Preprint, Institute of Statistics, RWTH Aachen University

A SURVEILLANCE PROCEDURE FOR RANDOM WALKS BASED ON LOCAL LINEAR ESTIMATION

Ansgar Steland

Institute of Statistics
RWTH Aachen University
Wüllnerstr. 3, D-52056 Aachen, Germany
steland@stochastik.rwth-aachen.de

We study the problem of detecting a change in the trend of time series whose stochastic part behaves as a random walk. Interesting applications in finance and engineering come to mind. Local linear estimation provides a well established approach to estimating level and derivative of an underlying trend function, provided the time instants where observations are available get dense asymptotically. Here we study the estimation principle for the classic time series setting where the distance between time points is fixed. It turns out that local linear estimation is applicable to our detection problem, and we identify the underlying (asymptotic) parameters. Assuming that observations arrive sequentially, we propose surveillance procedures and establish the relevant asymptotic theory, particularly, an invariance principle for the sequential empirical local linear process.

AMS 2000 subject classifications: Primary 62L10, 62E20; secondary 62G08, 60G50.

Keywords: Change-point, Non-standard asymptotics, Invariance principle, Nonparametric Regression, Sequential analysis, Weak dependence.

1. INTRODUCTION

This paper addresses the problem of detecting a change in the drift of a time series under the assumption that the series behaves as a random walk in large samples. We use sequential local linear estimation and associated surveillance procedures. The problem of detecting a change in the drift has several interesting applications, particularly, in econometrics and engineering. For example, economic theory and statistical analysis suggest that financial data such as log prices of assets, log exchange rates, log consumer price levels, and the great ratios are or could be random walks. We refer to the studies of Frankel [1986] and Edison [1987]. In engineering, random walks appear, e.g., as models for production processes affected by degradation, see Birnbaum and Saunders [1969], Doksum and Hoyland [1992] Taguchi [1985], and Taguchi et al. [1989]. The problem arises in computer science as well. For instance, in queuing models for devices as buffers (e.g. routers) in communication networks, the potential workload is a random walk with drift inducing the real workload process, cf. Whitt [2002]. Detecting changes in the drift can be used to define strategies to select the speed at which the buffer processes the input which reduces power consumption.

In some cases, it may be necessary to check the random walk assumption. For that purpose, various statistical tests have been proposed in the econometrics and statistics literature. We refer to Dickey and Fuller [1979], Kwiatkowski et al. [1992], and Breitung [2002], among others. Sequential monitoring procedures related to this problem have been recently investigated by Steland [2007*a*] and Steland [2007*b*]. For further references on these issues we refer to the cited articles.

Given that the random walk assumption for a time series is justified, we address the problem of detecting an unknown (non random) *change-point* where the drift is no longer given by a known model. Having in mind the above applications which provide sequential data streams, we are interested in sequential surveillance procedures which analyse at each time instant the available data to detect the change as quick as possible. Sequential detectors based on the Nadaraya-Watson estimator have been applied to this problem by Steland [2005]. That approach has the nice property that it covers an approximation of the classic EWMA control chart procedure, a common and easy-to-use tool in sequential analysis, as a special case. The well known boundary bias problem of the Nadaraya-Watson estimator in the regression setting suggests the use of procedures based on local linear estimation. Thus, the detectors studied here are based on the sequence of local linear estimates evaluated at the current time instant. Surveillance procedures based on that approach seem to date back to Schmid and Steland [2000], where they have been used to detect changes in the level and the slope of a process mean. Monte Carlo simulations revealed that these control charts have excellent detection properties, even when applied to stationary time series under conditional heteroskedasticity.

