JUMP-PRESERVING MONITORING OF DEPENDENT TIME SERIES USING PILOT ESTIMATORS

Ansgar Steland[1]
Ruhr-Universität Bochum, Germany
ansgar.steland@ruhr-uni-bochum.de

Abstract. An important problem of the statistical analysis of time series is to detect change-points in the mean structure. Since this problem is a one-dimensional version of the higher dimensional problem of detecting edges in images, we study detection rules which benefit from results obtained in image processing. For the sigma-filter studied there to detect edges, asymptotic bounds for the normed delay have been established for independent data. These results are considerably extended in two directions. First, we allow for dependent processes satisfying a certain conditional mixing property. Second, we allow for more general pilot estimators, e.g., the median, resulting in better detection properties. A simulation study indicates that our new procedure indeed performs much more better.
AMS Classification: 62L10, 62P20, 68U10
Keywords: *Image processing; Nonparametric regression; Quality Control, Structural Change.*

1. Introduction

When analyzing time series, e.g., financial prices or returns of stocks or bonds, or univariate statistics calculated from multivariate time series, we are often concerned with non-stationary time series. Indeed, often the non-stationary components (trends or heteroskedasticity) are the most informative characteristics of a series. For example, financial return series are often stationary but affected by conditional heteroscedasticity, and detecting deterministic increases of the dispersion is important, since the dispersion, called volatility, is a direct measure of the risk associated with an investment in that asset. Thus, the application of sequential monitoring procedures is of considerable interest. Even nowadays volatility is often simply measured by

Date: October 2003.
[1]Address of correspondence: Ansgar Steland, Ruhr-Universität Bochum, Fakultät für Mathematik, Mathematik 3 NA 3/71, Universitätsstr. 150, D-44780 Bochum, Germany.

the empirical standard deviation. This approach causes artifacts if the distribution has fat tails and outliers, and monitoring rules based on it may yield a substantial delay. This also applies to classic smoothing procedures. In this article we study a class of jump-preserving estimators and related monitoring procedures which try to detect abrupt changes faster than related classic procedures.

Methods for analyzing and detecting departures from stationarity have received considerable interest in the statistics and econometrics literature. Let us first briefly discuss some a posteriori approaches before motivating our proposal in greater detail. The problem of a posteriori change-point (structural break) detection with a focus on linear models has been extensively discussed in the econometric literature, e.g. Andrews (1988), Andrews and Ploberger (1994), Bai and Perron (1998), or Vogelsang (1998). Statistical methods based on U statistics can be found in Ferger (1994a). Kim and Hart (1998) studied an approach based on Fourier methods. A posteriori methods based on kernel-weighted averages have been investigated by Brodsky and Darkhovsky (1993, 2000), Ferger (1994b, 1994c, 1995, 1996), and Huskova and Slaby (2001). To estimate change-points and the regression function we also refer to Müller (1992) and Wu and Chu (1993). An iterative method for a posteriori estimation of a function with discontinuities via local polynomials with adaptive bandwidth choice has been studied by Spokoiny (1998), and Polzehl and Spokoiny (2003). The basic idea is to use a classic kernel smoother with small bandwidth at the first iteration, and to increase the bandwidth locally, if the local models are not statistically significant different. This is achieved by defining new weights which penalize large values of an appropriate test statistic.

From a sequential perspective the aim is to detect changes from an assumed so-called *in-control model* (the null hypothesis) as soon as possible instead of estimating the change-point with high precision, which requires sufficient observations after the change. The data is analyzed sequentially and a signal is given, if the data provides sufficient evidence that the process is *out-of-control* (alternative hypothesis). The main tools of sequential monitoring are control charts which are given by a control statistic depending only on past and current data and a stopping time based on that control statistic. Usually, the stopping time is simply the index of the first observation where the control statistic exceeds a control limit (critical value), when calculated from current and past data. A control chart is the graphical representation of that procedure. Control charts based on nonparametric kernel estimators related to the approach presented here have been studied by Brodksy and Darkhovsky (1993, 2000), Wu (1996), Schmid and Steland (2000), and Steland (2003a, 2003b, 2003c). In these articles the (normed) delay of stopping rules of the form $\inf\{n \in \mathbb{N} : \widehat{m}_{nh} > c\}$ are studied, where $\widehat{m}_{nh}$ is a Nadarya-Watson type smoother based on data $Y_1, \dots, Y_n$ evaluated at the current time point t_n of the observation Y_n, and the delay of a stopping time is $\max\{0, T - q\}$ where T denotes the stopping time and q the change-point.

When abrupt changes as jumps or strong (nonlinear) trends appear, the smoothing property of kernel estimators can be a drawback. Therefore, in this article we consider detection rules wich are based on a class of so-called jump-preserving estimators. These estimators rely on an idea which has been developped for image processing purposes. The (deterministic) weights are substituted by stochastic jump-preserving weights. In its simplest form these weights put a clipping rectangle over the time series and average observations located in the rectangle. In this way one obtains estimators which smooth the data nonparametrically if there are no jumps, but reproduce jumps more accurately than smoothers. This idea dates back to Lee's (1983) sigma filter. The sigma filter uses the current observation to locate the clipping rectangle. For a recent discussion in the fixed-sample situation and extensions to certain M estimators we refer to Chiu et al. (1998), Winkler and Liebscher (2002), and Rue et al. (2002). The sigma filter can be regarded as a special case of the vertically weighted regression approach of Pawlak and Rafajłowicz (2000). For recent applications see Skubalska-Rafajłowicz (1994) and Kryzak, Rafajłowicz and Skubalska-Rafajłowicz (2001). Jump-preserving medians are discussed in Pawlak, Rafajłowicz and Steland (2003).

Steland (2002a) studies detection rules based on the classic sigma filter approach for independent data. Simulations indicated that under certain circumstances detection rules based on the sigma filter are better than classic methods as the EWMA procedure, if there are large jumps. Whereas the sigma filter uses the current observation to locate the window and is therefore quite wigly, in this paper we allow for more general pilot estimators as locators which may depend on an arbitrary but fixed number of past observations. By using more stable pilot estimators to locate the local window of relevant observations, we aim at decreasing the false alarms rate due to extreme observations as outliers. Our simulations indicate that the new procedure is considerably better than the classic sigma filter.

The theoretical contribution of the paper is to establish an upper bound for the normed delay. The result allows for dependent time series satisfying a conditional strong mixing properties and deals with general pilot estimators. We also allow for a certain type of local nonparametric alternatives. Further, we compare the performance of the proposed method with the sigma filter by a simulation study.

The organization of the paper is as follows. In Section 2 both the model and the method studied here are discussed in detail. Section 3 discusses the mixing conditions and general assumptions required for the main results of the paper. Our theoretical results, which deal with the tail behavior of the statistic and an upper bound for the normed delay, are presented in Section 4. We also report about a simulation study conducted to study the performance of the proposed refinement for some special cases. Finally, we illustrate the application of the proposed jump-preserving estimators by analyzing the volatility of the DEM-CZK exchange rate.

2. Data model and the jump-preserving procedure

This section explains the model framework and the proposed method for change-point detection in detail. The change-point model introduced in Section 2.1 assume a certain type of local alternatives which will provide meaningful asymptotic results. Our proposed detection rules relies on a jump-preserving estimator, whose statistical motivation is carefully explained in Section 2.2. The detection rule and related notions are introduced in Section 2.3. Strategies for the choice of the method's parameters are discussed in Section 2.4.

2.1. **Data model.** Assume we observe a local non-stationarity $\mathbb{R}$-valued process $\{Y_{th} : t \in \mathcal{T}, h > 0\}$ in continuous time $\mathcal{T} = [t_0, \infty)$ given by

$$\widetilde{Y}_{th} = m(t;h) + \widetilde{\epsilon}_t, \qquad (t \in \mathcal{T}), \tag{1}$$

where

$$m(t;h) = m_0([t-t_q]/h), \qquad t \in \mathbb{R}, h > 0,$$

for some function $m_0 : \mathbb{R} \to \mathbb{R}$ wich is continuous at 0, and $\{\widetilde{\epsilon}_t\}$ is a zero-mean stationary innovation process. We assume that $m_0(t) = 0$ for $t \leq 0$, and $m_0(t) \geq 0$ for $t \geq 0$. m_0 is called *generic alternative*. If $m_0(t) > 0$ for $t > 0$, $t_q = \inf\{s > 0 : m(s;h) > 0\}$ is called *change-point*. By continuity of m_0 at the origin, we have for each fixed t

$$m([t-t_q]/h) \to m_0(0) = 0, \qquad h \to \infty.$$

In this sense $m(t;h)$ defines a sequence of local alternatives which converge to the null model given by $H_0 : m_0 = 0$.

