_i, Y_i) be independent and binormally distributed, with parameters say $\mu_1, \sigma_1, \mu_2, \sigma_2, \rho$. One may now construct a five-dimensional monitoring process $M_n(t)$ whose limit distribution, under the hypothesis of no change in the parameters, corresponds to five independent Brownian bridges; we omit the algebraic details here. One may then single out for example the fifth of these, to look for possible changes in the ρ parameter.

Remark: So far we have discussed monitoring processes in the context of parametric models. One may also construct similar methods to monitor statistical parameters more generally, for example, nonparametrically checking the constancy of the skewness parameter for an observed sequence. Suppose α_i is such a parameter of interest, connected to the distribution of Y_i, and that the hypothesis $\alpha_1 = \cdots = \alpha_n$ is to be checked. Assume there is an estimator $\widehat{\alpha}_j$ for the common α value, depending on the first j of Y_i data, satisfying the standard requirement that $\widehat{\alpha}_n - \alpha = n^{-1}\sum_{i=1}^n I(Y_i) + R_n$, where $I(y)$ is the influence function with variance say τ^2, and where $n^{1/2}R_n \to_p 0$. This ensures that $A_n(t) = [nt]^{1/2}(\widehat{\alpha}_{[nt]} - \alpha)$ goes to a $N(0, \tau^2)$ for each positive t, and more generally, under mild extra regularity, that the process A_n is asymptotically zero-mean Gaußian with covariance structure $(s/t)^{1/2}\tau^2$ for $s \le t$. It follows that

$$B_n(t) = n^{-1/2}[nt](\widehat{\alpha}_{[nt]} - \widehat{\alpha}_n) = n^{-1/2}[nt]^{1/2}A_n(t) - n^{-1}[nt]A_n(1)$$

tends to $\tau W^0(t)$ in $D[0,1]$. Hence $M_n(t) = B_n(t)/\widehat{\tau}$ is a Brownian bridge in the large-sample limit, employing any reasonable estimator $\widehat{\tau}$ of τ. This makes previous techniques apply for testing the constancy hypothesis.

This apparatus may be used in the parametric models above when there is a sub-parameter to be concentrated on, like for example the shape parameter α in the Gamma model. As another example, suppose pairs (X_i, Y_i) are independent and that one is interested in monitoring their correlation coefficients ρ_i. Let $\widehat{\rho}_j$ be the usual estimator based on the first j pairs of data. Then $n^{-1/2}[nt](\widehat{\rho}_{[nt]} - \widehat{\rho}_n)/\widehat{\tau}$ tends to a Brownian bridge under the hypothesis of no change, for an appropriate scale estimate $\widehat{\tau}$. Tests can now be constructed as above. It should be pointed out that the convergence to a Brownian bridge in this and similar examples, though of course mathematically correct, may be slower than for our (2.3) type processes, particularly for smaller t.

3. Weighted monitoring processes

The tests developed above are all of 'omnibus type', constructed without any particular attention to the types of departures from H_0 that could be deemed more plausible than others. When specific alternatives to constancy are envisaged better tests can be constructed. This section works with rich classes of goodness-of-fit processes, emerging by integrating weight functions w.r.t. the basic monitoring process.

Consider

$$V_n(t) = \int_0^t K_n(s) \circ \mathrm{d}M_n(s) = \frac{1}{\sqrt{n}} \sum_{i \le [nt]} K_n(\tfrac{i}{n}) \circ \widehat{J}^{-1/2} u(Y_i, \widehat{\theta}),$$

where $K_{n,1}, \ldots, K_{n,p}$ are suitable weight functions, and $a \circ b$ for two vectors indicates coordinate-wise multiplication. In the interest of concise presentation we postpone until the remark ending this section discussing the exact regularity conditions needed for the intended martingale and stochastic integration calculus to work; these conditions at a minimum require the weight functions to tend in probability to predictable functions $K_1, \ldots, K_p$. Under such conditions one finds an appropriate generalisation of (2.4),

$$V_n \to_d V, \quad \text{with independent components } V_j = \int_0^t K_j(s)\,\mathrm{d}W_j^0(s), \tag{3.1}$$

under H_0. These are again normal processes, and calculations yield

$$\mathrm{cov}\{V_j(t_1), V_j(t_2)\} = \int_0^{t_1 \wedge t_2} K_j^2 \,\mathrm{d}s - \int_0^{t_1} K_j \,\mathrm{d}s \int_0^{t_2} K_j \,\mathrm{d}s. \tag{3.2}$$

Many tests can be constructed using these $\int K_n \circ \mathrm{d}M_n$ processes, along the lines of Section 2.3 for the special case of constant weight functions. The difference is that the limit processes become less tractable, but this is not a serious obstacle in view of the practically simple option of simulating these when needed. As an example, consider the supremum type test which for component j uses

$$U_{n,j} = \max_{0 \le t \le 1} |V_{n,j}(t)| = \frac{1}{\sqrt{n}} \max_{1 \le l \le n} \Big| \sum_{i \le l} K_{n,j}(\tfrac{i}{n}) (\widehat{J}^{-1/2})_{(j)} u(Y_i, \widehat{\theta}) \Big|,$$

which goes to $U_j = \max_{0 \le t \le 1} |\int_0^t K_j \,\mathrm{d}W_j^0|$ under H_0. This distribution must then be approximated by simulation. This remark applies also to sums over some or all components, like $\sum_{j=1}^p U_{n,j}$, and to other types of omnibus tests based on the $\int_0^t K_n \circ \mathrm{d}M_n$ processes.

A large class of tests which are more easily applied, in that no simulation of limit distributions is called for, is that of the chi squared tests. Focus first on a single component, say j. Divide again $[0,1]$ into m cells I_k, and let

$$\Delta V_{n,j,k} = \int_{I_k} \mathrm{d}V_{n,j}(s) = \frac{1}{\sqrt{n}} \sum_{i/n \in I_k} K_{j,n}(\tfrac{i}{n})(\widehat{J}^{-1/2})_{(j)} u(Y_i, \widehat{\theta}),$$

the increment over cell I_k. Then the vector of these, under H_0, tends to the vector $(\Delta V_{j,1}, \ldots, \Delta V_{j,m})^{\mathrm{t}}$, say, which is zero-mean normal with covariance matrix of the form $\Sigma_j = D_j - c_j c_j^{\mathrm{t}}$. Here D_j is diagonal with $d_{j,k} = \int_{I_k} K_j^2 \,\mathrm{d}s$ while vector c_j has $c_{j,k} = \int_{I_k} K_j \,\mathrm{d}s$. And

$$\Sigma_j^{-1} = D_j^{-1} + D_j^{-1} c_j c_j^{\mathrm{t}} D_j^{-1} / (1 - c_j^{\mathrm{t}} D_j^{-1} c_j),$$

giving the simple χ^2 test

$$Q_{n,j} = \sum_{k=1}^{m} \frac{(\Delta V_{n,j,k})^2}{\widehat{d}_{j,k}} + \Big(1 - \sum_{k=1}^{m} \frac{\widehat{c}_{j,k}^2}{\widehat{d}_{j,k}}\Big)^{-1} \Big(\sum_{k=1}^{m} \frac{\widehat{c}_{j,k}}{\widehat{d}_{j,k}} \Delta V_{n,j,k}\Big)^2. \tag{3.3}$$

