Random Walks with Drift – A Sequential Approach

By Ansgar Steland

Ruhr-Universität Bochum

First Version received October 2002

Abstract. In this paper, sequential monitoring schemes to detect nonparametric drifts are studied for the random walk case. The procedure is based on a kernel smoother. As a by-product we obtain the asymptotics of the Nadaraya–Watson estimator and its associated sequential partial sum process under non-standard sampling. The asymptotic behaviour differs substantially from the stationary situation, if there is a unit root (random walk component). To obtain meaningful asymptotic results, we consider local nonparametric alternatives for the drift component. It turns out that the rate of convergence at which the drift vanishes determines whether the asymptotic properties of the monitoring procedure are determined by a deterministic or random function. Furthermore, we provide a theoretical result about the optimal kernel for a given alternative.

Keywords. Control chart; nonparametric smoothing; sequential analysis; unit roots; weighted partial sum process.

1. Introduction

Many economic time series are non-stationary, and analysts have to take account of that fact. A time series can be trend-stationary or have a random walk component (difference-stationarity). In the first case shocks are temporary, whereas shocks to a random walk are permanent. For unit-root tests, we refer to Dickey and Fuller (1979), Phillips and Perron (1988), Bierens (1997), and Breitung (2002). Often, in particular for financial data, the unit-root hypothesis cannot be rejected; and we are also interested in detecting as soon as possible a change-point where the time series is affected by an additional deterministic drift term. The problem discussed in this article should not be mixed up with the so-called random walk hypothesis which addresses a different issue, namely whether future values are predictable using past values. For that problem, we refer to French and Roll (1986), Fama and French (1988), Lo and MacKinlay (1988), Poterba and Summers (1988) and Jegadeesh (1991).

An important property of a random walk is that there are stochastic trends which can be mixed up with deterministic trends. Nevertheless, a random walk, i.e. a stochastic trend can be overlaid by a deterministic trend component. Hence we study the problem to detect a nonparametric drift component in a random walk. We assume that the observations $Y_{N,1}, Y_{N,2}, \ldots, Y_{N,N}$ arrive sequentially and

$$Y_{N,n+1} = Y_{N,n} + m_{N,n} + u_n, \quad n = 1, \ldots, N, \quad N \in \mathbb{N},$$
where \( u_n \) are independent and identical distributed (i.i.d.) innovations with \( E(u_n) = 0 \) and \( 0 < \text{var}(u_n) < \infty \). For the weak distributional limits presented in this paper, the i.i.d. assumption can be relaxed by a weak condition discussed in detail in Section 2, which allows, e.g. for correlated time series with generalized autoregressive conditionally heteroscedastic (GARCH) effects. To study asymptotic properties analytically, we model the deterministic drift \( m_{N,n} \) more explicitly. However, the detection procedure will not depend on a specification of the alternative, but decides after each new observation whether to continue with observations or whether to stop and reject the null hypothesis of no drift. In any case, we stop no later than after the \( N \)th observation, where \( N \) is calculated in advance.

Whereas \textit{a posteriori} methods aim at estimating consistently the time point where the mean changes and therefore employ data before and after the change-point, sequential monitoring methods use only past and current data, aiming at the detection of a change as soon as possible. The \textit{a posteriori} approach is well studied. For example, Kim and Hart (1998) propose a nonparametric approach to test for a change in a mean function when the data are dependent. Predictive tests for structural change with unknown change-points have been studied by Ghysels \textit{et al.} (1997). The analysis of multiple structural changes in linear models has been discussed, e.g. in Bai and Perron (1998). Yakir \textit{et al.} (1999) use the data after the change for optimization. Hušková and Slabý (2001) studied nonparametric multiple change-point detection based on kernel-weighted means similar to those studied in this article. Kernel-weighted averages have also been discussed by Ferger (1994b,c, 1995, 1996) and Brodsky and Darkhovsky (1993, 2000), where the latter examines \textit{a posteriori} and monitoring procedures. Sequential monitoring procedures to control for the derivative of a process mean have been studied in Schmid and Steland (2000). Results for jump-preserving smoothers can be found in Chiu \textit{et al.} (1998), Pawlak and Rafajlowicz (2000, 2001), Rue \textit{et al.} (2002), Steland (2002c, 2004a, 2005a), and Pawlak \textit{et al.} (2004). For the application of \textit{U}-statistics, we refer to Ferger (1994a, 1997), Gombay and Horvath (1995), and Horváth and Hušková (2003).

The contribution of this paper is to study sequential smoothers to monitor random walks to detect deterministic drifts, and to contrast the results with those of former studies on stationary processes (Steland, 2004b, 2005b). Whereas in the stationary case the normed delay of the procedure converges to a deterministic constant, for a random walk the relevant (kernel-weighted) partial sums have a different convergence rate. Hence, the statistics have to be scaled appropriately to obtain well-defined limit distributions. Furthermore, depending on the rate of convergence of the local alternative, we obtain a deterministic or stochastic limit under the alternative. As a by-product, we provide the asymptotic law of the Nadaraya–Watson estimator. Our approach via kernel-weighted sequential partial sum processes yields asymptotic results for both the \textit{classic fixed sample design} and the \textit{sequential sampling design}. Compared with classic nonparametric regression, the monitoring framework as suggested by Wald (1947), Siegmund

© Blackwell Publishing Ltd 2005
(1985), Brodsky and Darkhovsky (1993), and many others, assumes observations at fixed time points.

The paper is organized as follows. Section 2 discusses the random walk model with local drift and the proposed monitoring procedure. Section 3 gives a brief discussion of the asymptotics for a stationary AR(1) process. Section 4 provides the new results about the control statistic under the random walk model for both the null hypothesis and the alternative. We also discuss general time designs in Section 5, which allow to thin a time series with respect to time. The results are applied in Section 6 to derive the related results for the sequential stopping procedures. Section 7 studies the question of optimal kernel choice. Finally, in Section 8 we study the accuracy of the asymptotic distributions by simulations.

2. MODEL, METHOD AND ASSUMPTIONS

We aim at detecting a nonparametric trend starting at a so-called change-point (break-point) in the presence of a pure random walk without drift. In this section, we explain in detail the model, the proposed method and required assumptions.

2.1. Non-stationary time series model

Assume the data \( Y_1, \ldots, Y_N, N \in \mathbb{N}, \) arrive sequentially,

\[
Y_{N,n+1} = Y_{N,n} + m_{N,n} + u_n, \quad 1 \leq n \leq N, \quad N \in \mathbb{N},
\]

where \( \{u_n\} \) is a sequence of innovation terms with \( E(u_n) = 0 \) and common variance \( 0 < \sigma^2 < \infty. \) We assume that the observation \( Y_n \) is taken at time \( t_n, \) where \( \{t_n\} \) denotes a deterministic and ordered sequence of time points. For convenience, we assume \( t_n = n \in \mathbb{N}. \) Generalizations to other designs are straightforward and discussed in Section 5.

We study a detection procedure which does not depend on a specification of the drift \( m_{N,n}. \) The null hypothesis (in-control model) is that \( m_{N,n} \) vanishes for all \( N, n \in \mathbb{N}, \) and in this case \( Y_{N,n} = Y_n. \) The alternative says that starting at a change-point \( t_q \) specified below the mean changes. Our limit theorems work under the following sequence of alternative models (out-of-control models) for the drift term. We assume

\[
m_{N,n} = m_0 \left( \frac{t_n - t_q}{h_N} \right) h_N^b, \quad n \in \mathbb{N}, \quad h > 0,
\]

where \( h = h_N \) is a sequence of positive constants with

\[
\frac{N}{h_N} \to \zeta \in [1, \infty), \quad \text{as} \ N \to \infty.
\]
m₀, called generic alternative, is a continuous function such that m₀(t) = 0 for t ≤ 0 and m₀(t) ≥ 0 for t > 0; m₀ = 0 corresponds to the null hypothesis. The function m₀ is given by nature and unknown to us. However, in many applications it may be possible to define, e.g. a worst-case scenario in terms of m₀. Then our results can be used to assess the performance of the procedure under that scenario. β ∈ (-1, 0] is a tuning parameter which controls the rate of convergence. If m₀(t) > 0, t ∈ (0, t'), for some t' > 0, then there is a change at time t_q. t_q is called change-point. In this paper we address the following two change-point models.

2.1.1. Change-point model CP1
Having in mind applications where it is reasonable to assume that a change may occur at a fixed given date, e.g. when a firm publishes its balance sheet, it is assumed that t_q = q ∈ N is a fixed integer. Consequently, if m₀ does not vanish, for each fixed N there is a change, but the percentage of pre-change observations tends to 0, as N tends to ∞. It will turn out that, in this case, the asymptotic limit depends on the function m₀, but not on the change-point.