We assume that $\{\widetilde{Y}_t\}$ is sampled at a sequence of fixed and ordered time points $t_1 < \cdots < t_n,\ n \in \mathbb{N}$. Thus, we observe

$$Y_{nh} = m_{nh} + \epsilon_n, \qquad (n \in \mathbb{N}).$$

where $Y_n = \widetilde{Y}_{t_n}$, $m_{nh} = m_0([t_n - t_q]/h)$, and $\epsilon_n = \widetilde{\epsilon}_{t_n}$. Clearly, $q = \lfloor t_q \rfloor + 1$ is the change-point in discrete time. Throughout the paper we shall assume that $\{\epsilon_n : n \in \mathbb{N}\}$ is a stationary process in discrete time $\mathbb{N}$ with distribution F_ϵ, which satisfies a conditionally strong mixing property discussed below in detail. The mixing assumption will ensure that auto-correlations die out sufficiently fast.

Using the terminology of statistical process control, the null hypothesis $H_0 : m_0 = \mathbf{0}$ corresponds to the in-control model for the process. If $H_1 : m_0(t) > 0$ for all $t \in (t, t^*)$ for some $t^* > 0$, the out-of-control model holds and the process gets out of control starting at time q.

2.2. **Jump-preserving estimation.** Let us briefly discuss the statistical motivation of the proposed stopping rule. For that purpose we will first derive a fixed-sample jump-preserving estimator of the mean $m(t) = EY_t$. Evaluating that estimator at the current time t_n gives the control statistic.

For fixed t let $f_t(y) = f_\epsilon(y - m(t))$ denote the density function of Y_t and denote by E_t expectations w.r.t. $f_t(y),\ t \in \mathcal{T}$. Let $w(y,t)$ be a non-negative weighting function satisfying $\int\int w(y,t) y^2 f(y,t)\, dydt < \infty$. For each $t \in \mathcal{T}$ consider the weighted squared loss

$$Q_t(m) = E_t[w(t,Y)(Y - m)^2]$$

when estimating Y by a constant m. Minimizing $Q_t(m)$ w.r.t. m provides the solution

$$m^*(t) = E_t[Y w(Y,t)]/E_t[w(Y,t)]$$

which is a function of time. The relationship between the optimizer m^* and the mean function m is given by the fact that m satisfies the algebraic fix-point equation,

$$m^*(t) = E_t[Y k(Y - m^*(t))]/E_t[k(Y - m^*(t))], \tag{2}$$

provided we put

$$w(t,y) = k(y - m^*(t)) \tag{3}$$

with $k \geq 0$ being an integrable function satisfying the moment conditions

$$\int z k(z) f_\epsilon(z)\, dz = 0 \qquad \text{and} \qquad \int k(z) f_\epsilon(z)\, dz = 0, \tag{4}$$

(Pawlak and Rafajłowicz, 2000). Note that (4) holds if $k(-z) = z$ and $f_\epsilon(-z) = f_\epsilon(z)$ for all $z \in \mathbb{R}$. Thus, $m = m^*$ if the weight function w is chosen according to (3). As a consequence, an estimator for $m(t)$ can be obtained by estimating the r.h.s. of (2). Here and in the sequel we shall estimate expectations $E_t f(Y)$ by kernel smooths with respect to time,

$$\widehat{E}_t f(Y) = \sum_{i=1}^n K_h(t_i - t) f(Y_i) / \sum_{j=1}^n K_h(t_j - t),$$

where K denotes a further non-negative weighting functions (usually a density), $K_h(z) = K(z/h)/h$, and $h > 0$ is a bandwidth determining the amount of smoothing w.r.t time. This provides the estimator

$$\frac{\sum_{i=1}^n K_h(t_i - t) k_M(Y_i - m(t)) Y_i}{\sum_{j=1}^n K_h(t_j - t) k_M(Y_j - m(t))}.$$

Note that this estimator still depends on the unknown quantity $m(t)$. Since we are interested in estimating $m_n = m(t_n)$, it is quite natural to evaluate the estimator at $t = t_n$ and to use some pilot estimate $\widetilde{m}_n$ for m_n. This suggests to employ the estimator $\widehat{m}_n = \widehat{m}_{M,h,n}$,

$$\widehat{m}_{M,h,n} = \frac{\sum_{i=1}^n K_h(t_i - t_n) k_M(Y_i - \widetilde{m}_n) Y_i}{\sum_{i=1}^n K_h(t_i - t_n) k_M(Y_i - \widetilde{m}_n)}.$$

To simplify notation we shall stress the dependence on (M, h) only when necessary. Observe that this estimator can be written as a weighted average $\sum_i \widehat{w}(t_i, t_n) Y_i$ with *adaptive weights*

$$\widehat{w}(t_i, t_n) = \frac{K_h(t_i - t_n) k_M(Y_i - \widetilde{m}_n)}{\sum_j K_h(t_j - t_n) k_M(Y_j - \widetilde{m}_n)}$$

depending on the data. The choice $\widetilde{m}_n = Y_n$ provides Lee's (1983) proposal for image processing purposes, the so-called *sigma filter*, $\widehat{m}_{SF,n}$. This estimator has been studied by Godtliebsen (1991), Godtliebsen & Spjøtvoll (1991), and recently by Chiu et al. (1998). An attractive feature of the sigma filter is the fact that jumps are detected with probability 1 with no delay, provided the distribution of the error terms has bounded support (Pawlak and Rafajłowicz, 2000). Whereas M controls for the estimator's sensitivity w.r.t. jumps, the bandwidth h determines the memory of the estimator. In the sequel we shall assume that the parameter $M > 0$ is either fixed or chosen to optimize the detection procedure as described below. Asymptotic properties will be established for $h \to \infty$. In Steland (2002a) a control chart based on the sigma filter has been studied in some detail. In some situations the procedure performed badly which may be due to the large variance of $\widehat{m}_n$ when M is small. To reduce the variance in this paper we propose to employ more stable pilot estimators than the rough guess Y_n of the sigma filter. However, the initial estimator should be chosen carefully to ensure that the resulting estimator is well-defined. Indeed, for small M and kernels k with bounded support it may happen that there is no observation to which k_M assigns a positive weight. Therefore, it is reasonable to focus on preliminary estimators $\widetilde{m}_n$ for m_n ensuring the consistency condition

$$\widetilde{m}_n \in \{Y_1, \dots, Y_n\}.$$

A natural choice is

$$\widetilde{m}_n = \mathrm{Med}\{Y_n, Y_{n-1}, Y_{n-2}\},$$

providing our final proposal

$$\widehat{m}_n = \widehat{m}_{M,h,n} = \frac{\sum_{i=1}^n K_h(t_i - t_n) k_M(Y_i - \mathrm{Med}\{Y_n, Y_{n-1}, Y_{n-2}\}) Y_i}{\sum_{i=1}^n K_h(t_i - t_n) k_M(Y_i - \mathrm{Med}\{Y_n, Y_{n-1}, Y_{n-2}\})},$$

again denoted by $\widehat{m}_n$. In our simulation study and the data analysis we focus on that choice of the pilot estimator. However, the theoretical results apply to more general pilot estimators being functions of a finite number, say, l, of past observations, i.e.,

$$\widetilde{m}_n = \widetilde{m}(Y_n, Y_{n-1}, \dots, Y_{n-l}).$$

The essential point is that the pilot estimator should be very sensitive w.r.t. level changes.

Let us briefly discuss how this estimator works, in particular to understand how the estimator automatically chooses the sample size for estimation. For that purpose assume that K has support $[-1, 1]$ such that $\widehat{m}_{M,h,n}$ takes into account $Y_{n-h+1}, \dots, Y_n$, and that k is the uniform kernel. If there is no jump or trend in the data, the majority of these h observations will satisfy the constraint $k_M(Y_i - \widetilde{m}_n) = 1$, i.e., $|Y_i - \widetilde{m}_n| \leq M$. If a jump or abrupt change occurs, the distance between the pilot estimator and the observations in the time window will tend to be large. Consequently, many of the past data points are neglected and the effective sample size reduces drastically.

2.3. **Detection procedure.** To detect arbitrary deviations from the zero-mean null hypothesis (in-control model), it is natural to apply the following monitoring scheme. We provide an out-of-control signal if $|\widehat{m}_{M,h,n}|$ exceeds a prespecified control limit (critical value) c, i.e., consider a two-sided control chart. The corresponding run length is defined as

$$N_h(c, M) = \inf\{n \in \mathbb{N} : |\widehat{m}_{M,h,n}| > c\},$$

where $\inf \emptyset = \infty$. The related upper one-sided control chart to detect positive mean functions gives a signal if $\widehat{m}_{M,h,n} > c$. The normed delay is now defined as

$$\rho_h(c, M) = \frac{\max(0, N_h(c, M) - q)}{h}.$$

Another prominent performance measure is the average run length (ARL) defined as

$$ARL_F[c, M, h] = E_F[N_h(c, M)],$$

where E_F means that the expectation is calculated under a fixed distribution F for the data. Recall that for dependent processes F is determined by all finite-dimensional distribution functions.