Here $\widehat{c}_{j,k}$ and $\widehat{d}_{j,k}$ are natural estimates of the respective quantities, typically using $K_{n,j}$ instead of K_j. We have $Q_{n,j} \to_d \chi_m^2$ under H_0, unless $K = K_j$ is constant, which causes the final term to vanish and a limiting χ_{m-1}^2; see Test 1 of Section 2.4.

The above is valid for each of the components of V_n. Summing over some or all components gives a grander test,

$$Q_n = \sum_{j=1}^{p} Q_{n,j} \to_d \chi_{mp}^2 \quad \text{under } H_0. \tag{3.4}$$

This holds since the individual transformed Brownian bridges are independent.

Remark: Various sets of regularity conditions can be put up to ensure result (3.1). References include Gännsler and Häusler (1979), Rootzén (1980) and Jacod and Shiryayev (1987). The $K_{n,j}$ functions would either have to be predictable (essentially, left-continuous processes with values at time s not depending on outcomes of variables to be seen after time s), or to be well enough approximated by predictable processes. It would often suffice to have $K_{n,j}(s)$ of the form $K_{n,j}(s, \widehat{\alpha})$, where $\widehat{\alpha}$ is $n^{1/2}$-consistent for a certain α, and where $K_{n,j}(s, \alpha)$ is predictable; see Hjort (1990, Section 2.1). Next, $K_{n,j}(s)$ is required to converge in probability to a predictable limit function $K_j(s)$, and we should have $\int_0^t K_n K_n^{\mathrm{t}} \,\mathrm{d}s \to_p \int_0^t K K^{\mathrm{t}} \,\mathrm{d}s$ for each t. Finally a Lindeberg type condition is needed. We refer to Rootzén (1980) rather than spending too many efforts discussing the details of his conditions applied to our context. We note that Rootzén's methods and conditions also apply to the regression framework of Section 5.

4. Local power and optimal weight functions

The previous calculations have only been under the constancy hypothesis. This section works out limiting distribution results for various local alternatives, and also derives the optimal form of the weight functions when specific alternatives are being envisaged. We also learn about the expected shapes of the monitoring processes, under various alternatives. This is useful when it comes to assessing the type of change that has occurred, in cases where tests reveal that parameters have not been constant.

4.1. Limiting distributions for local alternatives. Consider alternatives in the vicinity of H_0, of the form $\theta_i = \theta_0 + \delta \circ h(\frac{i}{n})/n^{1/2} + O(1/n)$, for departure functions $h = (h_1, \ldots, h_p)^{\rm t}$ of suitable shapes and degrees of departure $\delta = (\delta_1, \ldots, \delta_p)^{\rm t}$. Then Y_i comes from

$$f(y, \theta_i) \doteq f(y, \theta_0)\,\{1 + u(y, \theta_0)^{\rm t}\, \delta \circ h(\tfrac{i}{n})/\sqrt{n}\}. \tag{4.1}$$

Consider again the basic score process $\psi_n(t, \theta_0) = n^{-1/2} \sum_{i \le [nt]} u(Y_i, \theta_0)$. Presently it has mean function

$$\mathrm{E}\psi_n(t, \theta_0) \doteq \frac{1}{n} \sum_{i \le [nt]} \int f(y, \theta_0) u(y, \theta_0) u(y, \theta_0)^{\rm t}\, \mathrm{d}y\, \delta \circ h(\tfrac{i}{n}) \to J \int_0^t \delta \circ h(s)\, \mathrm{d}s.$$

The covariance function is different from what it is under H_0, but only by an $O(1/n)$ effect. Further details ensure $\psi_n(t, \theta_0) \to_d J \int_0^t \delta \circ h(s)\, \mathrm{d}s + Z_0(t)$, where Z_0 is as with $\delta = 0$, that is, it is zero-mean Gaußian with covariance function $\min(t_1, t_2)\, J$.

Convergence of the estimated cumulative score process can now be assessed outside the null hypothesis. Some analysis, appropriately generalising arguments used in Section 2, leads to

$$\begin{aligned} \psi_n(t, \widehat\theta) &= \psi_n(t, \theta_0) - t\psi_n(1, \theta_0) + o_p(1) \\ &\to_d J\Big(\int_0^t \delta \circ h\, \mathrm{d}s - t \int_0^1 \delta \circ h\, \mathrm{d}s\Big) + Z_0(t) - tZ_0(1). \end{aligned}$$

It follows that

$$M_n(t) = \widehat{J}^{-1/2} \psi_n(t, \widehat\theta) \to_d J^{1/2} \int_0^t \delta \circ (h - \bar h)\, \mathrm{d}s + W^0(t), \tag{4.2}$$

under $\theta_i = \theta_0 + \delta \circ h(\frac{i}{n})/\sqrt{n}$ circumstances. Here $\bar h = \int_0^1 h(s)\, \mathrm{d}s$, and W^0 is a vector of p independent Brownian bridges W_j^0. The above generalises result (2.4), which corresponds to the case of h being constant.

Without going too much into the technical details we note that the methods and results of Rootzén and others, as explained in the remark ending the previous section, yield results for the weighted processes $\int_0^t K_{n,j}\,\mathrm{d}M_{n,j}(s)$ studied in Section 3, in the present local alternatives framework. Thus for each $V_{n,j}$ we have, under (4.1) circumstances,

$$\int_0^t K_{n,j}(s)\,\mathrm{d}M_{n,j}(s) \to_d \int_0^t K_j(s)(J^{1/2})_{(j)}\delta\circ(h(s)-\bar h)\,\mathrm{d}s+\int_0^t K_j(s)\,\mathrm{d}W_j^0(s) \quad (4.3)$$

modulo regularity assumptions discussed before. Result (4.3) generalises that of (3.1).

4.2. Weight functions and optimal local power. Using results above one may calculate approximate power for various tests based on M_n and $\int K_n\,\mathrm{d}M_n$, against various alternatives of interest. Given departure functions $h_1,\ldots,h_p$, results (4.2)–(4.3) lead to expressions for limiting power, depending on the degrees of departure $\delta_1,\ldots,\delta_p$. This approach is rather complicated but nevertheless quite useful when it comes to exploring the performance of several of the supremum and integration based tests portrayed in Section 2.3. It lends itself most easily to the chi square type tests, however.