2.1.2. Change-point model CP2
This approach, which is well established in the literature, assumes that the change occurs after a fixed fraction of the data, i.e.

\[ t_q = t_{Nq} = \lfloor N\vartheta \rfloor, \quad \text{for some } \vartheta \in (0, 1). \]

Here and in the sequel we denote by \( \lfloor x \rfloor \) the largest integer less than or equal to x ∈ ℝ. Under this model, the asymptotic limit will depend on the change-point parameter \( \vartheta \), too.

Remark 1. Let us briefly discuss our approach to define local alternatives nonparametrically more precisely. We may write

\[ m_{N,a} = m_N(t_a; \beta), \quad \text{if } m_N(t; \beta) = m_0(\lfloor t - t_q \rfloor / h)h_N^\beta. \]

Hence, since N/h → ζ, for each fixed t we have

\[ \lim_{N \to \infty} m_N(t; \beta) = m_0(0) = 0. \]

Provided m₀(t) is twice differentiable at t = 0 with m₀''(0) < ∞, we have

\[ m_N(t; \beta) = m_0'(0)(t - t_q)h^{\beta-1} + O(h^{\beta-2}). \]

Thus, the underlying drift tends to zero at rate h^{\beta-1}, point-wise.

Although we do not discuss the case of dependent innovations in detail in this article, our results work under the general Assumption A.

Assumption A. The stationary sequence \{u_i\} ensures that the partial sum process \( N^{-1/2} \sum_{i=1}^{\lfloor N\tau \rfloor} u_i, \ r \in [0, 1] \), converges weakly to scaled Brownian motion, \( \sigma B(r) \), for some constant 0 < σ < ∞ which is determined by
\[ \sigma^2 = \lim_{N \to \infty} E \left( N^{-1/2} \sum_{i=1}^{N} u_i \right)^2. \]

It is worth discussing assumption A. First, note that it covers weakly dependent innovations as arising in stationary autoregressive moving average (ARMA) or GARCH models, provided certain additional conditions are fulfilled. In particular, Basrak et al. (2002) have shown that GARCH\((p,q)\) models,

\[ Y_n = \sigma_n \epsilon_n, \quad \sigma_n^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{n-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{n-j}^2, \]

where \(\{\epsilon_n\}\) are independent and identically distributed (i.i.d.) with \(E\epsilon_n = 0\) and \(E\epsilon_n^2 = 1\), are strictly stationary and strongly mixing with geometric rate, if

\[ \alpha_0 > 0, \quad \sum_i \alpha_i + \sum_j \beta_j < 1 \quad \text{and} \quad E \ln^+ |\epsilon_1| < \infty, \]

provided the series is started with its stationary distribution. For a general sufficient condition for Assumption A in terms of the \(\alpha\)-mixing coefficients \(\{\alpha(k) : k \in \mathbb{N}\}\) of \(\{u_n\}\) we refer to Herrndorf (1985), which in particular yields Assumption A provided there exists some \(\delta > 0\) such that

\[ E|u_1|^{2+\delta} < \infty \quad \text{and} \quad \sum_{k=1}^{\infty} \alpha(k)^{\delta/2+\delta} < \infty. \]

Finally, note that this assumption is often considered as a nonparametric definition of an \(I(0)\) process (e.g. Davidson, 2002).

### 2.2. The monitoring procedure

We monitor the time series by a sequential kernel smoother

\[ \hat{m}_n = \frac{\sum_{i=1}^{n} K_h(t_i - t_n) Y_i}{\sum_{i=1}^{n} K_h(t_i - t_n)} \]

which employs only past and current data. The associated kernel-weighted sequential partial sum process is defined as

\[ \hat{m}_N(s) = \frac{\sum_{i=1}^{\lfloor Ns \rfloor} K_h(t_i - t_{\lfloor Ns \rfloor}) Y_i}{\sum_{i=1}^{\lfloor Ns \rfloor} K_h(t_i - t_{\lfloor Ns \rfloor})}, \quad s \in [0, 1]. \]

where \(h\) is a bandwidth parameter given in advance and \(K_h(z) = h^{-1}K(z/h)\) the rescaled version of the smoothing kernel \(K\). If \(K\) vanishes outside the interval \([-1, 1]\), \(h\) equals the number of past observations used by the procedure. To obtain meaningful results, namely weak limits, under alternatives, it turns out that the smoothing parameter \(h\) should converge to \(\infty\), as \(N \to \infty\), i.e. \(h = h(N) \uparrow\). It
turns out that \( h(N) \) and the sequence \( h_N \) appearing in the definition of the local alternatives should satisfy \( \lim_{N \to \infty} h(N)/h_N = c \) for some constant \( c > 0 \). That constant can be absorbed in the unknown function \( m_0 \). Thus, for simplicity, we assume \( h(N) = h_N \). The asymptotic framework is parameterized in the maximum sample size \( N \to \infty \) under the condition (2).

Note that the random function \( \hat{m}_N(\cdot) \) is an element of the Skorokhod space \( D[0, 1] \), consisting of all right-continuous functions with left-hand limits. We denote convergence in distribution of random variables and random vectors by \( \Rightarrow \). Weak convergence in the space \( D[0, 1] \) is denoted by ‘\( \Rightarrow \)’.

The time series is now monitored by the truncated stopping rule

\[
S_N = \inf\{1 \leq n \leq N : T_n(N) > c\}, \quad T_n(N) = c(h, N)\hat{m}_n,
\]

with \( \inf \emptyset = N \). Here \( c(h, N) \) is a scaling function to be chosen later, and \( T_n(N) \) is the rescaled sequential smoother. Note that \( S_N \) is the index of the first time point where the kernel smoother exceeds the threshold (critical value) \( c \).

The monitoring procedure is truncated, i.e. we stop monitoring at \( N \). Note that the definition of \( S_N \) does not depend on any model specification of the alternative.

Concerning the smoothing kernel we make Assumption K.

**Assumption K.** \( K \) is assumed to be a Lipschitz continuous probability density with mean 0 and finite variance. Let \( L \) be the Lipschitz constant, i.e.

\[
|K(z_1) - K(z_2)| \leq L|z_1 - z_2|
\]

holds true for all \( z_1, z_2 \in \mathbb{R} \).

For results under the alternative we need the following conditions.

**Condition M.** \( m_0 \) is a piecewise continuous function.

**Condition KM.** For the function

\[
I(x) = \int_0^x K(s - x) \int_0^s m_0(r)dr ds,
\]

assume \( |I(x)| < \infty \) for all \( x \geq 0 \), \( I \in C(\mathbb{R}_0^+) \), \( K(\cdot) \cdot \int_0^x m_0 \) has bounded variation, and there exists some \( x^* > 0 \) such that \( I(x^*) > c \).

2.2.1. A nuisance-free procedure

It will turn out that the limiting distribution of \( \hat{m}_n \) depends on the nuisance parameter \( \sigma^2 \). A simple candidate is the naive estimator
where $\Delta Y_i = Y_i - Y_{i-1}$, $i = 2, \ldots, n$. Recall that $\hat{\sigma}_n^2$ is consistent for $\sigma^2$ under the null hypothesis, if $\{\Delta Y_n\}$ is a linear process,

$$\Delta Y_n = \sum_{j=-\infty}^{\infty} \psi_j Z_{n-j}$$

where $\{Z_i\}$ are i.i.d. $(0, \eta^2)$ with $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ (Brockwell and Davis, 1991, Propn 7.3.4).

A better choice may be the Gasser’s estimator which is based on a local linear fitting procedure (Gasser et al., 1986). Define the pseudo-residuals

$$\tilde{e}_n = 0.5\Delta Y_{n-1} + 0.5\Delta Y_{n+1} - \Delta Y_n$$

and note that

$$E\tilde{e}_n^2 = ED^2(n, h) + \left(\frac{3}{2}\right)\sigma^2, \quad \text{where } D(n, h) = \left(\frac{1}{2}\right)(m_{n-1} - m_n + m_{n+1} - m_n).$$

By eqn (1), $D^2(n, h) = O(h^{2\beta-2})$, if $h \rightarrow \infty$, provided $m_0$ is twice continuously differentiable. This yields Proposition 1.

**Proposition 1.** If $m_0$ is twice continuously differentiable, the estimator

$$\hat{\sigma}_n^2 = \frac{2}{3(n-2)} \sum_{i=2}^{n-1} \tilde{e}_i^2$$

is asymptotically unbiased, as $h \rightarrow \infty$, if $\{u_n\}$ are i.i.d. with an existing second moment.

Thus, whereas the estimator $\hat{\sigma}_n^2$ tends to overestimate the variance, $\tilde{\sigma}_n^2$ may produce more reliable estimates. A related estimator is Rice’s estimator given by

$$1/(2[n-2]) \sum_{i=2}^{n-1} \Delta^2 Y_{i+1}.$$ 

Thus, we may use the asymptotically nuisance-free control statistic $T_n^*(N) = s_n^{-1} T_n(N)$, where $s_n$ is one of estimators discussed above.