2.4. **Choice of parameters.** To design the detection procdure one has to specify the parameters c, h, and M. Recall that the bandwidth h determines the degree of smoothing, whereas the parameter M is related to the height of a jump we want to detect immediately. Since it is reasonable to measure jumps in terms of the scale, one can choose M proportional to a (robust) scale estimate. This approach was used in the data analysis of Section 5.2. In other situations it may be advisable to choose M to optimize the detection properties at a certain target out-of-control model of interest. Indeed, our simulations support the conjecture that in many cases the performance is a concave function of M. That approach was chosen in our simulation study. Another approach is to set the critical value equal to a target level shift as in Wu (1996). For example, for a time series of prices c may be a psychological price. However, as in practical applications the average run length is often used to evaluate the performance, one can also choose c such that the procedure ensures a certain ARL as long as the in-control model (null hypothesis) holds.

3. Dependencies and general assumptions

To deal with the dependence structure of the innovation process $\{\epsilon_n\}$, and hence of $\{Y_n\}$, we shall exploit mixing conditions which are frequently used in nonparametric statistics for dependent data (cf. Fan and Gijbels, 1996, ch. 6]. Mixing conditions impose conditions on the approximation error which results when joint probabilities, say, $P[A \cap B]$, A, B events with a sufficiently large time lag, are approximated by $P[A]P[B]$. For our purposes we need a mixing property for certain conditional probabilities. To state the mixing condition we need some further notation. Assume that the $\mathbb{R}$-valued innovation sequence $\{\epsilon_n\}$ is defined on a common probability space $(\Omega, \mathcal{A}, P)$ and denote for $1 \leq i < j \leq \infty$ by

$$\mathcal{I}_i^j = \sigma(\epsilon_i, \ldots, \epsilon_j)$$

the information set (σ-algebra) of all information contained in the process $\{\epsilon_n\}$ during the time period $[t_i, t_j]$. We shall write

$$P_i^j[\,\cdot\,] = P[\,\cdot\,|\mathcal{I}_i^j] \qquad \text{and} \qquad E_i^j[\,\cdot\,] = E[\,\cdot\,|\mathcal{I}_i^j]$$

for the conditional probability and expectation, respectively, obtained by conditioning on the information $\mathcal{I}_i^j$. We write $A \in \mathcal{I}_i^j$, if A is an event only depending on the information set $\mathcal{I}_i^j$, i.e., A is determined by $Y_i, \ldots, Y_j$. For fixed $n \in \mathbb{N}$ and $l \in \mathbb{N}$ denote by

$$\mathcal{I}_i^j(n, l) = \sigma(\epsilon_k - E_{n-l}^n[\epsilon_k] : k = 1, \ldots, j), \qquad (1 \leq i < j \leq n - l),$$

the information set of the innovations $\epsilon_i, \ldots, \epsilon_j$ when centered at their conditional expectations with respect to $\mathcal{I}_{n-l}^n$.

We shall call a stationary process $\{\epsilon_n : n \in \mathbb{N}\}$ conditionally strong mixing, if $\lim_{k\to\infty} \alpha(k) = 0$, where

$$\alpha(k) = \sup_{n \geq 1} \max_{1 \leq t \leq n-l-2} \sup_{A,B} |P[A \cap B|\mathcal{I}_{n-l}^n] - P[A|\mathcal{I}_{n-l}^n]P[B|\mathcal{I}_{n-l}^n]|$$

where $\sup_{A,B}$ means that the supremum is taken over all $A \in \mathcal{I}_1^t(n, l)$ and $B \in \mathcal{I}_{t+k}^{n-l-1}(n, l)$. Hence, for any actual time t events B of the *near* future $\mathcal{I}_{t+k}^{n-l-1}(n, l)$, i.e., not depending on the whole future, are asymptotically independent from the past events $A \in \mathcal{I}_1^t(n, l)$, when conditioning on $\mathcal{I}_{n-l}^n$. The property of conditionally strong mixing will ensure that we may apply results known for strong mixing processes to some sequence which is centered at their conditional expectations w.r.t. P_{n-l}^n, when conditioning on events A measurable w.r.t. the information set $\sigma(\epsilon_{n-l}, \dots, \epsilon_n)$.

The following assumptions concern the process $\{\epsilon_n\}$ of error terms.

(A1) There is a constant $\mathcal{C}^+$ not depending on (n, l) with

$$\sup_{\nu \in \mathbb{N}} \max_{i=1,\dots,\nu-l-1} E_{\nu-l}^{\nu} \epsilon_i^+ \leq \mathcal{C}^+$$

where $z^+ = \max(0, z)$.

(A2) ϵ_n is symmetrically distributed around 0 w.r.t. P_{n-l}^n for all $n \in \mathbb{N}$.

(A3) $\{\epsilon_n\}$ is a stationary sequence of random variables satisfying a conditionally strong mixing property with

$$\lim_{k \to \infty} k^2 \alpha(k) = 0.$$

(A4) *Cramér's condition* holds for $\{\epsilon_n\}$ w.r.t. to the conditional probability P_l, i.e., there is a constant $c > 0$ such that

$$\sup_{\nu \in \mathbb{N}} E_{\nu-l}^{\nu} \sup_{i=1,\dots,n-l-1} \exp(c|Y_i|) < \infty.$$

Concerning the kernels K and k we require the following regularity conditions.

(A5) The kernel K is a differentiable density with bounded derivative, symmetric around 0, and satisfies

$$\max_{z \in \mathbb{R}} K(z) = K(0) < \infty \qquad \text{and} \qquad \int K(z)^2 \, ds < \infty.$$

(A6) The kernel k is non-negative, bounded, integrable, and symmetric around 0 with

(i) $k(z) \geq k_{\min} > 0 \; \forall z \in \mathbb{R}$.

(ii) $\max_{z \in \mathbb{R}} k(z) = k(0)$.

For simplicity of presentation we also assume

(A7) $t_\nu = \nu$ for all $\nu \in \mathbb{N}$ (equidistant time design.)

The pilot estimator is required to be a member of the following class of estimators.

(A8) The pilot estimator, $\widetilde{m}_n$, is $\mathcal{I}_{n-l}^n$-measurable for each n, i.e., we may assume $\widetilde{m}_n = \widetilde{m}_n(Y_n, \dots, Y_{n-l})$. Further, $\widetilde{m}_n \stackrel{d}{=} -\widetilde{m}_n$ if applied to a sationary process.

Remark 3.1. *Note that the pilot estimator* $\widetilde{m}_n = \mathrm{Med}\{Y_{n-2}, Y_{n-1}, Y_n\}$ *satisfies (A8).*

Finally, we need the following condition about both K and m_0.

(A9) The generic alternative $m_0 : [0, \infty) \to [0, \infty)$ satisfies

$$\int_0^x K(s-x) m_0(s)\, ds < \infty \qquad (\forall x > 0).$$

It is worth to consider an example for a time series which satisfies the conditionally strong mixing condition. Indeed, (A1)-(A4) hold true for m-dependent Gaussian processes.

Example 3.1. *Let $\{\epsilon_n : n \geq 1\}$ be an i.i.d. sequence of Gaussian innovations with common variance $\sigma^2 > 0$. Let $\{Y_n : n \geq 1\}$ be an m-dependent $\mathcal{F}_1^n$-adapted linear process generated by $\{\epsilon_n\}$, i.e., there are deterministic coefficient vectors $\boldsymbol{\vartheta}_n = (\vartheta_{n1}, \ldots, \vartheta_{nn})' \in \mathbb{R}^n$ such that*

$$Y_n = \sum_{i=1}^n \vartheta_{ni} \epsilon_i = \boldsymbol{\vartheta}_n' \boldsymbol{\epsilon}_n \qquad (n \geq 1),$$

where $\boldsymbol{\epsilon}_n = (\epsilon_1, \ldots, \epsilon_n)'$. Obviously, $\{Y_n\}$ is m-dependent iff. $\vartheta_{ni} = 0$ for all $1 \leq i < n - m$ which implies $\boldsymbol{\vartheta}_i' \boldsymbol{\vartheta}_j = 0$ whenever $|i - j| > m$. Fix $n \geq 1$. Consider $\mathbf{Y}_n = (Y_1, \ldots, Y_n)'$ and partition $\mathbf{Y}_n = (\mathbf{Y}_{n,1}', \mathbf{Y}_{n,2}')'$ with $\mathbf{Y}_{n,1} = (Y_1, \ldots, Y_{n-l-1})'$ and $\mathbf{Y}_{n,2} = (Y_{n-l}, \ldots, Y_n)'$. The joint distribution of $\mathbf{Y}_n = (Y_1, \ldots, Y_n)'$ is given by

$$\mathbf{Y}_n = \begin{bmatrix} \mathbf{S}\boldsymbol{\epsilon}_n \\ \mathbf{T}\boldsymbol{\epsilon}_n \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \begin{bmatrix} \mathbf{SS}' & \mathbf{ST}' \\ \mathbf{TS}' & \mathbf{TT}' \end{bmatrix}\right)$$

where $\mathbf{S} = (\boldsymbol{\vartheta}_1', \ldots, \boldsymbol{\vartheta}_{n-l-1}')'$ and $\mathbf{T} = (\boldsymbol{\vartheta}_{n-l}', \ldots, \boldsymbol{\vartheta}_n')'$. By m-dependence only the lower right sub-matrix of the covariance matrix does not vanish. Further, the conditional distribution, P_{n-l}^n, of $\mathbf{Y}_{n,1}$ given $\mathbf{Y}_{n,2}$ is

$$\text{(5)} \qquad \mathbf{Y}_{n,1} | \mathbf{Y}_{n,2} \sim \mathcal{N}(\mathbf{ST}'(\mathbf{TT})^{-1}\mathbf{Y}_{n,2}, \sigma^2[\mathbf{SS}' - \mathbf{ST}'(\mathbf{TT}')^{-1}\mathbf{TS}'])$$