Consider the local power of the chi squared tests constructed in Section 3. Focus on a single component j first. Now study the χ^2 tests based on quadratic forms in $\Delta V_{n,j,k}=\int_{I_k}K_{n,j}(s)\,\mathrm{d}M_{n,j}(s)$. The vector of such increments tends to

$$\Big(\int_{I_1}K_jH_j\,\mathrm{d}s+\Delta V_{j,1},\ldots,\int_{I_m}K_jH_j\,\mathrm{d}s+\Delta V_{j,m}\Big)^{\mathrm{t}},$$

where $H_j(s)=(J^{1/2})_{(j)}\delta\circ(h(s)-\bar h)$. For the test of (3.3) one therefore finds a noncentral chi squared limit,

$$Q_{n,j}\to_d \chi^2_m(\lambda_j),\quad \text{where } \lambda_j=a_j^{\mathrm{t}}(D_j-c_jc_j^{\mathrm{t}})^{-1}a_j, \quad (4.4)$$

with $a_{j,k}=\int_{I_k}K_jH_j\,\mathrm{d}s$. The excentre parameter λ_j can also be written

$$\sum_{k=1}^m\frac{(\int_{I_k}K_jH_j\,\mathrm{d}s)^2}{\int_{I_k}K_j^2\,\mathrm{d}s}+\Big(1-\sum_{k=1}^m\frac{(\int_{I_k}K_j\,\mathrm{d}s)^2}{\int_{I_k}K_j^2\,\mathrm{d}s}\Big)^{-1}\Big(\sum_{k=1}^m\frac{\int_{I_k}K_j\,\mathrm{d}s}{\int_{I_k}K_j^2\,\mathrm{d}s}\int_{I_k}K_jH_j\,\mathrm{d}s\Big)^2.$$

The bigger $\lambda_j=\lambda_j(K_j)$, the greater power of the tests. The optimal choice of K_j can be proved to be

$$K_j(s)=H_j(s)=(J^{1/2})_{(j)}\{\delta\circ(h(s)-\bar h)\} \quad (4.5)$$

(or proportional to this choice), see Appendix I. It attains the maximum possible value $\lambda_j(H_j)=\max_{K_j}\lambda_j(K_j)=\int_0^1 H_j(s)^2\,\mathrm{d}s$, and the corresponding optimal local power in (4.4).

The above is valid for each component of the vector $V_n(t) = \int_0^t K_n \circ \mathrm{d}M_n$. For combined χ^2 tests the optimal weight function is to make K_n proportional to a consistent estimate of

$$K(s) = J^{1/2}\{\delta \circ (h(s) - \bar{h})\},$$

where $h_1, \ldots, h_p$ are the departure functions of interest. This gives χ^2_{mp} tests of the form (3.4), with limiting maximal power determined by the excentre parameter for the χ^2_{mp} distribution, namely

$$\lambda = \sum_{j=1}^{p} \int_0^1 \{(J^{1/2})_{(j)}(\delta_j h_j(s) - \delta_j \bar{h}_j)\}^2 \,\mathrm{d}s = \int_0^1 (\delta \circ (h - \bar{h}))^{\mathrm{t}}\, J(\delta \circ (h - \bar{h}))\,\mathrm{d}s.$$

For comparison, the simpler test statistic using constant K_j functions, corresponding to A_n^2 of Section 2.3, has a $\chi^2_{(m-1)p}(\mu)$ limit distribution with excentre parameter $\sum_{j=1}^{p} \sum_{k=1}^{m} (\int_{I_k} H_j \,\mathrm{d}s)^2/|I_k|$.

4.3. Shape of M_n and V_n plots. Result (4.2) also provides useful information about the expected shape of M_n plots under different circumstances. Take the δ_js to be equal, for simplicity. We see then that the expected $M_{n,j}$ plot is proportional to $(J^{1/2})_{(j)}$ times the vector of $\int_0^t (h - \bar{h})\,\mathrm{d}s$. If h_j is a change point departure function, say zero on $[0, a)$ and then equal to b on $[a, 1]$, this integral is proportional to the triangular function with value $-(1-a)t$ and $a(t-1)$ on respectively $[0, a]$ and $[a, 1]$. If on the other hand h_j describes a linear trend of change, as in $h_j(s) = cs$, then the expected shape of the appropriate monitoring component is $-\frac{1}{2}ct(1-t)$, a symmetric parabola. Illustrations 1 and 2 of Section 6 give examples of such behaviour.

It is also possible to estimate the position of break points, for any of the parameters, in cases where initial tests indicate that the (2.1) hypothesis does not hold. Suppose for simplicity of discussion that there is only one parameter to consider, and that this parameter has a jump at an unknown position a. Then M_n can be represented as a triangular function plus noise, as explained above. An estimate of a emerges by fitting the M_n to a triangular shape and looking for its top point.

If the alternative to H_0 of (2.1) is that of a linear trend, say $\theta_i = \theta_0 + (i/n)c$, then the optimal weight function to use, by result (4.5), is $K(s) = s - \frac{1}{2}$. Thus $V_n(t) = \int_0^t (s - \frac{1}{2})\,\mathrm{d}M_n(s)$ is most readily detecting the existence of such a trend. The limiting null distribution of $\max_t |V_n(t)|$ is that of $\max_t |\int_0^t (s - \frac{1}{2})\,\mathrm{d}W^0(s)|$, for example. Simulations showed that this distribution has median about 0.32 and upper 0.05 quantile point about 0.64, for example. In brief simulations for the situation examined in Illustration 2 of Section 6 below, the $\max_t |V_n(t)|$ test was indeed consistently better than the $\max_t |M_n(t)|$ test. This also suggests additional plots to be constructed for special tasks, like plotting $V_n(t)/\tau(t)$ in the mentioned situation, where $\tau(t)^2 = \frac{1}{3}(t - \frac{1}{2})^3 + \frac{1}{24} - \frac{1}{4}\{(t - \frac{1}{2})^2 - \frac{1}{4}\}^2$ is the limiting variance, as per (3.2).

5. Regression models

There are many situations where observations Y_i have relevant covariate information x_i, and where interest would focus on whether the precise form in which the distribution of Y_i depends on x_i somehow could have changed over time. Under suitable assumptions the previous no-covariate methodology can be readily extended to the regression framework.