Local linear estimation is, nowadays, a well established approach to the nonparametric estimation of conditional means, and well studied in terms of its efficiency properties, see, e.g., Fan and Gijbels [1996] and Antoch et al. [2002]. It has recently been used by Grégoire and Hamrouni [2002] for a posteriori (off line) estimation of a jump point in a smooth curve. Here a random regression design is assumed and estimation is based on a process which converges to a compound Poisson process. The estimation of the location of isolated jumps in piecewise continuous regression functions is a problem that has been studied for quite a long time. We refer to the papers of Wu and Chu [1993] and Müller and Song [1997], among others. Müller and Stadtmüller [1999] and Horváth and Kokoszka [2002] investigated the problem of testing smoothness against the alternative of at least one jump for the fixed design i/n for the regressor. The i.i.d. regression setting was recently studied by Antoch et al. [2007] assuming the classic i.i.d. regression framework. For methods to detect changes in autoregressive time series confer Hušková et al. [2007, 2008].

In the present paper we study the asymptotic distribution theory of sequential local linear estimation for random walks, where no restrictive i.i.d. assumption is imposed upon the increments; the case of stationary processes will be studied in a separate paper. We assume the classic framework of time series analysis where the observations may be correlated and are sampled at discrete fixed time points, which do not get dense as the maximum sample size increases. At first glance, this seems to be odd, but our interest is not in consistent estimation of the mean. Instead, we address the question of sequentially detecting a change in the mean. For that purpose, consistent estimation is not required, and assuming an arbitrary nonparametric mean function, it is impossible within the classic time series framework. Assuming that the observations are taken at time points from an arbitrary fine grid with fixed mesh leads to interesting non-standard asymptotic laws, particularly for the random walk case studied here. Our simulation study presented below in detail indicates that the approximations obtained by these asymptotic laws are remarkably accurate even for very small sample sizes, which is beneficial for practical applications. Related results for procedures employing the Nadaraya-Watson estimator in the same framework have been obtained by Steland [2005]. Truncated change-point detection for renewal counting processes has been studied by Gut and Steinebach [2002].

From a theoretical viewpoint, the question arises as to whether the local linear approach is applicable at all, i.e., estimates well defined functional parameters of the trend function, asymptotically. The development of an appropriate asymptotic distributional theory matters as well. For that purpose, we consider a framework where the trend of the time series, given by a sequence resp. array of constants, is induced by an underlying trend function μ. We demonstrate that this framework allows to handle various change-point models for random walk time series which are of practical interest. In these models the underlying function μ has a simple and intuitive interpretation. It defines the change-point as well as the process mean (trend) after the change. The local linear estimation approach locally fits a straight line to the data. It turns out that, locally in large samples, the intercept and slope parameters are related to μ and its integral function, which justifies the application of local linear estimation in the present context. However, the proposed method itself does not require knowledge of μ as in many other approaches, e.g., likelihood based procedures.

We will represent the proposed detection procedures as functionals of an underlying bivariate sequential empirical process. The asymptotic theory given here deals with a new bivariate functional central limit theorem (FCLT) which describes the asymptotic behavior of that process, in distribution. The FCLT implies central limit theorems (CLTs) for the surveillance procedures. However, the interpretation of our results is not restricted to sequential analysis. The sequential empirical process represents the sequence of local linear estimates as a function on the unit interval $[0,1]$. When evaluated at $t = 1$, it coincides with the local linear estimates calculated from the full sample and calculated at the point n. Thus, the results of the present paper cover an interesting CLT for the classic local linear estimator at the boundary under non-standard conditions as well.

Surveillance procedures require to specify a control limit, which plays a similar role as the critical value in hypothesis testing. A common approach is to select the control limit to ensure that the average run length is larger or equal to a given constant, provided the null hypothesis holds. However, other design criteria can be chosen and have been studied in the literature, e.g., the point-wise false alarm rate, the median run length, or the type I error rate. Based on our asymptotic results, which provide an approximation for the unknown distribution of the detector, one can easily design a surveillance procedure according to various design criterion.

The paper is organized as follows. Section 2 introduces the model framework, the required assumptions, and discusses change-point models which are covered by our setting. Local linear estimation and the associated local linear detector are discussed in detail in Section 3. Section 4 presents the main theoretical results. We provide a rigorous justification of the local linear estimation principle for our setting and establish FCLTs for the sequential bivariate local linear processes. These results yield CLTs for the corresponding surveillance procedures. Detailed proofs are postponed to Section 6.