### 3. Asymptotics for Stationary AR Processes

Before turning our attention to the random walk case, let us briefly discuss the situation for a stationary AR(1) process. The asymptotic behaviour follows from general results obtained for stationary $\alpha$-mixing sequences of innovations, but the resulting formulas are slightly different and less explicit.

In this section we assume that $Y_{N,1}, \ldots, Y_{N,N}$ are observations arriving sequentially and
\[ Y_{N,n+1} = aY_{N,n} + m_{N,n} + u_n, \quad n = 1, \ldots, N, \quad N \in \mathbb{N}, \]

where the AR parameter \( a \) satisfies \(|a| < 1\), \( \{u_n\} \) is an i.i.d. sequence of innovations with \( E(u_n) = 0 \) and \( 0 < \text{var}(u_n) = \sigma^2 < \infty \). The deterministic drift component is given by

\[ m_{N,n} = m_0 \left( \frac{t_n - t_q}{h} \right), \]

with \( m_0 \) as in Section 1, but at this point we put \( \beta = 0 \). Note that \( Y_n = Y_{N,n}, \quad n \in \mathbb{N}, \) is stationary under \( H_0 \).

We have

\[ Y_{n+1} = \sum_{i=0}^{\infty} a^i m_{n-i} + \sum_{i=0}^{\infty} a^i u_{n-i}, \]

where \( \sum_{i=0}^{\infty} a^i u_{n-i} \) is a stationary process with autocovariance function

\[ r_0(k) = \frac{\sigma^2 a^k}{(1 - a^2)}, \quad |k| \in \mathbb{N}_0, \]

thus being \( \alpha \)-mixing with geometric rate.

Under the null hypothesis \( H_0 : m_0 = 0 \) we may apply Theorem 3.1 of Steland (2004b) to obtain weak convergence at the usual rate \( N^{1/2} \), i.e.

\[ \frac{h}{N^{1/2}} \hat{m}_N(s) \Rightarrow \mathcal{N}(s), \quad \text{in } D[0,1], \quad (4) \]

as \( N \to \infty \), where \( \mathcal{N}(s) \) is a centred Gaussian process with correlation kernel given by

\[ \text{cor}(\mathcal{N}(s), \mathcal{N}(t)) = \frac{C_\xi(s,t)}{\left( \int_0^s \int_0^t K(z - \xi s)dz \int_0^{s+t} K(z - \zeta t)dz \right)^2}, \]

for \( 0 \leq s \leq t \leq 1 \), with

\[ C_\xi(s,t) = \lim_{N \to -\infty} \sum_{i=1}^{[Ns]} \sum_{j=1}^{[Nt]} K_h(t_i - t_{i[Ns]})K_h(t_j - t_{i[Ns]}) \frac{\sigma^2 a^{|i-j|}}{1 - a^2}. \]

As a result of the Lipschitz continuity of \( K \), the sample paths of \( \mathcal{N}_\xi \) are continuous w.p. 1. Note that

\[ I(m_0) = \lim_{N \to \infty} \sum_{j=0}^{N} a^j m_0 \left( \frac{n - j}{n} \right) < \infty, \]

if \( \int m_0^2(s)ds < \infty \). Now a similar argument as in Theorem 3.3 of Steland (2004b) shows that under the alternative the process in eqn (4) diverges at the rate \( N^{1/2} \), since
\[
h^{-1} \hat{m}_N(s) = \frac{h}{N} \sum_{i=1}^{[Ns]} K_h(t_i - t_i(Ns)) \sum_{j=0}^{i} a^j m_0 \left( \frac{n-j}{h} \right) + o_P(1)
\]

\[
= O \left( I(m_0) \int_0^{\xi s} K(z - \xi s) dz \right).
\]

These results have also immediate implications for the sequential stopping rules. If

\[
\mu_{\xi}(s) = P - \lim_{N \to \infty} h^{-1} \hat{m}_N(s),
\]

it can be shown that for any fixed \(0 < \kappa < 1\)

\[
N^{-1} \inf\{[\kappa N] \leq n \leq N : hN^{-1/2} \hat{m}_n > c\} \xrightarrow{p} \inf\{\kappa \leq s \leq 1 : \mu_{\xi}(s) > c\},
\]

as \(N \to \infty\), i.e. the normed delay converges to a deterministic quantity.

4. ASYMPTOTICS FOR RANDOM WALKS

Now we study the asymptotic behaviour of the Nadaraya–Watson estimator \(\hat{m}_n\) under the random walk model as introduced in Section 2. Note that our asymptotic framework differs from the usual framework in nonparametric regression. We do not assume that the time points \(\{t_i\}\) get dense in any finite time interval or are distributed according to a density, which ensures that we may let the bandwidth \(h\) tend to 0 at a certain rate. Instead, we assume a fixed time design taking account of the fact that time series are commonly observed at a fixed time scale. Thus, as a by-product, we provide the asymptotic laws of the Nadaraya–Watson-type smoothing under the sampling design of the present paper. We formulate the results for equidistant observations, i.e. \(t_n = n\), and discuss more general time designs in Section 4.

The results about the Nadaraya–Watson process \(\hat{m}_N(s), s \in [0, 1]\), are preparations for the analysis of the stopping time \(S_N\), but since they are interesting in their own right we discuss them in detail here. In particular, the interesting relationship between the (qualitative) asymptotic behaviour and the convergence rate of the local alternative are properties of that underlying process.

4.1. Limit theory under the null hypothesis

We first study the asymptotic distributions under the null hypothesis that we deal with a random walk without drift. The limit distributions are centred Gaussian processes and centred normal distributions.
Theorem 1. Assume (A) and (K). Under the null hypothesis \( H_0 : m_0 = 0 \) we have
\[
hN^{-3/2}\hat{m}_N \xrightarrow{d} \frac{\sigma \int_0^1 K(\zeta(r-1))B(r)dr}{\zeta \int_0^1 K(\zeta(r-1))dr},
\]
as \( N \to \infty \). The associated partial sum process converges weakly
\[
hN^{-3/2}\hat{m}_N(s) \Rightarrow \mathcal{M}_\zeta(s) = \frac{\sigma \int_0^s K(\zeta(r-s))B(r)dr}{\zeta \int_0^s K(\zeta(r-s))dr}, \quad \text{in } D[0, 1],
\]
as \( N \to \infty \). The limit process is continuous w.p. 1.

Observe that for \( \sigma = 1 \), the limit process \( \mathcal{M}_\zeta(s) \) is distributed according to a \( N(0, \sigma^2_K) \) distribution with variance given by
\[
\sigma^2_K(s) = \frac{\int_0^1 K(\zeta(s-1))\left[\int_0^s tK(\zeta(t-1))dt + s \int_0^1 K(\zeta(t-1))dt\right]ds}{(\zeta \int_0^1 K(\zeta(r-s))dr)^2}
\]
which can be calculated explicitly for any given kernel (Shorack and Wellner, 1986, p. 42). Table I provides some values of \( \sigma^2_M = \sigma^2_K(1) \) for the Gaussian kernel, the Epanechnikov kernel given by \( K_{\text{Epan}}(z) = (3/4)(1-z^2) \), for \( z \in [-1, 1] \), and the (standardized) Laplace kernel, which is defined by \( K_{\text{Lap}}(z) = (1/\sqrt{2})e^{-\sqrt{2}|z|} \), \( z \in \mathbb{R} \).

Theorem 1 suggests the following confidence interval
\[
\hat{m}_N \pm z_{1-\alpha/2} \sigma_K h^{-1}N^{3/2}
\]
which has asymptotic coverage \( 1 - \alpha \) under \( H_0 \). It can be used to perform a preliminary-level \( \alpha \) test given data \( Y_1, \ldots, Y_N \) before establishing a monitoring procedure. The accuracy of that procedure is studied to some extent in section 7. However, comparing \( \hat{m}_Nh \) with the confidence limits \( z_{1-\alpha/2} \sigma_K h^{-1}N^{3/2} \) does not ensure well-defined statistical properties of the associated stopping rule in terms of the average run length or the normed delay.

Remark 2. Note that the event \( T_n(N) = hN^{3/2}\hat{m}_n > c \) stands for a false alarm at the \( n \)th time point, if \( m_0 = 0 \). It is straightforward to show

| Kernel         | \( \zeta \)         |
|----------------|---------------------|
|                | 10      | 5       | 4       | 2       | 1.5      | 1.2      | 1       |
| Gaussian       | 0.0089  | 0.0310  | 0.0449  | 0.1242  | 0.1913   | 0.2754   | 0.3775  |
| Laplace        | 0.0089  | 0.0316  | 0.0463  | 0.1443  | 0.2310   | 0.3353   | 0.4578  |
| Epanechnikov   | 0.0095  | 0.0359  | 0.0545  | 0.1857  | 0.2921   | 0.3968   | 0.4857  |

\( \zeta = \lim N/h \)

© Blackwell Publishing Ltd 2005
\[ P(hN^{3/2} \hat{m}_N > c) = O(h^{-2}N^{3/2}) = O(N^{-1/2}), \]
i.e. in our framework the point-wise false-alarm rate tends to 0, as \( N \to \infty \).