Since $\mathbf{ST}'$ is the $(n-l-1) \times (l+1)$-matrix with entries $\boldsymbol{\vartheta}_i' \boldsymbol{\vartheta}_j$, $1 \leq i \leq n-l-1$, $n-l \leq j \leq n$, by m-dependence only the lower right sub-matrix does not vanish, and therefore conditioning on $\mathbf{Y}_{n,2}$ only alters the corresponding lower right sub-matrix of $\mathbf{SS}'$, the covariance matrix of $\mathbf{Y}_{n,1}$. Thus, the sequence $\mathbf{Y}_{n-1} - E_{n-l}^n[\mathbf{Y}_{n-1} | \mathbf{Y}_{n-2}]$, which is centered at its conditional expectation, is m-dependent, and the conditionally strong mixing property follows. (5) also shows symmetry around 0, and existence of all moments.

4. Main results

In this section we provide the main results of this article. We first study the tail behavior of the control statistic $\widehat{m}_n$ under the in-control model (null hypothesis) of a stationary conditional strong mixing process. The result is interesting in its own right and is needed to establish an upper bound for the normed delay of the stopping time $N_h(c, M)$. That upper bound makes an assertion about the normed delay under the out-of-control model of a local nonparametric alternative. Considering the normed delay means that the delay is expressed as a percentage of the bandwidth parameter h. This provides the nice interpretation that for large h the delay is not greater than the upper bound - usually a number between 0 and 1 - times h with probability tending to 1. In particular if the support of K is $[-1,1]$, we get an impression after how many observations the deviation in the mean will be detected with high probability. Similar results, namely a.s. convergence, have been obtained for classic kernel smoothers by Brodsky and Darkhovsky (1993) assuming a simple level shift and have been extended to local nonparametric alternatives as considered here by Steland (2003b).

It turns out that the upper bound depends on the *kernel ratio* of k which is defined as

$$R_k = \sup_{z \in \mathbb{R}} k(z) / \inf_{z \in \mathbb{R}} k(z).$$

4.1. **Tail behavior.** The following result deals with the tails of the distribution of $\widehat{m}_n$ It makes a statement about the time-point-wise false-alarm rate of the corresponding detection rules, since the events $\{\widehat{m}_n > c\}$ and $\{|\widehat{m}_n| > c\}$ correspond to false-alarms if the process is in control, i.e., $m_0 = \mathbf{0}$. Furthermore, the result is needed to establish the upper bound for the normed delay. In addition to the statistic $\widehat{m}_n$ let us also also consider the version, $\breve{m}_n = \breve{m}_{M,h,n}$, without norming the weights, i.e.,

$$\breve{m}_n = \sum_{i=1}^{n} K_h(t_i - t_n) k_M(Y_i - \widetilde{m}_n) Y_i.$$

Theorem 4.1. *Assume (A1)-(A8) and* $m_0 = \mathbf{0}$*. If in addition*

$$n/h \to \zeta > 0 \qquad \text{as } n, h \to \infty,$$

then the following assertions hold true.

(i) *There exist constants* $b_1, b_2 > 0$ *with*

$$P[\widehat{m}_n > c], P[|\widehat{m}_n| > c] \;=\; O(h^{1/2} \exp(-b_1 \cdot h^{1/2})) + O(h\alpha(h^{1/2})) + O(e^{-b_s h})$$

for every candidate control limit $c > (l+1)(R_k - 1)\mathcal{C}^+/M^2$.

(ii) *There exist constants* $b_1, b_2 > 0$ *with*

$$P[\breve{m}_{R,n} > c], P[|\breve{m}_{R,n}| > c] \;=\; O(h^{1/2} \exp(-B \cdot h^{1/2})) + O(h\alpha(h^{1/2})) + O(e^{-b_2 h})$$

for each candidate control limit c *satisfying*

$$c > \kappa(R_k, M) = [k(0) - k_{\min}]\mathcal{C}^+/(2M^2),$$

where $I(\zeta) = \int_0^\zeta K(s)\, ds$.

A Borel-Cantelli argument yields the following corollary.

Corollary 4.1. *If* $\sum_k k^2\alpha(k) < \infty$*, then under the conditions of Theorem 4.1 for each* $c > (l+1)(R_k - 1)\mathcal{C}^+/M^2$ *we have*

$$\sum_n P[\widehat{m}_n > c], \sum_n P[|\widehat{m}_n| > x] < \infty$$

which implies

$$P[\widehat{m}_n > x,\ i.o.] = 0, \qquad P[|\widehat{m}_n| > x,\ i.o.] = 0.$$

4.2. **Upper bound for the normed delay.** We shall now establish an asymptotic bound for the normed delay, $\rho_h = \rho_h(c, M) = h^{-1}\max\{0, N_h - q\}$, as $h \to \infty$. Whereas for classic Nadaraya-Watson type kernel weights a.s. convergence to a deterministic function depending on the smoothing kernel and the generic alternative m_0 can be shown, for jump-preserving weights the situation is more delicate. However, an upper bound can be established.

Theorem 4.2. *Assume (A1)-(A8) and* $t_q = 1$*. Let* c *be some fixed control limit. Suppose* ρ_0 *satisfies the nonlinear equation*

$$\rho_0 = \inf\left\{\rho > 0 : \int_0^\rho K(s-\rho)m_0(s)\,ds = cR_k/2 + (R_k - 1)\mathcal{C}^+/(2M)\right\}$$

Then ρ_0 *is an asymptotic upper bound for* ρ_h *in the sense that*

$$P[\rho_h > \rho_0] = o(1),$$

as $h \to \infty$*.*

Again, the following corollary is straightforward.

Corollary 4.2. *In addition to the assumptions of Theorem 4.2 suppose that* $\sum_k k^2\alpha(k) < \infty$*. Then*

$$P[\rho_h > \rho_0,\ i.o.] = 0.$$

Remark 4.1. *The bound may be not as sharp as possible, but for the case considered here better bounds are not known for error distributions with infinite support.*

Remark 4.2. *The upper bound depends on the smoothing kernel* K *(w.r.t time) and the generic alternative* m_0 *only the through the function*

$$I(\rho; K, m_0) = \int_0^\rho K(s-\rho)m_0(s)\,ds.$$

For optimization of the functional $\tau(K) = \inf\{\rho > 0 : I(\rho; K, m_0) \geq c\}$ *w.r.t.* K*, we refer to Steland (2003b).*

4.3. **Proofs.** We will now provide the proofs of both theorems.

Proof (of Theorem 4.1). First note that by symmetry of k_M and since $\widetilde{m}_l \overset{d}{=} -\widetilde{m}_l$, we have for each $x \in \mathbb{R}$

$$\begin{aligned} P(\widehat{m}_n \le -x) &= P[\frac{\sum_i K_h(t_i - t_n) k_M(Y_i - \widetilde{m}_l) Y_i}{\sum_i K_h(t_i - t_n) k_M(Y_i - \widetilde{m}_l)} \le -x] \\ &= P[\frac{\sum_i K_h(t_i - t_n) k_M(\widetilde{m}_l - Y_i)(-Y_i)}{\sum_i K_h(t_i - t_n) k_M(\widetilde{m}_l - Y_i)} \le -x] \\ &= P[\widehat{m}_n > x]. \end{aligned}$$

Consequently, $P(|\widehat{m}_n| > c) = 2P(\widehat{m}_n > c)$. Assumption (A7) ensures that

$$\sum_{i=1}^{n} K_h(t_i - t_n) = I(\zeta) + O(1/h)$$

where $I(\zeta) = \int_0^\zeta K(s)\, ds$. Thus,

$$\sum_{i=1}^{n} K_h(t_i - t_n) k_M(Y_i - \widetilde{m}_n) \ge (k_{\min}/M)\{I(\zeta) + O(1/h)\}.$$

Using this lower bound and recalling that l denotes the number of lagged observations on which $\widetilde{m}_n$ depends, we obtain

$$\begin{aligned} & P[\widehat{m}_n > c] \\ (6) \qquad & \le P\left[\sum_{i=1}^{n-l-1} K([t_i - t_n]/h) k([Y_i - \widetilde{m}_n]/M) Y_i > \frac{c k_{\min}}{l+1}\{I(\zeta) + O(1/h)\} \cdot h\right] \\ & + (l+1) \max_{j=0,\ldots,l} P\left[Y_{n-j} > c \frac{1}{l+1} \frac{k_{\min}}{K(0)k(0)}\{I(\zeta) + O(1/h)\} \cdot h\right]. \end{aligned}$$