We take the Y_is to be conditionally independent given the sequence of x_is, and will analyse sampling behaviour in such a conditional framework, considering $x_1, \ldots, x_n$ as given. We shall however assume that the x_is arrive as an exchangeable or ergodic sequence, where averages stabilise in probability; they could for example themselves be i.i.d. outcomes from some random mechanism.

Assume that Y_i given x_i comes from a density of the form $f(y_i \,|\, x_i, \theta_i)$. Let $u(y \,|\, x, \theta)$ and $i(y \,|\, x, \theta)$ be the first and second derivatives of $\log f(y \,|\, x, \theta)$ with respect to the parameters. Under model conditions and the hypothesis H_0 that the parameters θ_i do not change, there is process convergence

$$\psi_n(t, \theta_0) = \frac{1}{\sqrt{n}} \sum_{i \le [nt]} u(Y_i \,|\, x_i, \theta_0) \to_d Z_0(t) \quad \text{in } D[0,1], \tag{5.1}$$

under mild regularity conditions. Here θ_0 is the common true parameter value. The variance matrix of $\psi_n(t, \theta)$ is $n^{-1}[nt]J_{[nt]}$, where

$$J_n = n^{-1} \sum_{i=1}^{n} V(x_i, \theta_0) \to_p J \quad \text{as } n \to \infty, \tag{5.2}$$

writing $V(x_i, \theta_0)$ for $\operatorname{Var} u(Y_i \,|\, x_i, \theta_0)$ (assumed to be finite). That J_n stabilises follows from our ergodicity assumption about the x_i sequence. It follows from this, and the Lindeberg-extended Donsker theorem, that the Z_0 limit has independent increments with covariance structure $\min(t_1, t_2)J$.

Likewise other arguments of Section 2 can be utilised and generalised to the present framework, showing that if $\widehat{\theta}$ is the maximum likelihood estimator, then

$$\begin{aligned} \psi_n(t, \widehat{\theta}) &= \frac{1}{\sqrt{n}} \sum_{i \le [nt]} u(Y_i \,|\, x_i, \widehat{\theta}) \\ &\doteq \psi_n(t, \theta_0) - tJ_{[nt]}J^{-1}\psi_n(1, \theta_0) \to_d Z(t) = Z_0(t) - tZ_0(1). \end{aligned}$$

Again, this process is Gaußian with zero mean and covariance structure $t_1(1 - t_2)J$ for $t_1 \le t_2$. And a canonical monitoring process emerges, of the form

$$M_n(t) = \widehat{J}^{-1/2}\psi_n(t, \widehat{\theta}) \quad \text{for } 0 \le t \le 1. \tag{5.3}$$

Here $\widehat{J}$ is any reasonable estimator of J_n, and required to be consistent for the limiting matrix J as n grows; a natural choice is $n^{-1}\sum_{i=1}^{n} V(x_i, \widehat{\theta})$. The limiting process $J^{-1/2}Z$ is that of p independent Brownian bridges, where p is the number of parameters in θ.

We note that the constructions and results of weighted monitoring processes, discussed in Sections 3 and 4, can be extended to the regression case, partly in view of the methodology of Rootzén (1980) and Gännsler and Häusler (1979).

Example 1: Let Y_i given x_i be normal $(x_i^{\rm t}\beta, \sigma^2)$, where x_i is p-dimensional and $\beta_1, \ldots, \beta_p, \sigma$ are unknown parameters. Then

$$J_n = \sigma^{-2}\begin{pmatrix} n^{-1}\sum_{i=1}^n x_i x_i^{\rm t} & 0 \\ 0 & 2 \end{pmatrix},$$

and the monitoring vector process takes the form

$$M_n(t) = \frac{1}{\sqrt{n}}\sum_{i\le[nt]}\begin{pmatrix} Z_i x_i \\ 2^{-1/2}(Z_i^2-1)\end{pmatrix}, \quad \text{where } Z_i = (Y_i - x_i^{\rm t}\widehat{\beta})/\widehat{\sigma}.$$

This appropriately generalises the process of Example 1, Section 2.4. The last component process of M_n can be used to look for changes in the σ parameter, for example; see the second illustration of Section 6.

Example 2: Let Y_i be Poisson with mean parameter $\exp(x_i^{\rm t}\beta)$, and let $\widehat{\beta}$ be the maximum likelihood estimator. The monitoring process takes the form

$$M_n(t) = \Big\{n^{-1}\sum_{i=1}^n \exp(x_i^{\rm t}\widehat{\beta})\,x_i x_i^{\rm t}\Big\}^{-1/2}\frac{1}{\sqrt{n}}\sum_{i\le[nt]}\{Y_i - \exp(x_i^{\rm t}\widehat{\beta})\}x_i \quad \text{for } 0\le t\le 1.$$

Remark: There are situations where ergodicity of the x_is cannot be assumed, for instance in the case where one of the covariates in question is the running time i itself. Without ergodicity one is faced with additional problems with no clear-cut general-purpose solution. The variance function of the score process $\psi_n(t,\theta_0)$ may not factorise into a scalar function of t and the information matrix J, which precludes a global standardisation (that is, premultiplying by $J^{-1/2}$) of the score increments $u(Y_i\,|\,x_i,\theta)$. Instead, a local standardisation should be used. However, the obvious candidate for such a standardisation, premultiplying by $V(x_i,\theta_0)^{-1/2}$, is not applicable if $V(x_i,\theta_0)$ is not of full rank. Further work is needed to solve such problems in a satisfactory manner. Note however that if θ is scalar, these problems do not emerge since the variance function of the score process trivially factorises.

6. Illustrations and applications

It is easy to illustrate the behaviour of our canonical and weighted monitoring processes, under the hypothesis of no change as well as under various discontinuity alternatives, via simulations from models of interest. We briefly provide some such as Illustrations 1 and 2 below. In addition an application is described using data from the Dutch Ombudsman, looking for changes in Poisson model parameters.

Illustration 1: Suppose data Y_i come from a Gamma (a_1, b_1) for $i = 1, \ldots, 100$ and from another Gamma (a_2, b_2) for $i = 101, \ldots, 200$. We take here the mean levels a_1/b_1 and a_2/b_2 to be the same, but scale up the standard deviation $a_2^{1/2}/b_2$ to be 1.25 times that of $a_1^{1/2}/b_1$. It is not easy to spot from just a plot of the 200 data points that anything has happened to the underlying parameters. However, the first component of the monitoring process $M_n(t)$, see Example 2 of Section 2.4, signals by its triangular shape and maximum size that something has happened around $t = \frac{1}{2}$, that is, around data point 100 of the 200. The upper 0.05 quantile of the distribution of $\max_{0 \le t \le 1} |W^0(t)|$ is 1.358, so $M_{n,1}$ clearly does not agree with the Brownian bridge behaviour it should have had under the hypothesis of no change.