4.2. Limit theory under local drifts

We now investigate the asymptotic behaviour under the (local) alternative model and both model specifications for the change-point. It turns out that the result depends qualitatively on the rate parameter \( \beta \) of the alternative. If \( \beta = -1/2 \), i.e., the alternative converges at the rate \( h^{-3/2} \) to the null model, we obtain a non-degenerate Gaussian limit with drift for the process \( hN^{-3/2} \hat{m}_N(s) \) studied in Theorem 1 under the null hypothesis. That process has a proper asymptotic null distribution. For a slowly converging alternative (\( \beta = 0 \)) corresponding to the rate \( h^{-1} \), we have to change the scaling function to obtain a limit. In this case we obtain stochastic convergence to a non-stochastic function. That function determines the asymptotic detection properties of the proposed procedure. We formulate the results for the partial sum processes \( \hat{m}_N(s) \), and putting \( s = 1 \) yields the asymptotic laws of the Nadaraya–Watson estimator.

**Theorem 2.** Assume (A), (K), (M), and (KM). Fix \( 0 < a < 1 \). Under the alternative \( H_1 : m_0 \geq s \) 0 the following assertions hold true.

(i) If \( \beta = -1/2 \), we have weakly in \( D[a, 1] \),

\[
\frac{h}{N^{3/2}} \hat{m}_N(s) \Rightarrow \sigma \left[ \int_0^s K(\zeta(r-s))B(r)\,dr \right] + \frac{\int_0^s K(\zeta(r-s)) \int_0^t \int_0^r m_0(t-\zeta \vartheta 1_{CP2})\,dt\,dr}{\zeta^{3/2} \int_0^s K(\zeta(r-s))\,dr},
\]
as \( N \to \infty \). Here, \( 1_{CP2} = 0 \) if change-point model CP1 holds, and \( 1_{CP2} = 1 \) under model CP2.

(ii) If \( \beta = 0 \), then

\[
\frac{h^{1/2}}{N^{3/2}} \hat{m}_N(s) \Rightarrow \sigma \left[ \int_0^s K(\zeta(r-s)) \int_0^t \int_0^r m_0(t-\zeta \vartheta 1_{CP2})\,dt\,dr \right] \frac{\zeta^{3/2} \int_0^s K(\zeta(r-s))\,dr}{\zeta^{3/2} \int_0^s K(\zeta(r-s))\,dr},
\]
as \( N \to \infty \). Again, \( 1_{CP2} = 0 \) if change-point model CP1 holds, and \( 1_{CP2} = 1 \) under model CP2.

**Remark 3.** Note that the asymptotic limit depends on the change-point parameter \( \vartheta \) if model CP2 holds. Under model CP1 the limit is free of \( t_q \), which is a consequence of \( t_q/h = o(1) \) and continuity of \( m_0 \).

**Remark 4.** It is worth noting that procedures based on the partial sum process \( \hat{m}_N(s) \) are able to detect a drift if the function
\begin{equation}
\mu \zeta(s) = \int_0^s K(\zeta(t-s)) \int_0^s \theta m_0(t-\zeta) 1_{CP2} \, dt \, dr
\end{equation}

is positive for some interval of \( s \)-values.

**Remark 5.** Note that Theorem 2(ii) implies that the statistic \( h N^{-3/2} \hat{m}_N \) diverges under local alternatives corresponding to \( \beta = 0 \) at the rate \( h^{1/2} \).

## 5. General Time Designs

Let us briefly discuss more general time designs for the choice of the time points \( t_n \). In some applications, the following monitoring approach may be possible and reasonable. We monitor the process at equidistant time points \( 1, 2, \ldots \) until either the procedure provides a signal, or we have reached the time horizon (maximum sample size) \( N \). Here we assume that the time unit is chosen appropriately, e.g. 1 day or 1 week. Intuitively, to detect a change as soon as possible it should be better to use more recent observations \( Y_i \), i.e. with \( t_n - t_i \) small, than past observations where \( t_n - t_i \) is large. To some extent, this is achieved by the smoothing kernel, which downweights past data, but a real thinning of the data can only be achieved by an appropriate selection resp. design of the time points. This means, at the current time \( t_{n,n} = n \) one chooses past time points \( 0 < t_{n,1}, \ldots, t_{n,n-1} < t_{n,n} \) where observations are taken. This allows to start with monthly observations and use daily observations at the end of the (current) sample. We consider two different approaches corresponding to the two change-point models CP1 and CP2.

### 5.1. Generalized time designs for the CP1 model

Assume that

\begin{equation}
t_{n,i} = n F_T^{-1} \left( \frac{i}{n} \right), \quad i = 1, \ldots, n, \quad n \in \mathbb{N},
\end{equation}

where \( F_T \) is a continuously differentiable d.f. with support \([0, 1]\). Clearly, if \( F_T \) is the d.f. of the uniform distribution on \([0, 1]\), we obtain \( t_{n,i} = i \). Nonlinear choices of \( F_T \) allow to ensure that past or more recent observations dominate the sample. Note that \( F_T^{-1} \) defines a sampling scheme which is rolled over the time axis: at each time \( n \), the time points \( t_{n,1}, \ldots, t_{n,n-1} \) are chosen according to the scheme (6).

Under model CP1, a Taylor expansion yields \( t_{nq} = (F_T^{-1})'(0)q + o(1) \) provided \( F_T^{-1} \) is continuously differentiable. Thus, if \( (F_T^{-1})'(0) > 0 \), the underlying (asymptotic) change-point equals \( (F_T^{-1})'(0)t_q \), whereas for \( (F_T^{-1})'(0) = 0 \) the sequence of change-points vanishes asymptotically, i.e. the detection problem is made easier as \( n \) increases.
The associated Nadaraya–Watson process is given by
\[ \hat{m}_N(s) = \frac{\sum_{i=1}^{[Ns]} K_h(t_{[Ns],i} - [Ns])Y_i}{\sum_{i=1}^{[Ns]} K_h(t_{[Ns],i} - [Ns])}, \quad s \in [0, 1], \]
where again \([Ns]\) plays the role of the current time point. It is straightforward to check that the proofs of Theorems 1 and 2 still work. Now the limit process under the null hypothesis is given by
\[ M_{\tilde{\zeta},F_T}(s) = \int_0^s K(\tilde{\zeta}s[F_T^{-1}(r/s) - 1])B(r)dr \int_0^s K(\tilde{\zeta}s[F_T^{-1}(r/s) - 1])dr, \quad s \in [0, 1]. \]

The drift term appearing in Theorem 3 changes to
\[ \mu_{\tilde{\zeta},F_T}(s) = \frac{\int_0^s K(\tilde{\zeta}s[F_T^{-1}(r/s) - 1])\int_0^r m_0(\tilde{\zeta}sF_T^{-1}(t/s))dt dr}{\zeta^{3/2}\int_0^s K(\tilde{\zeta}s[F_T^{-1}(r/s) - 1])dr}, \quad s \in [0, 1]. \]

Remark 6. In practice, it may be necessary to use the time point \(t_{n,j}^{*} \in \{t_{n,1}^{*}, \ldots, t_{n,m}^{*}\}\) nearest to \(t_{n,i}\), where \(\{t_{n,j}^{*}\}\) denotes the finest discrete time scale available. Then, eqn (6) defines a selection rule for the time points \(\{t_{n,j}^{*}\}\).

5.2. Generalized time designs for the CP2 model

It is easy to see that the generalized time design above does not make much sense under model CP2. One may consider the following modification, which is easier to apply, but lacks the authentic idea to allow for schemes which use more observations near each current time \(n\). Assume
\[ t_{N,i} = NF_T^{-1}\left(\frac{i}{N}\right), \quad i = 1, \ldots, N, \quad (7) \]
where \(F_T\) is a continuously differentiable d.f. with support \([0, 1]\). Here, given the maximum sample size \(N\), the time design scheme is set up only once, i.e. the selected time points do not change with the current time \(n\). Since under model CP2 the change-point is given by \(t_q = t_{Nq} = [N\vartheta]\), we obtain
\[ t_{Nq} = NF_T^{-1}\left(\frac{[N\vartheta]}{N}\right) \]
yielding \(t_{Nq}/N \to F_T^{-1}(\vartheta)\). This means that the (asymptotic) change-point parameter is transformed by \(F_T^{-1}\), and it appears in the asymptotic limit. The associated Nadaraya–Watson process is now defined by
\[
\hat{m}_N(s) = \frac{\sum_{i=1}^{[Ns]} K_h(N/h[F_{T_i}^{-1}(i/N) - F_{T_i}^{-1}([Ns]/N)])Y_i}{\sum_{i=1}^{[Ns]} K_h(N/h[F_{T_i}^{-1}(i/N) - F_{T_i}^{-1}([Ns]/N)])}, \quad s \in [0, 1].
\]

A straightforward calculation shows that the drift term now changes to
\[
\mu_{\zeta, F_T}(s) = \int_0^s K(\zeta s[F_T^{-1}(r) - F_T^{-1}(s)]) \int_0^r m_0(\zeta [F_T^{-1}(t) - F_T^{-1}(\theta)]) dt dr, \quad s \in [0, 1].
\]

Note that Remark 6 is also applicable to the time design scheme (7).