By Cramér's condition the last term can be bounded by $b_1 \exp(-b_2 h)$ for some constants $b_1, b_2 > 0$. Thus, it remains to provide a similar bound for the first term. Denote by $F_{nl}(y)$ the distribution function of $\widetilde{m}_n$. By conditioning on $\mathcal{I}_{n-l}^n = \sigma(Y_{n-l}, \ldots, Y_n)$ the first term can be written as

$$(7) \qquad \int P_{n-l}^n \left[\sum_{i=1}^{n-l-1} K([t_i - t_n]/h) k([Y_i - z]/M) Y_i > \frac{c k_{\min}}{l+1}\{I(\zeta) + O(1/h)\} h\right] dF_{nl}(z).$$

The same argument provides

$$\begin{aligned} (8) \qquad & P_{n-l}^n[\breve{m}_n > c] \le \\ & \int P_{n-l}^n \left[\sum_{i=1}^{n-l-1} K([t_i - t_n]/h) k([Y_i - z]/M) Y_i > chM\right] dF_{nl}(z) \\ & + b_1' \exp(-b_2' h) \end{aligned}$$

for some constants $b_1', b_2' > 0$. Define

$$S_n(z) = \sum_{i=1}^{n-l-1} K([t_i - t_n]/h)\xi_i(z)$$

where $\xi_i(z) = k([Y_i - z]/M)Y_i$, $i = 1, \ldots, n-l-1$. Note that all conditional moments (under P_{n-l}^n) of $\xi_\nu(z)$ are uniformly bounded in $z \in \mathbb{R}$ and $\nu \in \mathbb{N}$. Having in mind (7) and (8), it is sufficient to show

$$\int P_{n-l}^n[S_n(z) > xh]\, dF_{nl}(z) = O(h^{1/2}\exp(-b_1 h^{1/2})) + O(h\alpha(h^{1/2})) \tag{9}$$

for some constant $b_1 > 0$.

To verify (10) we will use a blocking argument. Fix $0 < \gamma < 1$. Recall that $n/h \sim \zeta$ and partition the set $\{1, \ldots, n\}$ in blocks of length $p_h = \lfloor (\zeta h)^{1/2}\gamma \rfloor$. Then there are $m_h = \lfloor (n-l-1)/p_h \rfloor \sim (\zeta h)^{1/2}/\gamma$ blocks. The first step is to show

$$P[S_n(z) - E_{n-l}^n S_n(z) > xh] = O(h^{1/2}\exp(-b_1 h^{1/2})) + O(h\alpha(h^{1/2})). \tag{10}$$

where the O does not depend on $z \in \mathbb{R}$. The proof will then be completed by estimating the conditional mean $ES_{n-l}^n(z)$.

Split the sum $S_n(z)$ in p_h partial sums, each consisting of m_h summands taken from the m_h blocks, and a remainder term. This means,

$$\begin{aligned} S_n(z) &= \sum_{r=1}^{p_h} S_n^{(r)}(z) + R_n(z), \\ S_n^{(r)}(z) &= \sum_{k=1}^{m_h} \xi_{km_h+r}(z), \qquad (r = 1, \ldots, p_h), \end{aligned}$$

and $R_n(z) = \sum_{i=p_h m_h+1}^{n-l-1} \xi_i(z)$. First note that a similar argument as used in (7) ensures that $P[R_n(z) > x] = O(p_h \exp(-b_3 h))$ for some constant $b_3 > 0$, since the number of summands of $R_n(z)$ is not larger than p_h. Thus, w.l.o.g. assume $p_h m_h = n - l - 1$. Next observe that for each $x > 0$

$$P_{n-l}^n[S_n(z) - E_{n-l}^n S_n(z) > xh] \le \sum_{r=1}^{p_h} P_{n-l}^n[S_n^{(r)}(z) - E_{n-l}^n S_n^{(r)}(z) > xh/p_h].$$

We shall show

$$\max_{1 \le r \le p_h} \sup_{z \in \mathbb{R}} P_{n-l}^n[S_n^{(r)}(z) - E_{n-l}^n S_n^{(r)}(z) > xh/p_h] = O(\exp(-b_1 h^{1/2}))$$

for some constant $b_1 > 0$ yielding (10), since $m_h = O(h^{1/2})$.

Using Markov's inequality it follows that the absolute value of the difference between

$$P_{n-l}^n[S_n^{(r)} - E_{n-l}^n S_n^{(r)}(z) > xh/p_h]$$

and

$$\exp(-txh/p_h)\prod_{k=1}^{m} E_{n-l}^n \exp(tK([t_{km_h+r}-t_n]/h)[\xi_{km_h+r}(z)-E_{n-l}^n\xi_{km_h+r}(z)])$$

is not larger than $16(m_h-1)\alpha(p_h) = O(\alpha^{1/2}\alpha(h^{1/2}) = o(1)$, since $\{\xi_\nu(z)\}$ is conditionally strong mixing and satisfies (A3). By strict stationarity of $\{\xi_\nu(z)\}$

$$g_0 := \sup_{z\in\mathbb{R}}\sup_{\nu\in\mathbb{N}}\max_{i=1,\dots,\nu} E_{\nu-l}^\nu \xi_i^2(z) < \infty$$

Thus, for all $|t|\le T$ and $g > g_0$ we have

$$E_{n-l}^n \exp(tK([t_i-t_n]/h)\xi_i(z) \le \exp(K([t_i-t_n]/h)gt^2/2), \qquad (i=1,\dots,n; n\in\mathbb{N}).$$

Hence

$$\begin{aligned}&\prod_{k=1}^{p} E_{n-l}^n \exp(tK([t_{km_h+r}-t_n]/h)[\xi_{km_h+r}-E_{n-l}^n\xi_{km_h+r}(z)])\exp(txh/p_h)\\ &\quad\le \exp(K_r(p_h)gt^2/(2p_h)-txh/p_h),\end{aligned}$$

where

$$K_r(p_h) = \sum_{k=1}^{p_h} K([t_{km_h+r}-t_n]/h)^2.$$

By minimizing the function $t\mapsto K_r(p_h)gt^2/2 - txh/p_h$, we obtain

$$\begin{aligned}&P_{n-l}^n[S_n^{(r)}(z)-E_{n-l}^n S_n^{(r)}(z) > xh/p_h]\\ &\quad= \begin{cases} O(\exp(-(xh/p_h)^2/[2gK_r(p_h)]), & (xh/p_h)\le gTK_r(p_h)\\ O(\exp(-(xh/p_h)T/2), & (xh/p_h) > gTK_r(p_h)\end{cases}\end{aligned}$$

We have to study $K_r(p_h)$ in detail. Obviously, $\{(t_{km_h+r}-t_n)/h = (n-km_h-r)/h : k = 1,\dots,p_h\}$ is an equidistant partition of $[(l+1-r)/h,(n-m_h-r)/h]$ with associated size $m_h/h = O(h^{-1/2})$. Hence,

$$\left|(m_h/h)\sum_{k=1}^{p_h} K^2([t_{km_h+r}-t_n]/h) - \int_0^\zeta K^2(s)\,ds\right| = O(m_h/h).$$

Consequently, $h/(K_r(p_h)m_h)$ is bounded away from 0. Note that

$$\frac{(xh/p_h)^2}{2gK_r(p_h)p_h^2} = \frac{x^2}{2g}\frac{h}{K_r(p_h)m_h}\frac{m_h h}{p_h^2} \ge d_1 h^{1/2}$$

and

$$\frac{xhT}{2p_h} \ge \frac{\eta' T}{2}\frac{h}{p_h} \ge d_2 h^{1/2}$$

for constants $d_1, d_2 > 0$. Putting things together we see that with $b_1 = \max(d_1,d_2)$

(11) $$\max_r \sup_{z\in\mathbb{R}} P_{n-l}^n[S_n^{(r)}(z)-E_{n-l}^n S_n^{(r)}(z) > xh/p_h] = O(\exp(-b_1h^{1/2})) + O(h^{1/2}\alpha(h^{1/2})).$$

A similar bound can be obtained for $P^n_{n-l}[-[S^{(r)}_n(z) - E^n_{n-l}S^{(r)}_n(z)] > xh/p_h]$.