2. TIME SERIES MODEL AND ASSUMPTIONS

Let us now introduce the change-point model of the paper. Since we aim at studying the asymptotic distribution theory of the proposed method under local alternative models for the drift, we assume that we are given an array $\{Y_{Tt} : t = 1, \ldots, T,\ T \geq 1\}$ of real-valued random variables satisfying

$$Y_{Tt} = m_{Tt} + \epsilon_{Tt}, \qquad t = 1, \ldots, T,\ T \in \mathbb{N}, \tag{2.1}$$

where $\{\epsilon_{Tt}\}$ is an array of zero mean error terms. To phrase our assumptions on $\{\epsilon_{Tt}\}$, we need some notations. Let $\lfloor x \rfloor$ denote the largest integer which is smaller or equal to $x \in \mathbb{R}$. Notice that to each array, say, $\{Z_{Tt}\}$, of random variables we may attach a sequence of càdlàg processes, namely, $s \mapsto Z_{T,\lfloor Ts \rfloor}$, $s \in [0,1]$, $T \geq 1$. These processes attain values in the Skorohod space $D[0,1]$ of all right-continuous functions $[0,1] \to \mathbb{R}$ with existing left-hand limits. The weak convergence of a sequence of such processes will be denoted by $\Rightarrow$. For details we refer to Billingsley [1999].

The following assumption basically requires that the error terms behave asymptotically as a pure random walk.

Assumption (E): Assume $\{\epsilon_{Tt} : 1 \leq t \leq T, T \geq 1\}$ is an array of zero mean random variables with

$$T^{-1/2}\epsilon_{\lfloor Ts \rfloor} \Rightarrow \eta B(s), \tag{2.2}$$

as $T \to \infty$, for some constant $\eta \in (0, \infty)$.

Here and in what follows B denotes a standard Brownian motion (Wiener process).

REMARK 2.1 *Suppose that, in addition to Assumption (E), the first-order differences $\epsilon_{T,t} - \epsilon_{T,t-1}$ form an array of row-wise stationary sequences. Then $\{\epsilon_{T,t}\}$ satisfies a nonparametric definition for the notion of an integrated process of order* 1*.*

Notice that Assumption (E) is very weak and covers many important classes of time series. Let us consider the special case that the error terms form a pure random walk. That means, $\epsilon_{Tt} = \sum_{i=1}^{t} \xi_{Ti}$, where $\xi_{T1}, \ldots, \xi_{TT}$ are i.i.d. with $E(\xi_{Ti}) = 0$, for any $T \geq \mathbb{N}$. By virtue of Donsker's theorem, Assumption (E) is satisfied. However, (2.2) holds under much more general conditions on the increments. For example, if $\{\xi_{Ti}\}$ is a martingale difference scheme satisfying a (conditional) Lindeberg condition, or if $\xi_{Ti} = \xi_i$, $i \in \mathbb{N}$, $T \in \mathbb{N}$, is α-mixing with mixing coefficients which decrease sufficiently fast. For details we refer to Durrett [2005].

The constants m_{Tt} in Eq. (2.1) represent the trend and are assumed to be generated as follows.

Assumption (M): Suppose that

$$m_{Tt} = T^{1/2} \int_0^{t/T} \mu(r)\, dr + o(T^{1/2}), \qquad 0 \leq t \leq T,\ T \geq 1,$$

for some function $\mu : [0,1] \to \mathbb{R}$ with $\int_0^1 |\mu(r)|\, dr < \infty$.

It is worth noting that the function μ plays the role of the derivative of the trend, and its integral yields the level. At first glance, the scaling with $T^{1/2}$ and the remainder $o(T^{1/2})$ may look strange, but our asymptotic results imply that these are the correct rates yielding non-trivial limits depending on the μ.

Let us discuss two examples.