6. SEQUENTIAL DETECTION RULES

Let us now discuss the implications of the results of Section 4 for the stopping rule
\[
S_N = \inf\{0 \leq n \leq N : T_n(N) > c\}.
\]

Note that \(S_N\) can be written in terms of the sequential partial sum processes. Indeed,
\[
S_N = N \cdot \inf\{0 \leq s \leq 1 : c(h, N)\hat{m}_N(s) > c\}.
\]

For asymptotic results under local alternatives, we also consider the stopping rule
\[
S_N^{(a)} = \inf\{[Na] \leq n \leq N : c(h, N)\hat{m}_N\left(\frac{n}{N}\right) > c\}
\]

where \(a \in (0, 1)\) is a fixed constant. Again notice that \(S_N^{(a)}\) can be written as
\[
N \cdot \inf\{a \leq s \leq 1 : T_N(s) > c\}, \quad \text{where } T_N(s) = c(h, N)\hat{m}_N(s).
\]

6.1. Limit theory under the null hypothesis

Theorem 3 provides the null distribution of the stopping rules.

**Theorem 3.** Assume (A), (K), and \(H_0 : m_0 = 0\) (random walk without drift).

(i) If \(T_n(s) = c(h, N)\hat{m}_n(s)\) with scaling factor \(c(h, N) = hN^{-3/2}\), the normed stopping time \(S_N/N\) converges in distribution to the random variable
\[
S_* = \inf\left\{s \in [0, 1] : \frac{\sigma \int_0^s K(\zeta(r - s))B(r)dr}{\zeta \int_0^s K(\zeta(r - s))dr} > c\right\},
\]

as \(N \to \infty\).

(ii) The limiting laws of the nuisance-free versions correspond to the special case \(\sigma = 1\).
These results can be used to choose the threshold (critical value) \( c \) from the asymptotic distribution. For example, we may simulate trajectories from the limiting processes and determine, for each trajectory, the smallest \( s \) such that the threshold \( c \) is exceeded. This gives an approximation of the distribution of \( S_N \) which can be used to choose \( c \) to ensure that, e.g. the average run length equals a prespecified value.

6.2. Limit theory under local drifts

The following results summarize our findings under local alternatives and give interesting insights into the asymptotic properties of the procedure. In particular, we see how the smoothing kernel and the generic alternative \( m_0 \) jointly affect the performance of the procedures.

**Theorem 4.** Assume (A), (K), (M), and (KM) (random walk with local drift). Fix \( a \in (0, 1) \).

(i) Suppose \( \beta = -1/2 \). If \( T_N(s) = c(h, N)\hat{m}_N(s) \) with scaling factor \( c(h, N) = hN^{-3/2} \), the normed stopping time \( S_N^{(a)} \) converges weakly to the random variable

\[
S(\zeta) = \inf\{s \in [a, 1] : W_\zeta(s) > c\},
\]

where the stochastic process \( W_\zeta(s) \) is given by

\[
W_\zeta(s) = \frac{\sigma \int_0^s K(\zeta(r-s))B(r)dr + \int_0^s \frac{K(\zeta(r-s))}{\zeta^3/2} m_0(t - \zeta \vartheta_{1CP2})dt dr}{\zeta^3/2 \int_0^s K(\zeta(r-s))dr},
\]

as \( N \to \infty \).

(ii) Suppose \( \beta = 0 \). If \( T_N(s) = c(h, N)\hat{m}_N(s) \) with scaling factor \( c(h, N) = h^{1/2}N^{-3/2} \), the normed stopping time \( S_N^{(a)} \) converges in probability to the non-stochastic asymptotic normed delay

\[
S^*(\zeta; K; m_0) = \inf\left\{ s \in [a, 1] : \frac{\int_0^s K(\zeta(r-s)) \int_0^r m_0(t - \zeta \vartheta_{1CP2})dt dr}{\zeta^3/2 \int_0^s K(\zeta(r-s))dr} > c \right\},
\]

as \( N \to \infty \).

This theorem says that the stopping rule relying on the control statistic \( hN^{-3/2}\hat{m}_N \), which has a proper limit under \( H_0 \), has a nondegenerate limit distribution under local alternatives converging to 0 at the rate \( h^{-3/2} \). If, however, we consider alternatives with rate \( h^{-1} \), which is the appropriate rate in the stationary case (see Steland, 2004b), and change the scaling function, we obtain a deterministic limit \( S^*(\zeta; K; m_0) \), the asymptotic normed delay, as in the case of a stationary process.
7. OPTIMAL KERNEL CHOICE

Suppose the critical value \( c \) is a fixed constant chosen by the data analyst. For example, when analysing a time series representing financial risk measured in terms of a currency unit, \( c \) may be a psychological price. Then \( S_N \) stands for the time point where that price is reached for the first time.

Assuming the change-point model CP1, Theorem 4(ii) motivates to examine whether optimal kernels exist which minimize the asymptotic normed delay \( S^*(\xi; K; m_0) \) for a given alternative \( m_0 \) representing a worst-case scenario. Recall that this deterministic quantity appears as the limit if the alternative model converges to the null model at the rate \( h^{-1} \), whereas for the faster rate \( h^{-3/2} \) we obtained a stochastic limit. From a practical viewpoint, considering the conditions for a slower convergence to 0 may provide a better approximation to reality.

First, note that for a finite set of candidate kernels, \( \{K_1, \ldots, K_M\} \), we can simply plot the \( M \) corresponding curves

\[
y_l(s) = \frac{\int_0^s K_l(\xi(r-s)) \int_0^r m_0(t) dt dr}{\int_0^s K(\xi(s-r)) dr}, \quad l = 1, \ldots, M,
\]

and use the kernel which provides the smallest \( s \) where the critical value \( c \) is exceeded. For the case of detecting a drift in a stationary process, Steland (2002a) provides a real-data analysis of credit risk data, where this simple procedure yields a detection rule which signals the change one time point earlier. For a Bayesian view on the problem of kernel optimization, see Steland (2002b).

Although we can provide a solution to the problem of optimal kernel choice, the results seem to be of limited practical use, since we can identify the optimal kernel only for a finite interval around 0. Nevertheless, from a theoretical point of view it is interesting to know that both the asymptotic normed delay and the optimal kernel can be calculated explicitly for any given generic alternative \( m_0 \).

Let \( \mathcal{K} \) denote a class of probability densities with expectation 0, which is uniformly Lipschitz continuous, i.e.

\[
\sup_{K \in \mathcal{K}} |K(z_1) - K(z_2)| \leq L|z_1 - z_2|, \quad \forall z_1, z_2 \in \mathbb{R},
\]

holds for some constant \( L > 0 \). The problem is to find a kernel \( K^* \in \mathcal{K} \) such that the corresponding asymptotic normed delay, \( S^*(\xi; K^*; m_0) \), satisfies

\[
S^*(\xi; K^*; m_0) = \inf\{S^*(\xi; K; m_0) : K \in \mathcal{K}\}.
\]

Such a pair \((K^*, S^*(\xi; K^*; m_0))\) is called optimal. Using optimization techniques presented in detail in Steland (2004b), one can establish Theorem 5 which provides a way to calculate the optimal asymptotic normed delay and provides the optimal kernel \( K^* \).
Theorem 5. Suppose that for all $s \in [0, 1]$

$$0 < \int_0^s \left( \int_0^r m_0(t) \, dt \right)^2 \, dr < \infty.$$ 

(i) The optimal asymptotic normed delay is given by

$$S^*(\zeta, K^*; m_0) = \inf \left\{ s \in [0, 1] : \frac{\int_0^s (\int_0^r m_0(t) \, dt)^2 \, dr}{\int_0^s \int_0^r m_0(t) \, dt \, dr} > c \right\}.$$

(ii) The optimal kernel $K^*$ satisfies

$$K^*(z) = \frac{\int_0^{\zeta + S^*(\zeta, K^*; m_0)} m_0(t) \, dt}{2 \int_0^\infty \int_0^r m_0(t) \, dt \, dr}$$

for arguments $z \in [-\zeta S^*(\zeta; K^*; m_0), \zeta S^*(\zeta; K^*; m_0)]$.

8. Simulations

To study the accuracy of the asymptotic distributions of the detection procedures, we simulated random walks, $\{Y_n\}$, where $Y_0 = 0$, and $Y_{n+1} = Y_n + u_n$ with $\{u_n\}$ i.i.d. $N(0, \sigma^2)$, $\sigma = 1$. To estimate the nuisance parameter $\sigma^2$ we assumed that an additional pre-run random walk of length $h = 10$ was given.