— *Figure 1 around here, see page 21* —

Illustration 2: Consider a regression situation where data Y_i are normal $(a + bx_i, \sigma^2)$, where the regression coefficients do not change, but where σ_i slowly increases with time. Specifically, we simulate 200 points around the regression line $1.11 + 2.22\,x$, with xs being uniform on the unit interval, and with $\sigma_i = 1 + 0.5\,i/200$, increasing linearly from 1 to 1.5. It is quite difficult to spot from scatterplots or residuals that the standard deviation has been increasing linearly over time. But a look at the three monitoring plots quickly shows that the third component reaches outside ± 1.358, the 0.95-probability band, and that the two first components, corresponding to the $a + bx$ part, stay nicely within. This time the third monitoring process approximately forms a parabola departure from zero, indicating as per the theory of Section 4.3 that the non-constancy of the σ parameter might be in the form of a linear trend. As explained there alternative plots involving $V_{n,3}(t) = \int_0^t (s - \frac{1}{2})\,\mathrm{d}M_{n,3}(s)$, not shown in our article, are even better at detecting linear trends in the σ parameter.

— *Figure 2 around here, see page 22* —

An application: The first paragraph of article 78a of the Dutch Constitution reads as follows: "On request or on his own initiative the National Ombudsman shall investigate the actions of administrative authorities of the national government and of other administrative authorities designated by or pursuant to Act of Parliament."

In the Netherlands, criminals may receive psychiatric treatment in so-called TBS-institutions as part of their sentence. (This Dutch acronym for 'terbeschikking-stelling' indicates in this case 'to be put at the disposal of', by the authority, for psychiatric treatment.) The psychiatric treatment precedes the actual prison sentence. Criminals on a waiting list for placement in a TBS-institution are temporarily imprisoned under relatively poor conditions. After receiving eleven complaints between December 1995 and February 1996, the National Ombudsman decided to investigate on his own initiative the TBS waiting lists, and especially the waiting time involved. The investigation was reported on in National Ombudsman (1996). In the tables presented in Appendix II the number of TBS-sentences and the number of ended TBS-treatments are given for each month during the years 1984–1992. Figure 3 displays the corresponding monitoring plots.

— *Figure 3 about here, see page 23* —

The monitoring plot of the expected number of ended treatments exceeds the value 1.358, indicating that the hypothesis of constancy of this parameter should be rejected at the 5% significance level. Moreover, the plot resembles a triangular shape reaching its maximum deviation from the time axis in March 1990; as explained in Section 4.3, this is indicative of a change point. The plot suggests that around March 1990, there was a sudden decrease in the expected number of ended treatments. A possible explanation could be the increased complexity of the psychiatric problems of the clients within the TBS-system. Due to several policy changes in Dutch psychiatric care in the late eighties, it became easier for unwilling psychiatric patients to avoid admittance to psychiatric institutions. For these patients (among them extreme psychotic patients) the TBS-system started to act as a dust-bin.

7. Supplementing remarks

This section lists various comments pertaining to the use and further study of our methods.

7.1. Model-robust variance matrix estimation. In the framework of Section 2, suppose that under H_0 there is indeed a common density $f(y)$ for the Y_is, but that this unchanged density is not a member of the parametric class $f(y,\theta)$. Still the maximum likelihood estimator is meaningful, taking aim at the least false parameter value θ_0 which minimises the Kullback–Leibler distance from the true to the parametrised density, and there is convergence in distribution $\sqrt{n}(\widehat{\theta}-\theta_0)$ towards a normal $(0, J^{-1}KJ^{-1})$. Here $J = -\mathrm{E}_f i(Y_i,\theta_0)$ and $K = \mathrm{Var}_f\, u(Y_i,\theta_0)$; these coincide under model conditions but not in general.

To analyse the implications of such a model-robust viewpoint for our methods, consider first the $\psi_n(t,\theta_0)$ process of Section 2.1. It is clear that its limit Z_0 has covariance structure $\min(t_1,t_2)K$. The Taylor expansion and other arguments of Section 2.2 now show, mutatis mutandem, that $\psi_n(t,\widehat{\theta})$ is well approximated with $\psi_n(t,\theta_0) - tJ_{[nt]}J^{-1}\psi_n(1,\theta_0)$, with limit $Z(t) = Z_0(t) - tZ_0(1)$. This has covariance structure $t_1(1-t_2)K$ for $t_1 \le t_2$. It follows that the natural model-robust monitoring process is $M_n^*(t) = \widehat{K}^{-1/2}\psi_n(t,\widehat{\theta})$, that is, just like in (2.3), but for model-robust safety employing $\widehat{K}$ to estimate the variance matrix of the score function, rather than $\widehat{J}$.

In the normal-model case of Example 1, Section 2.4, this leads to using

$$M_n^*(t) = \frac{1}{\sqrt{n}} \sum_{i \le [nt]} \begin{pmatrix} 1 & \widehat{\kappa}_1 \\ \widehat{\kappa}_1 & 2 + \widehat{\kappa}_2 \end{pmatrix}^{-1/2} \begin{pmatrix} Z_i \\ Z_i^2 - 1 \end{pmatrix}, \quad \text{with } Z_i = (Y_i - \widehat{\mu})/\widehat{\sigma}$$

instead of the simpler one given there. Here $\widehat{\kappa}_1$ and $\widehat{\kappa}_2$ are sample-based estimates of skewness and kurtosis. Similarly, in the Poisson model Example 3 of Section 2.4, the model-robust viewpoint leads to using $M_n^*(t) = n^{-1/2} \sum_{i \le [nt]} (Y_i - \bar{Y})/\widehat{\sigma}$, with $\widehat{\sigma} = \{n^{-1} \sum_{i=1}^n (Y_i - \bar{Y})^2\}^{1/2}$ replacing the simpler model-based $\bar{Y}^{1/2}$.

7.2. An innovation approach. The monitoring process $M_n(t)$ has the attractive property of converging under parameter constancy to a vector of p independent Brownian bridges. This property is lost by stochastic integration, and the limit in distribution of the weighted monitoring process $\int K_n(s)\,\mathrm{d}M_n(s)$ becomes less tractable. In Koning (1999) and Hjort and Koning (1999b) this problem is overcome by transforming the monitoring process $M_n(t)$ into a process $\widetilde{M}_n(t)$ which converges under parameter constancy to a vector of p independent Brownian motions. This involves innovation transforms in the spirit of Khmaladze (1981). Hence, under suitable conditions on K_n the process $\int K_n(s)\,\mathrm{d}\widetilde{M}_n(s)$ also converges to a vector of p independent Brownian motions (albeit time-transformed). This solution takes a slight toll, in the sense that the rate of convergence of $\widetilde{M}_n(t)$ is a factor $\log n$ less than the rate of convergence of $M_n(t)$.