EXAMPLE 2.1 *Suppose* $\{m_{Tt}\}$ *satisfies (M) with* $\mu(r) = \mu_0$, $r \in [0,1]$, *for some constant* $\mu_0 \in \mathbb{R}$. *Obviously,* $m_{T,\lfloor Ts \rfloor} = \frac{\lfloor Ts \rfloor}{\sqrt{T}} \mu_0$, *and due to Assumption (E)*

$$\frac{1}{\sqrt{T}} Y_{T,\lfloor Ts \rfloor} = \frac{\lfloor Ts \rfloor}{T} \mu_0 + \frac{1}{\sqrt{T}} \epsilon_{T,\lfloor Ts \rfloor} \Rightarrow s\mu_0 + \eta B(s),$$

as $T \to \infty$.

More important are random walks with drift and correlated increments.

EXAMPLE 2.2 *Suppose that* $Y_{T,t}$ *is a random walk,*

$$Y_{Tt} = \sum_{i=1}^{t} \xi_{Ti}, \qquad 1 \le t \le T, \ T \ge 1, \tag{2.3}$$

with increments given by

$$\xi_{Tt} = \frac{1}{\sqrt{T}} \mu\left(\frac{t}{T}\right) + \zeta_{Tt}, \tag{2.4}$$

where the array $\{\zeta_{Tt}\}$ *satisfies Assumption (E). If we put* $\epsilon_{T,t} = \zeta_{T,t}$ *and define*

$$m_{T,t} = \frac{1}{\sqrt{T}} \sum_{i=1}^{t} \mu\left(\frac{i}{T}\right)$$

for $1 \le t \le T$, $T \in \mathbb{N}$, *then* $E(Y_{Tt}) = m_{Tt}$, *i.e., model equation (2.1) as well as Assumptions (E) and (M) hold. Further, in this case we have*

$$\begin{aligned} \frac{1}{\sqrt{T}} Y_{T,\lfloor Ts \rfloor} &= \frac{1}{\sqrt{T}} m_{T,\lfloor Ts \rfloor} + \frac{1}{\sqrt{T}} \epsilon_{T,\lfloor Ts \rfloor} \\ &\Rightarrow \int_0^s \mu(r)\,dr + \eta B(s), \end{aligned} \tag{2.5}$$

as $T \to \infty$.

More generally, the weak convergence result in (2.5) holds under Assumptions (E) and (M).

LEMMA 2.1 *Given Assumptions (E) and (M),*

$$\frac{1}{\sqrt{T}} Y_{T,\lfloor Ts \rfloor} \Rightarrow \int_0^s \mu(r)\,dr + \eta B(s),$$

as $T \to \infty$.

REMARK 2.2 *Notice that the limit process can be written as* $\int_0^s \mu(r)\,dr + \int_0^s \sigma(r)\,dB(r)$, *if* $\sigma(r) = \eta$, $r \in [0,1]$, *thus belonging to the class of Itô processes with deterministic drift and constant volatility.*

REMARK 2.3 *It is worth mentioning that Assumptions (E) and (M) are satisfied for discretely observed continuous time processes of the class*

$$\mathcal{Z}(t) = \int_0^t \mu(r)\,dr + \sigma B(t), \qquad t \in [0,1],$$

where the drift $\mu \in L_1$ *is deterministic and* $\sigma \in (0,\infty)$ *a constant. Suppose that* $\mathcal{Z}$ *is observed at the grid* $t/T, t = 0,\dots,1$, $T \geq 1$. *Then the array* $Z_{T,t} = T^{1/2}\mathcal{Z}(t/T)$ *clearly satisfies our assumptions.*

Let us now consider some change-point models of interest.

EXAMPLE 2.3 (LINEAR DRIFT AFTER AN UNKNOWN CHANGE-POINT)
This model corresponds to the choice

$$\mu(s) = \mathbf{\Delta 1}_{[\vartheta,1]}(s), \qquad s \in [0,1], \tag{2.6}$$

for some constant $\Delta \neq 0$ *and a fixed but unknown change-point parameter* $\vartheta \in (0,1)$, *yielding*

$$m_{Tt} = \begin{cases} 0, & 0 \leq t < T\vartheta, \\ \sqrt{T}\left(\frac{t}{T} - \vartheta\right)\Delta, & T\vartheta \leq t \leq T. \end{cases}$$

for $t = 1,\dots,T$. *This means, for large* T *the fraction of pre-change mean-zero observations is approximately* ϑ, *and after the change-point there is a linear drift in the data, which is determined by the parameter* Δ.