Figure 1 shows 20 realizations of the kernel-weighted sequential partial sum process, $\hat{m}_N(s)$, $s \in [0, 1]$, for $N = 100$ and $h = 50$ and its asymptotic approximation via the kernel-weighted integral over Brownian motion using $\zeta = 2$. The sequential detection procedure $S_N$ can be visualized by drawing a horizontal line (control limit) at $c$. The first intersection of the process and the control limit is the run length. To study the accuracy of the asymptotic null distribution, we performed simulations to assess the coverage of the confidence interval based on $\hat{m}_N$ and average run lengths (ARL) of the stopping rule $S_N$. We focus on the ARL, since it may the most common criterion to design monitoring procedures for practical applications. Note, however, that our results also allow to design procedures which control the type I error rate.

Table II reports the simulated coverage probabilities of the confidence interval defined in eqn (5) for a Gaussian kernel and a nominal coverage of 0.95 under the null hypothesis. The results for the Epanechnikov and Laplace kernel were in close agreement and are not reported here. Each value is estimated by 10,000 repetitions. The asymptotic variance is estimated using the estimator (3) and $\sigma^2_K$ as given in Table I. It can be seen that even for $h \ll N$ and small $N$ coverage is good.

In order to simplify the application of the proposed sequential monitoring procedure we provide curves to obtain approximate critical values to achieve a
prespecified ARL, \( E_0(S_N) \), under the null hypothesis \( H_0 : m_0 = 0 \). Figure 2 provides curves of the normed ARL \( a_0 = E_0(S_N)/N \) as a function of \( c \), i.e. \( a_0 = a_0(c) \). For given \((N, h)\) use the curve for \( \zeta \approx N/h \) and determine \( c \) graphically with \( Nc \approx a_0(c) \).

How accurate is that approximation? To gain some insight we compared the asymptotic distribution of the stopping time

\[
S_\zeta = \inf \left\{ 0 \leq s \leq 1 : \frac{\int_0^s K(\zeta(r-s))B(r)dr}{\int_0^s K(\zeta(r-s))dr} > c \right\}
\]

with the true distribution of the normed stopping time

\[
\frac{S_{Nh}}{N} = \inf \left\{ 1 \leq n \leq N : \frac{\hat{\sigma}_n^{-1}hN^{-3/2}\hat{m}_{nh}(s)}{N} > c \right\}
\]

in terms of the ARL. Each ARL was approximated using 10,000 trajectories.

![Figure 1. Twenty realizations of the kernel-weighted sequential process \( \hat{m}_{nh}(s), s \in [0, 1] \), (bold line) and its asymptotic limit (dashed line).](image)

| \( \zeta \) | 10     | 50   | 100  | 250  | 500   |
|-------|-------|------|------|------|------|
| 10    | 0.9502 | 0.9496 | 0.9523 | 0.9502 | 0.9471 |
| 5     | 0.9481 | 0.9514 | 0.9478 | 0.9489 | 0.9534 |
| 4     | 0.9475 | 0.9525 | 0.9474 | 0.9473 | 0.9515 |
| 2     | 0.9408 | 0.9468 | 0.9480 | 0.9458 | 0.9518 |
| 1.5   | 0.9350 | 0.9431 | 0.9516 | 0.9512 | 0.9485 |
| 1.2   | 0.9320 | 0.9453 | 0.9523 | 0.9477 | 0.9518 |
| 1     | 0.9301 | 0.9526 | 0.9470 | 0.9494 | 0.9504 |

**TABLE II**

| Coverage probabilities of a 0.95 Confidence Interval for Random Walks |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| \( \zeta \) | 10 | 50 | 100 | 250 | 500 |
| 10 | 0.9502 | 0.9496 | 0.9523 | 0.9502 | 0.9471 |
| 5 | 0.9481 | 0.9514 | 0.9478 | 0.9489 | 0.9534 |
| 4 | 0.9475 | 0.9525 | 0.9474 | 0.9473 | 0.9515 |
| 2 | 0.9408 | 0.9468 | 0.9480 | 0.9458 | 0.9518 |
| 1.5 | 0.9350 | 0.9431 | 0.9516 | 0.9512 | 0.9485 |
| 1.2 | 0.9320 | 0.9453 | 0.9523 | 0.9477 | 0.9518 |
| 1 | 0.9301 | 0.9526 | 0.9470 | 0.9494 | 0.9504 |

© Blackwell Publishing Ltd 2005
Figure 2. Normed ARL curves for $hN^{-3/2}\hat{\theta}_h$ using the Gaussian kernel. $\zeta$ attains the values 1 (bottom curve), 1.5, 2, 3, 4, 5, and 10 (top curve).

Figure 3 provides the results. For $h \in \{10, 20, 50, 100\}$, $N = \zeta h$, and $\zeta = 3$ (left panel) and $\zeta = 10$ (right panel) the corresponding normed-ARL curves are shown. It can be seen that the curve representing the asymptotic critical values are below the simulated true curves. This means, the asymptotic critical values yield conservative procedures. The accuracy seems to be better for large values of $\zeta$, i.e. if $h$ is small compared with $N$.

APPENDIX: PROOFS

In this paper we work with weak convergence (denoted by $\Rightarrow$) of elements of the space $(D[0, 1], d)$ where $d$ is the Skorokhod metric. For treatments of the general theory we refer to Billingsley (1968), Pollard (1984) and Van der Vaart and Wellner (1996).
Proof of Theorem 1. Put $Y_0 = 0$ and define
\[ X_N(r; s) = N^{-1} Y_{[N r]} K_h(t_{[N r]} - t_{[N s]}), \quad r, s \in [0, 1]. \]
Note that $X_N(r; s)$ is a constant on the intervals $[i/N, (i+1)/N)$ with value $N^{-1} Y_i K_h(t_i - t_{(N+1)/N})$, $i = 1, \ldots, N$. Therefore, the area under the curve $X_N(r; s)$, $r \in [0, s]$, is given by
\[ \int_0^s X_N(r; s) \, dr = \frac{1}{N^2} \sum_{i=1}^{[N s]} K_h(t_i - t_{(N+1)/N}) Y_i. \]
Using
\[ Y_{[N r]} = \sum_{i=1}^{[N r]} u_i, \quad \text{we have } h N^{1/2} X_N(r; s) = \frac{1}{\sqrt{N}} \sum_{i=1}^{[N r]} u_i \cdot K \left( \frac{t_i - t_{((N+1)/N)}}{h} \right). \]

Since by Assumption A the partial sum process $N^{-1/2} \sum_{i=1}^{[N r]} u_i$ converges weakly to scaled Brownian motion $\sigma B(r)$, we may apply the a.s. representation theorem of Skorokhod and Dudley (Pollard, 1984, p. 71) which ensures that there exist versions of the random elements which converge a.s. in the supnorm. This implies
\[ \left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{[N r]} u_i K \left( \frac{N o_i - [N o_1]}{h} \right) - \sigma B(o_1) K(\zeta(o_1 - o_2)) \right\|_{D([0,1] \times [0,1])} \to 0, \]
which proves weak convergence.
\[ hN^{1/2}X_n(r; s) \Rightarrow \sigma K(\zeta(r - s))B(r) \quad \text{in } D([0, 1] \times [0, 1]). \]

By continuity of \( K \), the process \( \sigma K(\zeta(r - s))B(r) \), \((s, r) \in [0, 1] \times [0, 1]\), has continuous and bounded sample paths w.p. 1. Consider the integral operator \( I \) which maps an element \( f \in D([0, 1] \times [0, 1]) \) to the element \( I(f) \in D[0, 1] \) given by

\[
I(f)(s) = \int_0^s f(r, s) dr, \quad s \in [0, 1].
\]

If \((f_n) \subset D([0, 1] \times [0, 1])\) is a convergent sequence with limit \( f \in C([0, 1] \times [0, 1])\), i.e. \( d(f_n, f) \to 0 \), as \( n \to \infty \), then we also have \( \|f_n - f\|_{\infty} \to 0 \), \( n \to \infty \), yielding \( \|I(f_n) - I(f)\|_{\infty} \to 0 \), as \( N \to \infty \), i.e. continuity of \( I \). Hence, the continuous mapping theorem yields

\[
\frac{h}{N^{3/2}} \sum_{i=1}^{[N_{01}]} K_h(t_i - t_{[N_{01}]}))Y_i = I(hN^{1/2}X_n(o_1) \Rightarrow \sigma \int_0^{o_1} K(\zeta(r - o_1))B(r) dr,
\]

weakly in \( D[0, 1] \), as \( N \to \infty \). Since additionally,

\[
\sum_{i=1}^{[N_{01}]} K_h(t_i - t_{[N_{01}]} \to \int_0^{\zeta s} K(r - \zeta s) dr = \zeta \int_0^s K(\zeta(r - s)) dr, \quad (8)
\]

as \( N \to \infty \), the assertions follow.