It remains to estimate $E^n_{n-l}S_n(z)$ uniformly in $z \in \mathbb{R}$. Using symmetry of $Y_\nu,\ n \in \mathbb{N}$, one can show that

$$
\begin{aligned}
(12)\qquad E^n_{n-l}S_n(z) &\geq h[k_{\min} - k(0)] \sup_{\nu\in\mathbb{N}} \max_i E^\nu_{\nu-l} Y_i^+ \{I(\zeta) + O(1/h)\},\\
(13)\qquad E^n_{n-l}S_n(z) &\leq h[k(0) - k_{\min}] \sup_{\nu\in\mathbb{N}} \max_i E^\nu_{\nu-l} Y_i^+ \{I(\zeta) + O(1/h)\},
\end{aligned}
$$

uniformly in $z \in \mathbb{R}$, where the maximum is taken over $i = 1, \dots, n-l-1$. Using (13) we obtain

$$
\begin{aligned}
&\int P^n_{n-l}[S_n(z) > xh]\, dF_{nl}(z)]\\
&\quad = \int P^n_{n-l}[S_n(z) - E^n_{n-l}[S_n(z)] > xhM - E^n_{n-l}[S_n(z)]]\, dF_{nl}(z)\\
&\quad \leq \int P^n_{n-l}[S_n(z) - E^n_{n-l}[S_n(z)] > (x - \mu'_\Sigma + O(1/h)) \cdot Mh]\, F_{nl}(z)\\
&= O(h^{1/2}e^{-b_1 h^{1/2}}) + O(h\alpha(h^{1/2})),
\end{aligned}
$$

where $\mu'_\Sigma = I(\zeta)[k(0) - k_{\min}]\mathcal{C}^+/M$. Applying this estimate with $x = c(l+1)^{-1}k_{\min}M^{-1}I(\zeta)$ to (7) we see that

$$
P(\widehat{m}_n > c) = O(h^{1/2}e^{-b_1 h}) + O(h^{1/2}\alpha(h)) + O(e^{-b_2}h),
$$

if $c > (l+1)(R_k - 1)\mathcal{C}^+/M^2$. Similarly, applying the estimate with $x = cM$ to (8) yields the result for $\breve{m}_n$ if $c > \kappa(R_k, M) \geq I(\zeta)[k(0) - k_{\min}]\mathcal{C}^+/M^2$, since $I(\zeta) \in [0, 1/2]$ for all $\zeta > 0$.

We are now in a position to verify Theorem 4.2.

Proof (of Theorem 4.2). Note that by definition of ρ_h and N_h we have for each $\varepsilon > 0$

$$
\{\rho_h - \rho_0 > \varepsilon\} \subset \{|\widehat{m}_{\lfloor(\rho_0+\varepsilon)h\rfloor}| \leq c\}.
$$

Put $n(h) = \lfloor(\rho_0 + \varepsilon)h\rfloor$ and recall that $Y_{hi} = \epsilon_i$ if $i = 1, \dots, q-1$ and $Y_{hi} = m(t_i; h) + \epsilon_i$ if $i = q, \dots, n(h)$. We have

$$
\begin{aligned}
& P(\rho_h - \rho_0 > \varepsilon) \\
& \leq P(|\widehat{m}_{n(h)}| \leq c) \\
& \leq P\Big(|\sum_i K_h(t_i - t_{n(h)}) k_M(Y_i - \widetilde{m}_{n(h)})| \leq c \frac{k(0)}{M} \sum_i K_h(t_i - t_{n(h)}) \Big) \\
& \leq P\Big(\Big| \sum_{i=1}^{n(h)} K_h(t_i - t_{n(h)}) k_M(Y_i - \widetilde{m}_{n(h)}) \epsilon_i \Big| > \\
& \sum_{i=q}^{n(h)} K_h(t_i - t_{n(h)}) k_M(Y_i - \widetilde{m}_{n(h)}) m(t_i; h) - \frac{ck(0)}{M} (1/2 + O(h^{-1})) \Big).
\end{aligned}
$$

Here we used the fact that for $y \geq 0$

$$
|x + y| \leq z \;\Leftrightarrow\; |y| = y \geq |x| - z, \; y \leq z - x, \; x, y, z \in \mathbb{R},
$$

and the estimate

$$
\sum_{i=1}^{n(h)} K_h(t_i - t_{n(h)}) k_M(Y_i - Y_{n(h)}) \leq (k(0)/M) \{1/2 + O(1/h)\}.
$$

Note that, since $q/h = o(1)$,

$$
\begin{aligned}
& \sum_{i=q}^{n(h)} K_h(t_i - t_n) k_M(Y_i - \widetilde{m}_{n(h)}) m(t_i; h) \\
& \qquad \geq (k_{\min}/M) \sum_{i=1}^{n(h)} K_h(t_i - t_{n(h)}) m(t_i; h) \\
& \qquad = \frac{k_{\min}}{M} \int_0^{\rho_0 + \varepsilon} K(s - \rho_0 - \varepsilon) m_0(s) \, ds + O(1/h) \\
& \qquad = \frac{k_{\min}}{M} \int_0^{\rho_0} K(s - \rho_0) m_0(s) \, ds + O(\varepsilon) + O(1/h).
\end{aligned}
$$

Thus, if

$$
-\frac{ck(0)}{2M} + \frac{k_{\min}}{M} \int_0^{\rho_0} K(s - \rho_0) m_0(s) \, ds = \kappa(R_k, M),
$$

which is guaranteed by the choice of ρ_0, we obtain that

$$
P(\rho_h - \rho_0 > \varepsilon)
$$

is not greater than

$$
P\Big(\Big| \sum_{i=1}^{n(h)} K_h(t_i - t_n) k_M(Y_i - \widetilde{m}_{n(h)}) \epsilon_i \Big| \geq \kappa(R_k, M) + \{O(\varepsilon) + O(1/h)\} \Big). \tag{14}
$$

Therefore, an application of Theorem 4.1 (ii) with $\zeta = \rho_0 + \varepsilon$ yields

$$P(\rho_h - \rho_0 > \varepsilon) = o(1),$$

because the sum in (14) equals the estimator $\widetilde{m}_{n(h)}$ applied to the sample $\epsilon_1, \dots, \epsilon_{n(h)}$.

5. Simulations and data example

To shed some light onto the properties of the jump-preserving monitoring procedure with median-based pilot estimation, we performed simulations for i.i.d. data. Since financial time series as the exchange rate series analyzed below, are often affected by conditional heteroscedasticity, we also considered GARCH models for both the simulation and the data analysis. Although GARCH processes are known to be strongly mixing with exponential rate (Basrak, Davis and Mikosch (2002); Carrasco and Chen (2002)), it is not clear whether they satisfy the conditional mixing condition required here. However, neither the simulation nor the data analysis requires the theoretical bound on the normed delay. The simulation study is designed to yield an implicit comparison with the classic EWMA control chart.

5.1. **Simulations.** We conducted a simulation study which was devoted to compare the classical sigma filter with the median-based improvement proposed here. Concerning the mean structure the focus is on deterministic peak-like deviations from mean-stationarity. The innovation process was modeled as (i) Gaussian white noise and, to take account of the fact that financial time series are often affected by conditional heteroscedasticity, as (ii) a GARCH(1,1) process. For better comparisons the same simulation model as in Steland (2002a) was used where time series are generated according to the model

$$Y_n = a \cdot \mathbf{1}(q \leq t_n < q + s) + \epsilon_n, \qquad (n \geq -39),$$

with $t_\nu = \nu$ for all $\nu \in \mathbb{N}$ and stationary zero-mean innovations $\{\epsilon_n\}$. q stands for the first change-point in discrete time and $q + s$ for the second one. Small values for s correspond to peaks whereas for $s \to \infty$ the classical change-point model is obtained.

To study time series with conditional heteroscedasticity a GARCH(1,1) model (Bollerslev, 1986; Engle & Bollerslev, 1986) given by

$$\epsilon_n = h_n \eta_n, \qquad with \qquad h_n^2 = \alpha_0 + \alpha_1 h_{n-1}^2 + \beta_1 \epsilon_{n-1}^2$$

for $n \geq 2$ and $h_1 = \alpha_0/(1 - \alpha_1 - \beta_1)$ was used. Here α_0, α_1, and β_1 are parameters specified by $\alpha_0 = \alpha_1 = .1$ and $\beta_1 = .85$.

We considered the improvement for various values of the parameters h, M, and a for a level shift of $s = 3$ periods. Further, for various values of s and h we simulated the minimal out-of-control ARLs where minimization was done over a finite set of M-values. M was chosen from the set $\mathcal{M} = \{.5, 1, 1.5, 2, 2.5, 3\}$ for the first setting and from $\mathcal{M} = \{.5, 1, 1.5, 2.5, 3, 3.5, 4\}$ for the second one.