Generalizing to multiple change-points of this type yielding a piecewise linear function underlying the sequence m_{Tt} is straightforward. The next example provides a change-point model for the setting studied in Example 2.2.

EXAMPLE 2.4 (CHANGE-POINT MODEL FOR THE DRIFT OF A RANDOM WALK)
Consider again the random walk model (2.3) with increments ξ_{Tt} *given by (2.4). Then*

$$\xi_{Ti} = \begin{cases} \zeta_{Ti}, & 0 \leq i < T\vartheta, \\ \frac{1}{\sqrt{T}}\mu(i/T) + \zeta_{Ti}, & T\vartheta \leq i \leq T. \end{cases}$$

Further, suppose

$$\mu(s) = 0,\ s \in [0, \vartheta), \quad \mu(s) > 0,\ s \in [\vartheta, \vartheta + \varepsilon),\ \textit{for some } \varepsilon > 0, \quad \int_\vartheta^1 |\mu(s)|\, ds < \infty.$$

Now the ξ_{Ti}s have mean zero before the change, and after the change their expectation equals $T^{-1/2}\mu(i/T)$. Hence $m_{Tt} = 0$ for $0 < t < T\vartheta$, and $m_{Tt} = T^{-1}\sum_{i \geq T\vartheta}^{t} \mu(i/T)$, if $T\vartheta \leq t \leq T$, with $m_{T,\lfloor T\vartheta \rfloor} > 0$ by assumption. Again, we obtain $Y_{T,\lfloor Ts \rfloor} \Rightarrow \int_\vartheta^s \mu(r)\, dr + \eta B(s)$, as $T \to \infty$.

In general, a change-point model for the trend can be phrased as

$$\mu(t) = \begin{cases} \mu_0(t), & \text{if } t < \vartheta\ , \\ \mu_0(t) + \Delta(t), & \text{if } \vartheta \leq t, \end{cases}$$

for $t \in [0, 1]$, where μ_0 denotes the in-control model (H_0) and Δ represents the departure from it under the out-of-control (alternative) model, i.e., after the change-point $\lfloor T\vartheta \rfloor$, which is parameterized by $\vartheta \in [0, 1]$. To guarantee Assumption (M) suppose that $\int |\mu_0(s)|\, ds,\ \int |\Delta(s)|\, ds < \infty$.

The aim is to sequentially test the null hypothesis $H_0 : \mu = \mu_0$ against the alternative hypothesis

$$H_1 : \exists \vartheta \in (0, 1) : \mu = \mu_0 \text{ on } [0, \vartheta) \text{ and } \mu = \mu_0 + \Delta \text{ on } [\vartheta, 1],$$

by means of a surveillance procedure (stopping time.)

REMARK 2.4 *Notice that w.l.o.g. one may assume $\mu_0 = 0$. Indeed, otherwise consider the transformed observations*

$$\widetilde{Y}_{Tt} = Y_{Tt} - \sqrt{T} \int_0^{t/T} \mu_0(r)\, dr, \qquad 1 \leq t \leq T,\ T \geq 1.$$

Now the r.v.s $\{\widetilde{Y}_{Tt}\}$ satisfy Assumption (M) with μ replaced by $\mu - \mu_0$, which vanishes under H_0.

3. SEQUENTIAL LOCAL LINEAR ESTIMATION AND RELATED SURVEILLANCE PROCEDURES

In the regression setting local polynomial fitting, studied by Stone [1977], Cleveland [1979], Tsybakov [1986] and Masry and Fan [1997], among others, reduces the bias of the Nadaraya-Watson estimator and adapts automatically to the boundary of design points. Here we examine a sequential version assuming the time series setting. The basic idea of local linear estimation is as follows. The trend, which is a function $m_T(t) = m_{Tt}$ of t, is approximated by a linear function, locally at the current time instant $t_n = n \in \mathbb{N}$. That is, one assumes the validity of the approximation

$$m_T(s) \approx \beta_{0n}(t_n) + \beta_{1n}(t_n)(s - t_n)$$

with (local) intercept $\beta_{0n} = \beta_{0n}(t_n)$ and slope $\beta_{1n} = \beta_{1n}(t_n)$. In the next section we provide an asymptotic justification of that approximation and identify the parameters β_{0n} and β_{1n} and their limits.