**Proof of Theorem 2.** A random walk with non-vanishing drift, \( Y_{n+1} = Y_n + m_n + u_n \), can be decomposed as

\[
Y_n = \tilde{Y}_n + \sum_{i=1}^{n-1} m_i, \quad n \in \mathbb{N}, \quad \text{where } \tilde{Y}_n = \sum_{i=1}^{n-1} u_i
\]

is a random walk based on the innovations \( u_n \) without drift. Hence,

\[
hN^{-3/2} \sum_{i=1}^{[N_{01}]} K_h(t_i - t_{[N_{01}]}))Y_i
\]

can be decomposed as

\[
hN^{-3/2} \sum_{i=1}^{[N_{01}]} K_h(t_i - t_{[N_{01}]}))\tilde{Y}_i + hN^{-3/2} \sum_{i=1}^{[N_{01}]} K_h(t_i - t_{[N_{01}]})) \sum_{j=1}^{i-1} m_{ij}.
\]

For the first term one may argue as in the proof of Theorem 1 to verify that

\[
hN^{-3/2} \sum_{i=1}^{[N_{01}]} K_h(t_i - t_{[N_{01}]}))\tilde{Y}_i \Rightarrow \sigma \int_0^{s} K(\zeta(r - s))B(r) dr, \quad (9)
\]

as \( N \to \infty \). Further, since \( m_n = m_0(\ell_1 - \ell_0)/h^{1/2} \), \( \beta = -1/2 \) implies
\[ \mu_N(s) = \frac{h}{N^{3/2}} \sum_{i=1}^{\lceil Ns \rceil} K_h(t_i - t_{\lfloor Ns \rfloor}) \sum_{j=1}^{i-1} m_0 \left( \frac{[t_j - t_i]}{h} \right) h^0 \]

\[ \rightarrow - \frac{1}{\zeta^{1/2}} \int_0^s K(\zeta(r - s)) \int_0^{\zeta r} m_0(t - \zeta \partial 1_{CP2}) dt \, dr, \]

as \( N \to \infty \), by (K) and (KM) uniformly in \( s \in [a, 1] \) cf. Steland, 2004b, Th. 3.3(ii).

Combining this fact with eqns (9) and (8) yields

\[ \frac{h}{N^{3/2}} \bar{m}_h(s) = \alpha \int_0^s K(\zeta(r - s)) B(r) dr + \frac{\zeta^{-1/2} \int_0^s K(\zeta(r - s)) \int_0^{\zeta r} m_0(t - \zeta \partial 1_{CP2}) dt \, dr}{\zeta \int_0^s K(\zeta(r - s)) dr}, \]

in \( D[a, 1] \), as \( N \to \infty \). In contrast, if \( \beta = 0 \) we obtain convergence to a deterministic quantity, if we change the scaling factor from \( hN^{-3/2} \) to \( h^{1/2}N^{-3/2} \). Indeed, in this case we have

\[ \frac{h^{1/2}}{N^{3/2}} \sum_{i=1}^{\lceil Ns \rceil} K_h(t_i - t_{\lfloor Ns \rfloor}) \bar{Y}_i = o_P(1), \]

uniformly in \( s \in [a, 1] \), and for the centering term

\[ h^{-1/2} \mu_N(s) = \frac{h^{3/2}}{N^{3/2}} h^{-2} \sum_{i=1}^{\lceil Ns \rceil} K \left( \frac{[t_i - t_{\lfloor Ns \rfloor}]}{h} \right) \sum_{j=1}^{i-1} m_0 \left( \frac{[t_j - t_i]}{h} \right) \]

\[ \rightarrow \zeta^{-1/2} \int_0^s K(\zeta(r - s)) \int_0^{\zeta r} m_0(t - \zeta \partial 1_{CP2}) dt \, dr, \]

yielding

\[ \frac{h^{1/2}}{N^{3/2}} \bar{m}_h(s) \rightarrow \frac{\zeta^{-1/2} \int_0^s K(\zeta(r - s)) \int_0^{\zeta r} m_0(t - \zeta \partial 1_{CP2}) dt \, dr}{\zeta \int_0^s K(\zeta(r - s)) \, dr}, \]

uniformly in \( s \in [a, 1] \), as \( N \to \infty \).

**Proof of Theorems 3 and 4.** We verify Theorem 3(i), i.e. assuming \( \beta = -1/2 \) and \( c(h, N) = hN^{-3/2} \). The other assertions are shown along these lines. Fix \( 0 < a < 1 \). By Theorem 2(i) the process \( c(h, N)\bar{m}_h(s) \) converges weakly in \( D[a, 1] \) to the non-stationary and a.s. continuous process

\[ W_\zeta(s) = \frac{\sigma \int_0^s K(\zeta(r - s)) B(r) dr + \left( \zeta^{-1/2} \int_0^s K(\zeta(r - s)) \int_0^{\zeta r} m_0(t) dt \right) dr}{\zeta \int_0^s K(\zeta(r - s)) \, dr}, \]

as \( N \to \infty \). Define the functional \( \varphi_c : D[0, 1] \to D[0, 1] \),

\[ \varphi_c(f) = \inf \{ a \leq s \leq 1 : f(s) > c \}, \quad f \in D[a, 1]. \]

Clearly, \( \varphi_c \), \( \mathcal{E}_c \) is continuous w.r.t. \( \| \cdot \|_\infty \) and \( d \), where

\[ \mathcal{E}_c = \{ f \in C[0, 1] : f(x') > c \text{ for some } x' \}. \]

By (K) and (M) we have \( W_\zeta \in C[a, 1] \) w.p. 1. Thus, since

\[ \frac{S_N^{(a)}}{N} = \varphi_c(c(h, N)\bar{m}_h(\cdot)), \]

© Blackwell Publishing Ltd 2005
the continuous mapping theorem yields
\[
S_N^{(a)} \Rightarrow \varphi_a(W_\zeta(\cdot)) = \inf\{a \leq s \leq 1 : W_\zeta(s) > c\},
\]
as \(N \to \infty\). Notice that
\[
\inf\{a \leq s \leq 1 : W_\zeta(s) > c\} > x \iff \sup_{0 \leq s \leq x} W_\zeta(s) \leq c.
\]
By a.s. continuity of \(W_\zeta\), Theorem 2 of Lifshits (1982) ensures that \(v_x = \mathcal{L}(\sup_{0 \leq s \leq x} W_\zeta(s))\) can have an atom only at the point
\[
\gamma_x = \sup_{0 \leq r \leq x : \text{Var}(W_\zeta(r)) = 0} EW_\zeta(r),
\]
vanishes on \((-\infty, \gamma_x)\), and is absolutely continuous on \((\gamma_x, \infty)\). Since \(\text{var}(W_\zeta(s)) > 0\) if \(s > 0\), \(v_x\) is absolutely continuous. Therefore, we obtain convergence in distribution, i.e.
\[
P(\inf\{a \leq s \leq 1 : c(h, N)\hat{m}_N(s) > c\} \leq x) \to P(\inf\{a \leq s \leq 1 : W_\zeta(s) > c\})
\]
as \(N \to \infty\), for all \(x\).

**Proof of Theorem 5.** Using standard arguments of functional optimization theory, we see that \(S^*(\zeta ; K; m_0)\) is minimized w.r.t. \(K\) if
\[
\tau(K) = \frac{\int_0^{s^*} K(\zeta(r - s^*)) \int_0^{r} m_0(t) \, dt \, dr}{\int_0^{s^*} K(\zeta(r - s^*)) \, dr}
\]
is maximized w.r.t. \(K \in \mathcal{K}\), where \(s^* = S^*(\zeta, K^*; m_0)\) denotes the optimal asymptotic normed delay (c.f. Steland, 2004b). Clearly, \(\tau(K) \geq 0\) is less than or equal to
\[
\sqrt{\int_0^{s^*} K(\zeta(r - s^*))^2 \, dr} \sqrt{\int_0^{s^*} (\int_0^{r} m_0(t) \, dt)^2 \, dr}
\]
with equality if and only if
\[
\frac{K(\zeta(r - s^*))}{\int_0^{s^*} K(\zeta(r - s^*)) \, dr} = \lambda \int_0^{r} m_0(t) \, dt, \quad r \in [0, s^*],
\]
for some \(\lambda\). Using
\[
\int_0^{\infty} K(\zeta(r - s^*)) \, dr = \frac{1}{2}
\]
gives
\[
\lambda^{-1} = 2 \int_0^{\infty} \int_0^{r} m_0(t) \, dt \, dr \int_0^{s^*} K(\zeta(r - s^*)) \, dr,
\]
i.e. the optimal (symmetric) kernel \(K^*\) satisfies
\[
K^*(\zeta(r - s^*)) = \frac{\int_0^{r} m_0(t) \, dt}{2 \int_0^{\infty} \int_0^{r} m_0(t) \, dt}, \quad r \in [-s^*, s^*].
\]
Consequently, using $K(\zeta(r - s')) = K(\zeta(s') - r)$ and substituting $z = \zeta s' - \zeta r$ gives the representation in the theorem for $z \in [-\zeta s', \zeta s']$. Plugging in $K^*$ as given in eqn (11) in eqn (10) yields immediately

$$
\tau(K^*) = \frac{\int_0^\tau K^*(\zeta(s') - r) \int_0^\tau m_0(t) \, dt \, dr}{\int_0^\tau K^*(\zeta(s') - r) \, dr} = \frac{\int_0^\tau (\int_0^\tau m_0(t) \, dt)^2 \, dr}{\int_0^\tau \int_0^\tau m_0(t) \, dt \, dr}.
$$

Therefore, the assertion for the optimal asymptotic normed delay follows.