The median-based modification of the sigma filter, $\widehat{m}_n$, was specified as follows. The kernel k was chosen as an uniform kernel whereas the kernel K which is used to smooth the data w.r.t. time was chosen as a Laplace density. In this case the weights $\widehat{w}(t_i, t_n)$ converge to the weights $\lambda(1-\lambda)^{n-i}$ of the EWMA control chart given by the recursion $Z_{n+1} = (1-\lambda)Z_n + \lambda Y_{n+1}$, $Z_0 = 0$, if the bandwidht h and the smoothing parameter of the EWMA chart are related by

$$h_\lambda = -\sqrt{2}/\log(1-\lambda), \qquad \lambda \in (0, 1],$$

and n tends to ∞. h_λ is called equivalent bandwidth. In this sense, the control chart based on $\widehat{m}_n$ provides for large M an approximation to the EWMA chart. Thus, if for a data constellation small to moderate values of M are better than large ones, the jump-preserving procedure outperforms the (approximation to the) EWMA. For this reason, we used equivalent bandwidths $h = h(\lambda)$, $\lambda \in \{.02, .04, .06, .08, .1, .2\}$, which translates to bandwidths ranging from 6.34 to 70. The critical values to ensure an in-control ARL of $\xi = 20$ and each out-of-control was estimated by a simulation using approximately 50,000 repetitions. We used an automatic algorithm to estimate the necessary number of repetitions providing estimates ranging from 5,000 to 100,000 (cut-off). We used a small in-control ARL, but our experiments indicate that the results do not depend qualitatively on ξ. To ensure that the control charts have sufficient past data for all values of h, we used a pre-run of 40 time units. Otherwise, for moderate to large values of h the variance of the control statistic would be rather large when monitoring starts. Each generated time series was truncated at $n_{\max} = 10,000$.

The results for Gaussian white noise and GARCH(1,1) innovations corresponding to the case $s = 3$, are given in Table 1. For each (a, h, M, λ)-combination the ratio of the ARL of the sigma filter divided by the ARL of the median-based improvement is given. Further, in the last column the performances of both control charts obtained by (first) optimizing over $M \in \mathcal{M}$ are shown. For brevity, we provide the results for $\lambda = 0.02, 0.04$, and 0.1 which give a sufficient impression. The results indicate that there is a considerable improvement for small, moderate, and large values of M and h, respectively.

Table 2 provides out-of-control ARLs of the sigma filter with median-based pilot estimation. The corresponding maximizing value of M is given in brackets. We provide the results for $\lambda = 0.02, 0.04$, and 0.1. It can be seen that even small jumps lasting only for short periods

λ	a	M						
		0.5	1.0	1.5	2.0	2.5	3.0	M_{opt}
				Gaussian white noise				
0.02	0.5	1.03	1.06	1.08	1.09	1.08	1.01	1.051
0.02	1	1.14	1.15	1.16	1.19	1.16	1.05	1.126
0.02	2.5	9.12	8.42	7.57	8.16	7.29	4.94	7.932
0.04	0.5	1.03	1.09	1.07	1.07	1.05	1.03	1.026
0.04	1	1.13	1.18	1.18	1.16	1.08	1.01	1.072
0.04	2.5	8.65	8.94	7.86	7.51	6.81	4.96	7.687
0.10	0.5	1.14	1.05	1.10	1.07	1.06	1.15	1.061
0.10	1	1.28	1.19	1.21	1.17	1.07	1.16	1.101
0.10	2.5	9.82	9.01	7.99	8.39	6.08	5.36	7.909
				Garch innovations				
0.02	0.5	0.99	1.03	1.03	1.05	1.07	1.04	1.052
0.02	1	1.02	1.08	1.09	1.10	1.12	1.07	1.096
0.02	2.5	2.56	2.64	2.86	2.84	2.92	2.68	2.883
0.04	0.5	1.05	1.03	1.04	1.03	1.05	1.05	1.046
0.04	1	1.09	1.09	1.09	1.10	1.11	1.09	1.091
0.04	2.5	2.78	2.79	2.91	2.93	2.84	2.63	2.753
0.10	0.5	1.03	1.04	1.03	1.06	1.06	1.06	1.058
0.10	1	1.09	1.11	1.08	1.14	1.12	1.11	1.107
0.10	2.5	2.76	2.93	2.83	2.96	3.07	2.59	2.911

TABLE 1. *Improvement of the control chart based on a sigma filter with median-based pilot estimation in terms of the out-of-control average run length when compared with the corresponding control chart based on a classical sigma filter, expressed as a ratio of ARLs. First change-point* $t_q = 40$*, second change-point* $t_q + s = 43$.

can be detected soon. Comparing the table entries with Table 3 of Steland (2002a), we see that the ARLs are considerable smaller. For example, the sigma filter detects a unit shift for two periods ($s = 2$) in Gaussian random noise after 8.93 periods on average, whereas the median-based modification detects it after 4.36 periods on average. There is also a slight tendency to favor smaller values of M.

We may summarize that the sigma filter can be improved considerably by median-based pilot estimation.

5.2. **A financial application.** To illustrate the application of the jump-preserving estimators studied in this paper we provide an empirical analysis of the DEM-CZK exchange rate for the period from 03/10/96 to 20/04/98. The time period from 02/11/95 to 02/10/96 was used to fit an in-control model and to design the control chart. We analyzed the volatility of the return series $\{R_t\}$. Fitting a GARCH(1,1) gives $\widehat{\alpha}_0 = .000003341$, $\widehat{\alpha}_1 = .409777$, and

λ	a	s									
		1		2		4		8		64	
					Gaussian white noise						
0.02	0.5	9.27	[3.00]	8.05	[3.00]	6.27	[3.00]	4.21	[3.00]	2.17	[3.50]
	1.0	6.19	[3.00]	4.36	[3.00]	2.36	[3.00]	1.06	[3.00]	0.82	[3.50]
	2.5	0.93	[3.00]	0.15	[3.00]	0.06	[3.00]	0.05	[3.00]	0.05	[3.00]
0.04	0.5	5.01	[3.50]	4.18	[3.50]	3.15	[3.50]	2.09	[3.50]	1.48	[3.50]
	1.0	3.19	[3.50]	2.15	[3.50]	1.09	[3.50]	0.59	[3.50]	0.54	[3.50]
	2.5	0.40	[3.50]	0.08	[3.00]	0.04	[3.50]	0.04	[3.50]	0.04	[3.50]
0.10	0.5	2.27	[3.50]	1.83	[3.50]	1.35	[3.50]	0.99	[3.50]	0.89	[3.50]
	1.0	1.19	[3.50]	0.71	[3.50]	0.39	[3.50]	0.30	[3.50]	0.29	[3.50]
	2.5	0.08	[4.00]	0.02	[3.50]	0.01	[3.50]	0.01	[3.50]	0.01	[4.00]
					Garch innovations						
0.02	0.5	9.80	[4.00]	8.84	[4.00]	7.38	[4.00]	5.53	[4.00]	2.99	[4.00]
	1.0	7.54	[4.00]	5.84	[4.00]	3.78	[4.00]	2.02	[4.00]	1.28	[4.00]
	2.5	2.23	[4.00]	0.88	[4.00]	0.30	[4.00]	0.19	[4.00]	0.20	[4.00]
0.04	0.5	6.45	[4.00]	5.63	[4.00]	4.75	[4.00]	3.47	[4.00]	2.32	[4.00]
	1.0	4.65	[4.00]	3.55	[4.00]	2.23	[4.00]	1.27	[4.00]	1.00	[4.00]
	2.5	1.20	[4.00]	0.43	[4.00]	0.18	[4.00]	0.15	[4.00]	0.15	[4.00]
0.10	0.5	3.52	[4.00]	3.06	[4.00]	2.46	[4.00]	1.86	[4.00]	1.58	[4.00]
	1.0	2.29	[4.00]	1.63	[4.00]	1.00	[4.00]	0.72	[4.00]	0.67	[4.00]
	2.5	0.46	[4.00]	0.18	[4.00]	0.10	[4.00]	0.09	[4.00]	0.09	[4.00]

TABLE 2. *Average run lengths of the sigma filter with median-based pilot estimation for detecting peaks of varying length.*

$\widehat{\beta}_1 = .131462$ with $SBC = -2155.99$. A common approach is to measure volatility in terms of the empirical standard deviation $s_t = [(1/19)\sum_{i=0}^{19}(R_{t-i} - \overline{R}_t)^2]^{1/2}$, $\overline{R}_t = (1/20)\sum_{i=0}^{19} R_{t-i}$, computed for the last 20 trading days. It is known that s_t tends to produce artifacts when isolated (short periods of) trading days with extreme changes are present, since each trading day located in the estimation window has the same weight. Consequently, 'outlying' returns may dominate and the volatility is over-estimated for the next 20 trading days. Alternatively, one may apply a sigma filter with crude or median-based pilot estimation computed from the sequence of squared returns. To allow comparisons we used uniform kernels and $h = 20$. The parameter M was chosen according to the following rule of thumb. Relying on the quartile distance QD of $|R_t|$ to measure dispersion of the in-control period, we put $M = QD/8$. This gives $M = 0.000269$. Upper control limits for one-sided upper control charts corresponding to an in-control ARL of 60 were calculated by simulating from the estimated in-control GARCH model. Figure 1 provides the results. The picture suggests that the jump-preserving volatility