Now one proceeds by fitting the local model to the data using weighted least squares where the weights ensure that squared residuals corresponding to observations near the current time t_n dominate the objective function. We shall use kernel weights

$$w_{ni} = K_h(t_i - t_n) / \sum_{j=1}^{n} K_h(t_j - t_n), \qquad i = 1, \ldots, n,$$

for some given smoothing kernel function $K : \mathbb{R} \to \mathbb{R}$. Notice that one may assume that K is symmetric. $h > 0$ is a bandwidth and $K_h(z) = K(z/h)/h$, $z \in \mathbb{R}$, is the rescaled kernel. For sake of simplicity of exposition we omit the dependence of w_{ni} on h in our notation.

3.1. *Assumptions on the kernel and the bandwidth*

Concerning the bandwidth we use the following condition.

Assumption (H). The bandwidth h is chosen as function of T, i.e., $h = h_T$ such that

$$\left| \frac{T}{h} - \xi \right| = O(T^{-1}). \tag{3.1}$$

for some known constant $\xi \in [1, \infty)$.

Particularly, one may put $h = T/\xi$. Before discussing this assumption, let us consider the rather weak assumptions on the smoothing kernel K.

Assumption (K). Suppose that K is Lipschitz continuous, bounded, i.e., $\|K\|_\infty < \infty$, and positive on $(-\xi, \xi)$.

Common choices for K are bounded probability densities with mean 0, unit variance and the additional property that $K(z)$ is decreasing in $|z|$.

Notice that our approach for bandwidth selection differs from the usual conditions in the literature aiming at estimation (conditional) means where one assumes $h \to 0$ and $Th \to \infty$. To discuss this issue, let us assume that K has support $[-1, 1]$. Then for estimation of $m(t_i)$ only observations Y_j with $|t_j - t_i| \leq h \to 0$ are used. The asymptotic theory for consistency of local linear smoothers, which can be written as kernel smoothers w.r.t.

to a specific smoothing kernel, relies on the following facts. Due to Parzen's lemma the expectation of

$$\widetilde{m}_n(t) = \frac{1}{nh} \sum_{i=1}^{n} K((t - t_i)/h) Y_i,$$

which is equal to

$$E(\widetilde{m}_n(t)) = \int h^{-1} K((t - z)/h) m(z) dF(z),$$

if $t_i \sim F$, converges to $m(t)$, as $h \to 0$. However, the variance of $\widetilde{m}_n(t)$ is $O((nh)^{-1})$, which yields the constraint $nh \to \infty$. Consequently, in this setting one has to assume that the data get dense as $T \to \infty$, thus requiring an appropriate sampling design (fixed design) or random sampling w.r.t. to a positive density.

However, when detection of changes for time series is the goal, consistency is not a must and in many cases the assumption that the data get dense as the sample size increases, is not realistic. When assuming that the observations are observed at deterministic fixed time points not depending on T, it makes sense to choose the bandwidth h proportional to T to ensure that the number of observations available during $[t - h, t + h]$ tends to ∞, as $T \to \infty$. This justifies our Assumption (3.1), which additionally imposes the convergence rate T^{-1}, ruling out artificial choices as $h = T/\xi + T^{-1/2}$. Finally, Assumption (H) is also used in Aue et al. [2009] to obtain asymptotic results for stopping times based on moving sum detectors when the errors satisfy a FCLT.

It is also worth mentioning that our simulation study presented below shows that the setting studied in this paper provides accurate approximations even for small sample sizes. Thus, from an applied viewpoint, our non-standard approach proves to be useful. However, data-adaptive bandwidth selection for detection procedures remains an interesting and important issue which will be studied in future papers.