7.3. Extensions to Markov and time series models. Our methods are not limited to the context of independence considered in earlier sections of this paper, but have a far larger generality. In principle, they could be applied to any statistical model in which we can define a cumulative score process $\psi_n(t, \theta_0)$ which satisfies (2.2), the starting point for our methodology. For this purpose martingale central limit theorems would often suffice for verification. Presently we verify this property for Markov models, and then comment shortly on Gaußian autoregressive type time series models.

For a one-step memory Markov model, taking for technical expediency the viewpoint that the value y_1 of the first variable Y_1 simply is fixed and given, and not informative for the Markov parameters, we may define the log-likelihood 'at time t', that is, corresponding to the sample $Y_1, \ldots, Y_{[nt]}$, as $\sum_{i=2}^{[nt]} \log f(Y_{i-1}, Y_i; \theta)$. Here $f(y, y'; \theta)$ is the density of the transition measure (cf. Billingsley, 1961, p. 4). Defining $u(y, y'; \theta)$ as the first derivative of $f(y, y'; \theta)$ with respect to θ, it can be shown that the cumulative score process

$$\psi_n(t; \theta_0) = \frac{1}{\sqrt{n}} \sum_{i \le [nt]} u(Y_{i-1}, Y_i; \theta_0)$$

is a martingale with asymptotic variance function tJ, where J is the information matrix (see Billingsley, 1961, p. 6). Under mild regularity conditions, application of Rootzén's theorem now yields (2.2).

For stationary Gaußian AR, ARMA and ARIMA time series models we may use the so-called conditional likelihood, see e.g. equation (7.1.2) in Box, Jenkins and Reinsel (1994), to define a cumulative score process which becomes a martingale. This again yields result (2.2) if Rootzén's theorem applies.

Acknowledgements. This work grew out of material originally presented by one of us at the Oslo May 1998 conference on discontinuous phenomena in statistics, sponsored by the the European Science Foundation via their Highly Structured Stochastic Systems programme. A.K. is also grateful for hospitality and partial support in connection with visits to the Department of Mathematics at the University of Oslo.

References

Billingsley, P. (1961). *Statistical Inference for Markov Processes.* University of Chicago press, Chicago.

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994). *Time Series Analysis: Forecasting and Control* (3rd ed.). Prentice Hall, Englewood Cliffs, N.J.

Durbin, J. (1973). Weak convergence of the sample distribution function when parameters are estimated. *Annals of Statistics* **1**, 279–290.

Gänssler, P. and Häusler, E. (1979). Remarks on the functional central limit theorem for martingales. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **50**, 237–243.

Hjort, N.L. (1990). Goodness of fit tests in models for life history data based on cumulative hazard rates. *Annals of Statistics* **18**, 1221–1258.

Hjort, N.L. and Koning, A.J. (1999a). Has the density changed? Unpublished manuscript, Department of Mathematics, University of Oslo.

Hjort, N.L. and Koning, A.J. (1999b). A general innovation approach towards goodness of fit testing. Manuscript, in progress.

Jacod, J. and Shiryaev, A.N. (1987). *Limit Theorems for Stochastic Processes.* Springer-Verlag, Berlin.

Khmaladze, E.V. (1981). Martingale approach in the theory of goodness of fit tests. *Theory of Probability and its Applications* **26**, 246–265.

Koning, A.J. (1999). Goodness of fit for the constancy of a classical statistical model over time. *Biometrika*, to appear.

Koning, A.J. and Does, R.J.M.M. (1997). Cusum charts for preliminary analysis of individual observations. *Econometric Institute Report* **9727/A**, Rotterdam.

Miller, R.G. and Siegmund, D. (1982). Maximally selected chi-square statistics. *Biometrics* **38**, 1011–1016.

de Nationale Ombudsman (1996). *Openbaar rapport 96.00922.* de Nationale Ombudsman, 's-Gravenhage.

Ploberger, W. and Krämer, W. (1992). The CUSUM test with OLS residuals. *Econometrica* **60**, 271–285.

Pollard, D. (1984). *Convergence of Stochastic Processes.* Springer, New York.

Ripley, B.D. (1987). *Stochastic Simulation.* Wiley, New York.

Rootzén, H. (1980). On the functional limit theorem for martingales. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **51**, 79–93.

Schruben, L.W. (1982). Detecting initialization bias in simulation output. *Operation Research* **30**, 569–590.

Schruben, L.W. (1983). Confidence interval estimation using standardized time series. *Operation Research* **31**, 1090–1108.

Sullivan, J.H. Woodall, W.H. (1996). A control chart for preliminary analysis of individual observations. *Journal of Quality Technology* **28**, 265–278.

Appendix I: optimal K choice

Here an optimality problem raised in Section 4 is solved. The result leads to the optimal choice of weight functions $K_{n,j}(s)$ when basing chi squared tests on increments of $\int_0^t K_{n,j}(s)\,\mathrm{d}M_{n,j}(s)$; see also Section 3. Specifically, we maximise the $\lambda_j = \lambda_j(K_j)$ parameter of equation (4.4) over all K_j functions. We may omit the index j in what follows. We show that $\lambda(K) \le \int_0^1 H^2\,\mathrm{d}s$, for any nontrivial weight function K; note that this bound is then achieved with K proportional to H.

Introduce the vector function $\ell = (KJ_1, \ldots, KJ_m)^{\mathrm{t}}$, where $J_k(t)$ is indicator for t belonging to I_k, and let $L = \ell - c$, where $c = \int_0^1 \ell(s)\mathrm{d}s$. Note next that

$$\int_0^1 LH\,\mathrm{d}s = \int_0^1 \ell H\,\mathrm{d}s = \Big(\int_{I_1} KH\,\mathrm{d}s, \ldots, \int_{I_m} KH\,\mathrm{d}s\Big)^{\mathrm{t}}.$$

Moreover, one shows that $\int_0^1 LL^{\mathrm{t}}\,\mathrm{d}s$ equals $D - cc^{\mathrm{t}}$ (in the notation of Section 3, but with subscript j omitted). It follows that

$$\lambda(K) = \Big(\int_0^1 LH\,\mathrm{d}s\Big)^{\mathrm{t}} \Big(\int_0^1 LL^{\mathrm{t}}\mathrm{d}s\Big)^{-1} \int_0^1 LH\,\mathrm{d}s.$$

That $\lambda(K) \le \int_0^1 H^2\,\mathrm{d}s$ now follows from a generalised version of the Cauchy–Schwartz inequality: the matrix

$$\Omega = \begin{pmatrix} \int_0^1 H^2\,\mathrm{d}s & \int_0^1 L^{\mathrm{t}} H\,\mathrm{d}s \\ \int_0^1 LH\,\mathrm{d}s & \int_0^1 LL^{\mathrm{t}}\,\mathrm{d}s \end{pmatrix}$$

is nonnegative definite, implying that $\Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21}$ is nonnegative definite too. This proves the claim.