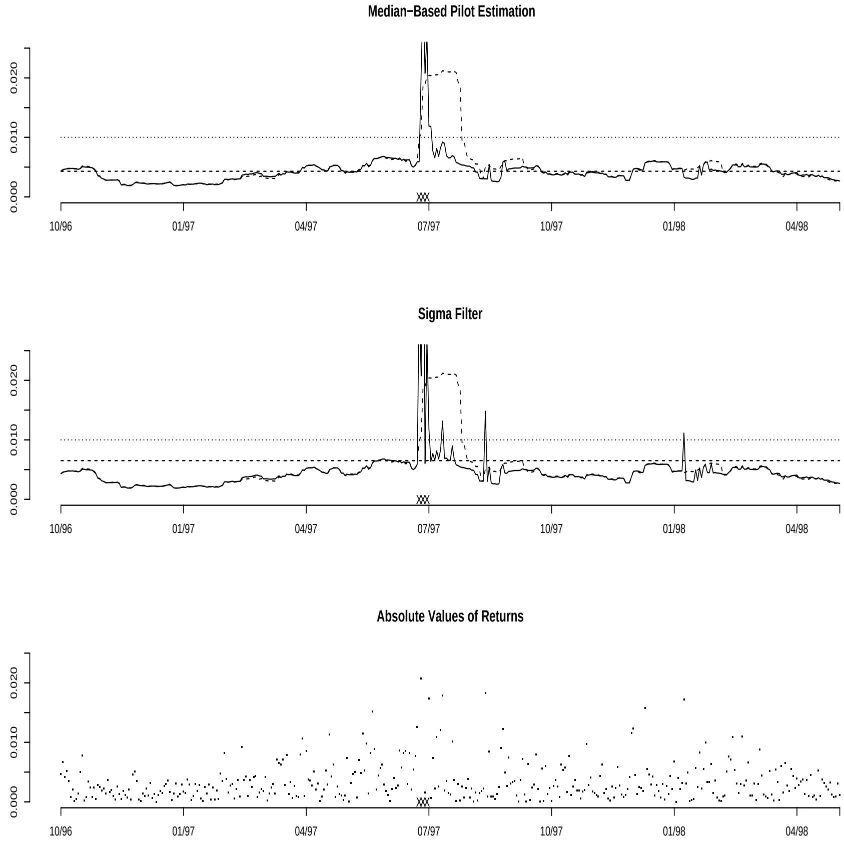

Figure 1. *DEM-CZK FX Rates.* Jump-preserving volatility estimator with median-based pilot estimation (top panel), sigma filter, and absolute values of daily FX returns (bottom panel). Superimposed (dotted lines) is the classical empirical standard deviation of the returns (moving window estimator with 20 observations.) Outlying returns are marked (X.) Control limits and a control limit corresponding to a 1% change are added.

estimates behave similar as s_t when volatility is stable and smooth, but nicely reproduce jump-like changes without producing artifacts as s_t. The jump-preserving monitoring procedures react quickly to increases. Note that the median-based modification indicates the increase of volatility much more earlier than the sigma filter whose control limit is considerably higher.

Acknowledgements

The author thanks an anonymous referee for helpful remarks and is grateful to Prof. E. Rafajłowicz, Polytechnika Wrocław, Poland, and Prof. M. Pawlak, University of Manitoba, Canada, for helpful discussions. The financial support of Deutsche Forschungsgemeinschaft (SFB 475, ,,Reduction of complexity in multivariate data structures") is gratefully acknowledged.

References

[1] Andrews, D. W. K. (1988), Tests for parameter instability and structural change with unknown change point, *Econometrica*, **61**, 817-858.

[2] Andrews, D. W. K. and Ploberger, W. (1994), Optimal changepoint tests for normal linear regression, *Journal of Econometrics*, **70**, 9-38.

[3] Bai, J. and Perron, P. (1998), Estimating and testing linear models with multiple structural changes, *Econometrica*, **66**, 47-78.

[4] Basrak B., Davis R.A., and Mikosch T. (2002). Regular variation of GARCH processes. *Stochastic Processes and Their Applications*, **99**, 1, 95-115.

[5] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31**, 307-327.

[6] Brodsky, B. E. and Darkhovsky B. S. (1993), *Nonparametric Methods in Change-Point Problems*, Mathematics and its applications, **243**, Kluwer Academic Publishers, Dordrecht.

[7] Brodsky, B. E. and Darkhovsky B. S. (2000), *Non-Parametric Statistical Diagnosis Problems and Methods*, Mathematics and its applications, **509**, Kluwer Academic Publishers, Dordrecht.

[8] Carrasco M. and Chen X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory*, **18**, 1, 17-39.

[9] Chiu, C. K., Glad, I. K., Godtliebsen, F., and Marron, J. S. (1998), Edge-preserving smoothers for image processing, *Journal of the American Statistical Association*, **93**, 526-556.

[10] Engle, R. F., Bollerslev, T. (1986), Modelling the persistence of conditional variances, *Econometric Reviews*, **5**, 1-50.

[11] Fan, J. and Gijbels (1996). *Local polynomial modelling and it applications*. London, Chapman and Hall.

[12] Ferger, D. (1994a). Change-point esstimators in case of small disorders. *Journal of Statistical Planning and Inference*, **40**, 33-49.

[13] Ferger, D. (1994b). An extension of the Csörgö-Horvath functional limit and its applications to changepoint problems. *Journal of Multivariate Analysis*, **51**, 2, 338-351.

[14] Ferger, D. (1994c). Nonparametric detection of changepoints for sequentially observed data. *Stochastic Processes and their Applications*, **51**, 2, 338-351.

[15] Ferger, D. (1995). Nonparametric tests for nonstandard change-point problems. *The Annals of Statistics*, **23**, 5, 1848-1861.

[16] Ferger, D. (1996). On the asymptotic bahevior of change-point estimators in case of no change with applications to testing. *Statistics and Decision*, **14**, 137-143.

[17] Godtliebsen, F. (1991), Noise reduction using Markov random fields, *Journal of Magnetic Resonance*, **92**, 102-114.

[18] Godtliebsen, F. and Spjøtvoll, E. (1991), Comparison of statistical methods in MR imaging, *International Journal of Imaging Systems and Technology*, **3**, 33-39.

[19] Härdle, W. (1990), *Applied Nonparametric Regression*, Cambridge University Press, Cambridge.

[20] Kim, J. H. and Hart, J. D. (1998), Tests for change in a mean function then the data are dependent, *Journal of Time Series Analysis*, **19**, 399-424.

[21] Kryzak, A., Rafajłowicz, E., and Skubalka-Rafjłowicz, E. (2001). Clipped median and space-filling curves in image filtering. *Technical Report*, ICT Technical University Wrocław, Poland.

[22] Lee, J. S. (1983), Digital image smoothing and the sigma filter, *Computer Vision, Graphics and Image Processing*, **24**, 255-269.

[23] Müller, H. (1992). Change-points in nonparametric regression analysis. *Annals of Statistics*, **25**, 2512-2546.

[24] Pawlak, M. and Rafajłowicz, E. (2000), Vertically weighted regression - a tool for nonlinear data analysis and constructing control charts, *Journal of the German Statistical Association.* **84**, 367-388.

[25] Pawlak, M, Rafajłowicz, E., and Steland, A. (2003). On detecting jumps in time series - Nonparametric setting. *Journal of Nonparametric Statistics*, to appear.

[26] Polzehl, J. and Spokoiny, V.G. (2003). Image denoising: Pointwise adaptive approach. *Annals of Statistics*, **31**, 1, 30-57.

[27] Rafajłowicz, E. (1996), Model free control charts for continuous time production processes, Technical University of Wrocław, Technical Report No. 56/96.

[28] Rue, H., Chu, C.K., Godtliebse, F., and Marron, J.S. (2002). M-smoother with local linear fit. *Journal of Nonparametric Statistics*, **14**, 155-168.

[29] Schmid, W. und Steland, A. (2000). Sequential control of non-stationary processes by nonparametric kernel control charts. *Journal of the German Statistical Association*, **84**, 3, 315-336.

[30] Spokoiny, V.G. (1998). Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. *Annals of Statistics*, **26**, 4, 1356-1378.

[31] Steland, A. (2002). Measuring credit risk: Can we benefit from sequential nonparametric control? *To appear: Proceedings of the 2001 Symposium on Operations Research.*

[32] Steland, A. (2003a). Nonparametric monitoring of financial time series by jump-preserving estimators. *Statistical Papers*, **43**, 361-377.

[33] Steland, A. (2003b). Optimal sequential kernel smoothers under local nonparametric alternatives for dependent processes, *Journal of Statistical Planning and Inference*, in revision.

[34] Steland, A. (2003c). Sequential control of time series by functionals of kernel-weighted empirical processes under local alternatives, *Metrika*, to appear.

[35] Volgesang, T. J. (1998), Trend function hypothesis testing in the presence of serial correlation, *Econometrica*, **66**, 1, 123-148.

[36] Winkler, G. and Liebscher, V. (2002), Smoothers for discontinuous signals, *Journal of Nonparametric Statistics*, **14**, 1-2, 203-222.

[37] Wu Y. (1996). A less sensitive linear detector for the change point based on kernel smoohting method. *Metrika*, **43**, 43-55.

[38] Wu, J. and Chu, C. (1993). Kernel type estimators of jump points and values of a regression function. *Annals of Statistics*, **21**, 1545-1566.