ACKNOWLEDGEMENTS

The author is grateful to an anonymous referee and an associate editor for valuable suggestions which improved the paper. The support of the Deutsche Forschungsgemeinschaft (DFG) (SFB 475, Reduction of Complexity in Multivariate Data Structures) is acknowledged.

NOTE

Corresponding author: Ansgar Steland, Ruhr-Universität Bochum, Fakultät für Mathematik, Mathematik 3 NA 3/71, Universitätsstr. 150, D-44780 Bochum, Germany. E-mail: ansgar.steland@rub.de

REFERENCES

BAI, J. and PERRON, P. (1998) Estimating and testing linear models with multiple structural changes. *Econometrica* 66, 47–78.

BASRAK, B., DAVIS, R. A. and MIKOSCH, T. (2002) Regular variation of GARCH processes. *Stochastic Processes and Their Application* 99, 95–115.

BIERENS, H. J. (1997) Testing the unit root with drift hypothesis against nonlinear trend stationarity, with an application to the US price level and interest rate. *Journal of Econometrics* 81, 29–64.

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

BREITUNG, J. (2002) Nonparametric tests for unit roots and cointegration. *Journal of Econometrics* 108, 343–63.

BROCKWELL, P. J. and DAVIS, R. A. (1991) *Time Series: Theory and Methods*. Springer Series in Statistics, New York.

BRODSKY, B. E. and DARCOVSKY, B. S. (1993) *Nonparametric Methods in Change-Point Problems*. Mathematics and its Applications 243, Kluwer Academic Publishers, Dordrecht.

BRODSKY, B. E. and DARCOVSKY, B. S. (2000) *Non-Parametric Statistical Diagnosis Problems and Methods*. Mathematics and its Applications, 509, Kluwer Academic Publishers, Dordrecht.

CHU, C. K., GODTLIEBEN, F. and MARRON, J. S. (1998) Edge-preserving smoothers for image processing. *Journal of the American Statistical Association* 93, 526–56.

DAVIDSON, J. (2002) Establishing conditions for the functional central limit theorem in nonlinear and semiparametric time series processes. *Journal of Econometrics* 106, 243–69.

DICKEY, D. A. and FULLER, W. A. (1979) Distribution of the estimates for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74, 427–31.

FAMA, E. and FRENCH, K. (1988) Permanent and temporary components of stock prices. *Journal of Political Economy* 96, 246–73.

FERGER, D. (1994a) On the power of nonparametric change-point tests. *Metrika* 41, 277–92.
FERGER, D. (1994b) An extension of the Csörgö-Horváth functional limit theorem and its applications to changepoint problems. *Journal of Multivariate Analysis* 51, 338–51.
FERGER, D. (1994c) Nonparametric detection of changepoints for sequentially observed data. *Stochastic Processes and their Application* 51, 359–72.
FERGER, D. (1995) Nonparametric tests for nonstandard change-point problems. *The Annals of Statistics*, 23, 1848–61.
FERGER, D. (1996) On the asymptotic behavior of change-point estimators in case of no change with applications to testing. *Statistics and Decision* 14, 137–43.
FERGER, D. (1997) Optimal bounds for the Prokhorov distance of the Miller-Sen process and Brownian motion. *Theory of probability and its Applications* 42, 155–62.
FRENCH, K. and ROLL, R. (1986) Stock return variances: the arrival of information and the reactions of traders. *Journal of Financial Economics* 17, 5–26.
GASSER, T. and Rafałowicz, E. and JENNE-N-STEINMETZ, C. (1986) Residual variance and residual pattern in nonlinear regression. *Biometrika* 73, 625–33.
GHYSELS, E., GUAY, A. and HALL, A. (1997) Predictive tests for structural change with unknown breakpoint. *Journal of Econometrics* 82, 209–33.
GOMBAY, E. and HORVÁTH, L. (1995) An application of the U-statistics to change-point analysis. *Acta Scientiarum Mathematicarum (Szeged)* 60, 345–57.
HERRNDORF, N. (1985) A functional central limit theorem for strongly mixing sequences of random variables. *Zeitschrift für Wahrscheinlichkeitstheorie and verwandte Gebiete* 69, 541–50.
HORVÁTH, L. and ŠKOVÁČ, M. (2003) Testing for changes using permutations of U-statistics. *Journal of Statistical Planning and Inference* 128, 351–71.
HUŠKOVÁ, M. and SLABÝ, A. (2001) Permutation tests for multiple changes. *Kybernetika* 37(5), 605–622.
JEGADEESH, N. (1991) Seasonality in stock price mean reversion: evidence from the U.S. and U.K. *Journal of Finance* 46, 1427–44.
KIM, J. H. and HART, J. D. (1998) Test for a change in a mean function when the data are dependent. *Journal of Time Series Analysis* 19, 399–424.
LIPSITS, M. A. (1982) On the absolute continuity of distributions of functionals of random processes. *Theory of probability and its applications* 27, 600–7.
LO, A. and MACKINLAY, A. C. (1988) Stock market prices do not follow random walks: evidence from a simple specification test. *Review of Financial Studies* 1, 41–66.
PASZEK, M. and Rafałowicz, E. and JENNE-N STEINMETZ, C. (2000) Vertically weighted regression – a tool for nonlinear data analysis. *Journal of the German Statistical Association* 84, 367–88.
Pawlak, M. and Rafałowicz, E. (2001) Jump preserving signal reconstruction using vertical weighting. *Nonlinear Analysis* 47, 327–38.
Pawlak, M., Rafałowicz, E. and Steland, A. (2004) On detecting jumps in time series: Nonparametric setting. *Journal of Nonparametric Statistics* 16, 329–47.
PHILLIPS, P. C. B. and PERRON, P. (1988) Testing for a unit root in time series regression. *Biometrika* 75, 335–48.
POLLARD, D. (1984) *Convergence of Stochastic Processes*. Springer, New York.
POTERBA, J. and SUMMERS, L. (1988) Mean reversion in stock prices: evidence and implications. *Financial Economics* 76, 1142–51.
RUE, H., CHU, C. K., GODTLIEBSEN, F. and MARRON, J. S. (2002) M-smoother with local linear fit. *Journal of Nonparametric Statistics* 14, 155–68.
SCHMID, W. and STELAND, A. (2000) Sequential control of non-stationary processes by nonparametric kernel control charts. *Journal of the German Statistical Association* 84, 315–36.
SHORACK, G. R. and WELLNER, J. A. (1986) *Empirical Processes With Applications to Statistics*. Wiley, New York.
SIEGMUND, D. (1985) *Sequential Analysis*. Springer, New York.
STELAND, A. (2002a) Sequential credit risk: can we benefit from sequential nonparametric control? *Selected Papers Operation Research Proceedings* 2002, 337–44.
STELAND, A. (2002b) A Bayesian view on detecting drifts by nonparametric methods. *Economic Quality Control* 17, 177–86.
STELAND, A. (2002c) Nonparametric monitoring of financial time series by jump-preserving estimators. *Statistical Papers* 43, 361–77.
STELAND, A. (2004a) Jump-preserving monitoring of dependent time series using pilot estimators. *Statistics and Decision* 21, 343–66.

© Blackwell Publishing Ltd 2005
Steland, A. (2004b) Sequential control of time series by functionals of kernel-weighted empirical processes under local alternatives. *Metrika* 60, 229–49.

Steland, A. (2005a) On the distribution of the clipping median under a mixture model. *Statistics and Probability Letters* 70, 1–13.

Steland, A. (2005b) Optimal sequential kernel smoothers under local nonparametric alternatives for dependent processes. *Journal of Statistical Planning and Inference* 132, 131–47.

Van der Vaart, A. and Wellner, J. A. (1996) *Weak Convergence of Empirical Processes*. Springer, New York.

Wald, A. (1947) *Sequential Analysis*. Wiley, New York.

Yakir, B., Krieger A. M. and Pollak, M. (1999) Detecting a change in regression: first-order optimality. *Annals of Statistics* 27, 1896–913.