Appendix II: Dutch TBS data

Number of TBS-sentences

	'84	'85	'86	'87	'88	'89	'90	'91	'92
Jan	1	7	8	7	8	9	8	5	4
Feb	5	11	7	2	9	9	12	6	12
Mar	10	10	14	3	11	9	10	8	3
Apr	13	8	4	7	5	2	9	6	13
May	6	4	4	7	7	9	11	14	6
Jun	5	5	7	5	9	7	9	9	7
Jul	15	6	8	10	9	10	8	9	14
Aug	5	8	2	4	3	11	3	6	11
Sep	5	8	9	8	4	6	9	11	8
Oct	9	9	7	7	8	6	3	17	8
Nov	6	16	14	6	8	10	7	14	14
Dec	10	14	10	10	9	6	6	12	17

Number of ended TBS-treatments

	'84	'85	'86	'87	'88	'89	'90	'91	'92
Jan	10	6	5	6	10	10	2	4	4
Feb	7	9	9	10	7	8	2	4	6
Mar	4	6	7	10	5	10	6	9	6
Apr	5	11	4	9	6	5	5	8	6
May	11	7	8	3	10	8	12	8	6
Jun	3	3	8	5	4	7	8	6	6
Jul	8	11	4	8	4	7	0	12	4
Aug	6	5	5	7	3	6	4	4	7
Sep	4	3	3	5	6	6	2	13	6
Oct	2	7	4	10	4	13	9	10	2
Nov	6	5	5	5	6	6	8	6	5
Dec	10	8	8	6	12	9	5	7	6

TABLE: Data given in de Nationale Ombudsman (1996), p. 82–83. See the discussion in Section 6.

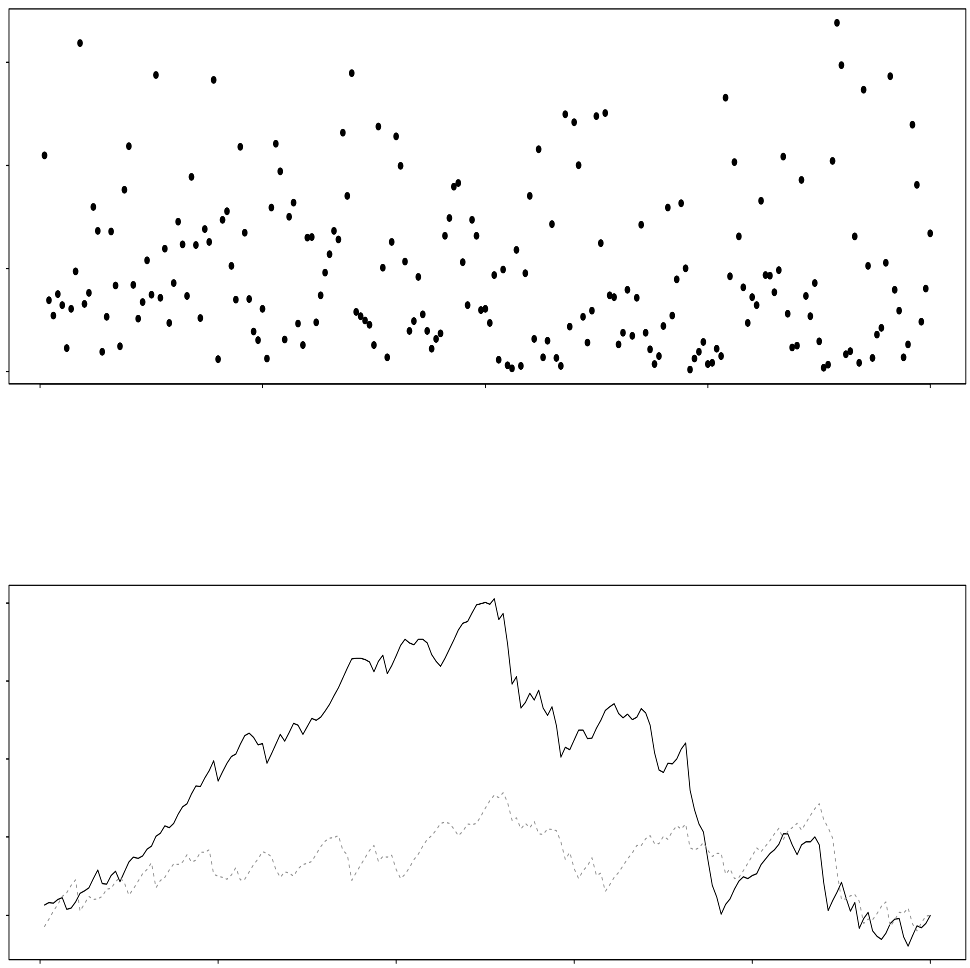

FIGURE 1: The first 100 and the following 100 data points have been drawn from two different Gamma densities; these have the same mean level, but the second has standard deviation 1.25 times bigger than that of the first. This aspect is barely visible from the data figure, but is being brought out by the monitoring processes; the maximum absolute value of the first of these exceeds the null-distribution 0.95 point of 1.358, for example. The triangular shape correctly indicates that the non-constancy is in form of a break point about half-way through the data.

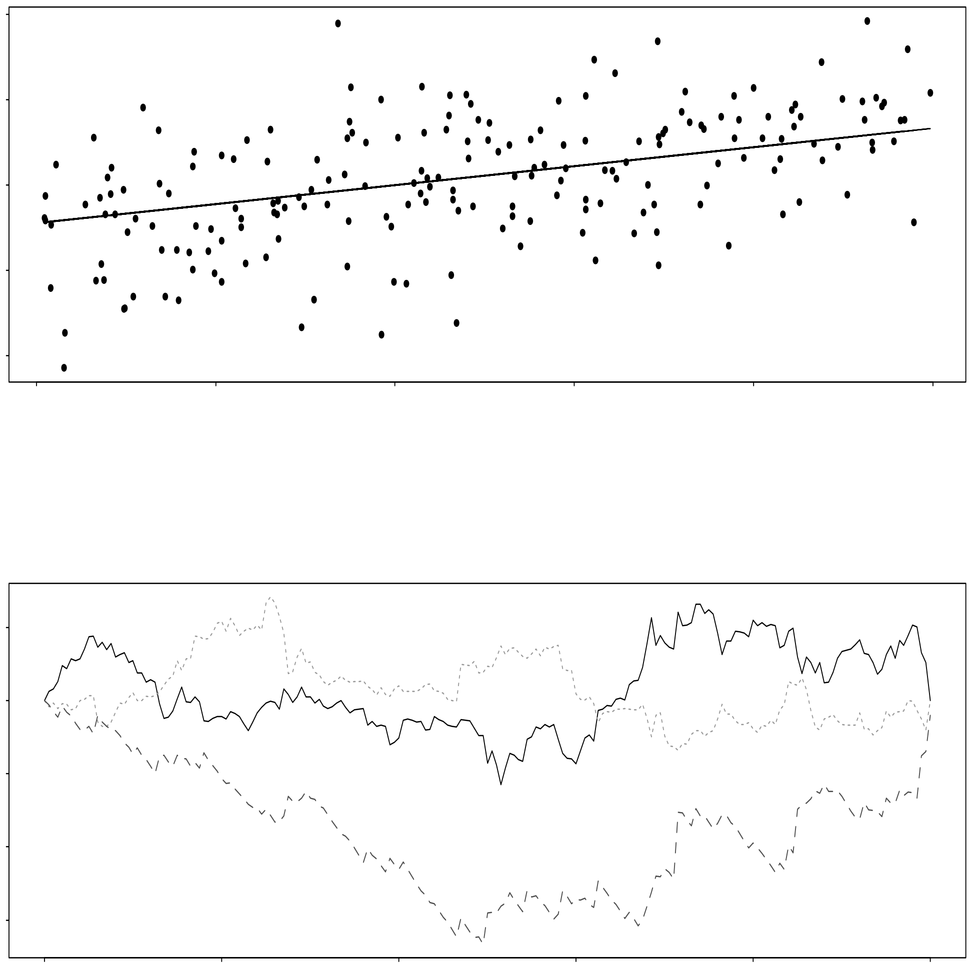

FIGURE 2: Here $n = 200$ pairs are generated by letting x_is be independent (and not sorted) uniformly on $(0, 1)$ and then using the the $y_i = a + bx_i + \varepsilon_i$ model, with normal errors $N(0, \sigma_i^2)$ and using $\sigma_i = 1 + 0.5\, i/n$. The monitoring process plots pick up the aspect that σ is not constant, in that its maximum absolute value exceeds the 1.358 value, for example. Also its approximately parabolic shape helps identify the type of non-constancy of the σ parameters.

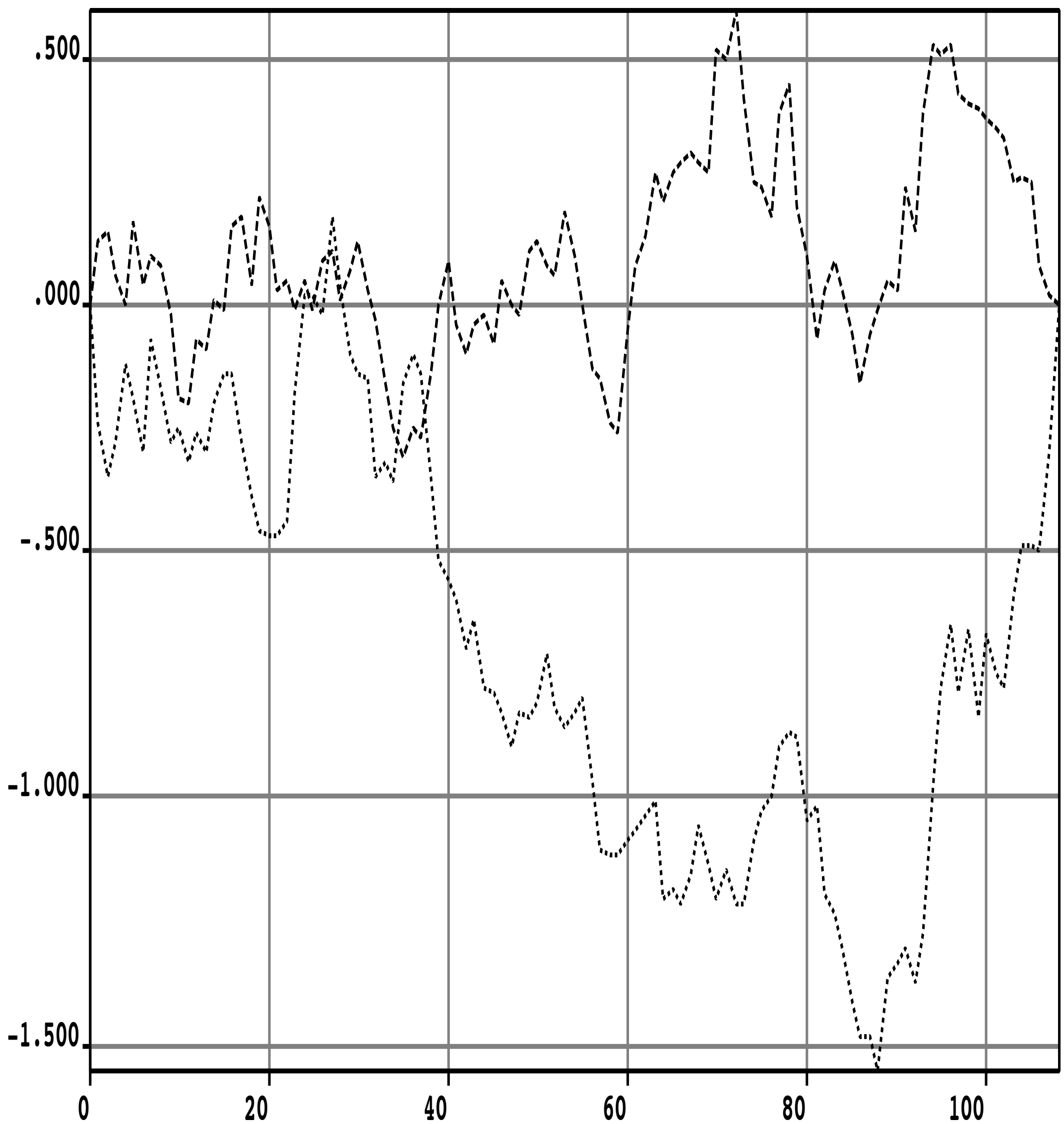

FIGURE 3: Monitoring plots for checking the constancy of Poisson parameters, for the two sets of Dutch Ombudsman data (see Appendix II). The plot for the expected number of TBS-sentences does not indicate any departure from the hypothesis of constancy, whereas the plot for the expected number of ended treatments indicates that this parameter has not been constant over the time period studied. The triangular shape indicates a sudden decrease around March 1990.