3.2. *Sequential local linear estimation and related detectors*

To simplify exposition we shall omit the dependence of Y_t on T in our notation. Let us introduce the inner product

$$(\mathbf{x}, \mathbf{y})_n = \mathbf{x}' \mathbf{W}_n \mathbf{y}, \qquad \mathbf{x}, \mathbf{y} \in \mathbb{R}^n,$$

and denote the associated norm by $\|\mathbf{x}\|_n = \sqrt{(\mathbf{x}, \mathbf{x})_n}$, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. The minimizers of the objective function

$$\sum_{i=1}^{n} w_{ni} (Y_i - \beta_0 - \beta_1 (t_i - t_n))^2, \qquad (\beta_0, \beta_1) \in \mathbb{R}^2,$$

are given by

$$\widehat{\beta}_{1n} = \frac{(\mathbf{Y}_n - \overline{Y}_n \mathbf{1}_n, \mathbf{t}_n - \bar{t}_n \mathbf{1}_n)_n}{\|\mathbf{t}_n - \bar{t}_n \mathbf{1}_n\|_n^2},$$
$$\widehat{\beta}_{0n} = \overline{Y}_n - \widehat{\beta}_{1n} \bar{t}_n,$$

where, with some abuse of notation, $\overline{Y}_n = \sum_{i=1}^n w_{ni} Y_i$, $\bar{t}_n = \sum_{i=1}^n w_{ni} t_i$, $\mathbf{1}_n = (1, \dots, 1)' \in \mathbb{R}^n$, and

$$\mathbf{Y} = (Y_1, \dots, Y_n)',$$
$$\mathbf{t}_n = (t_1 - t_n, \dots, t_{n-1} - t_n, 0)',$$
$$\mathbf{W}_n = \operatorname{diag}(w_{n1}, \dots, w_{nn}).$$

The proposed local linear detectors are based on the sequential empirical processes associated to the sequences $\widehat{\beta}_i(n)$, $n = 2, \dots, T$, $i = 0, 1$. For $s \in [0, 1]$ define

$$\widehat{\beta}_{0T}(s) = \widehat{\beta}_{0,\lfloor Ts \rfloor}, \qquad \text{and} \qquad \widehat{\beta}_{1T}(s) = \widehat{\beta}_{1,\lfloor Ts \rfloor}.$$

To ensure that the procedure is based on a sufficient amount of information, we define the start of monitoring, denoted by k, as a fraction of the maximal sample size T, i.e., we put

$$k = \lfloor T\kappa \rfloor$$

for some $\kappa \in (0, 1)$.

To detect deviations of the process mean from an assumed in-control model $\mu_0 = 0$ corresponding to the null hypothesis $H_0 : \mu = 0$, consider the stopping rule

$$L_T = L_T(c_L) = \inf\{k \le n \le T : T^{-1/2} \widehat{\beta}_{0T}(n/T) > c_L\} \tag{3.2}$$

for some control limit c_L. The scaling factor $T^{-1/2}$ is, indeed, the correct convergence rate when the errors form a random walk, as will be shown in the next section.

The control limit should be selected to ensure that the procedure has well defined statistical properties in large samples. To simplify exposition and due to its ease of interpretation, we consider the case that the control limit is chosen to guarantee an asymptotic nominal type I error rate, i.e.,

$$\lim_{T \to \infty} P_0(L_T(c_L) < T) = \alpha \tag{3.3}$$

for some given $\alpha \in (0, 1)$. Here, P_0 indicates that the probability is calculated assuming $\mu = \mu_0$. The asymptotic results given in the next section indeed allow us to select c_L, such that (3.3) is satisfied.

One may also study a surveillance procedure which gives a signal if the estimated slope parameter is too large. For that purpose consider

$$S_T = S_T(c_S) = \inf\{k \le n \le T : T^{-1/2}\widehat{\beta}_{1T}(n/T) > c_S\},$$

for a control limit c_S. Clearly, the above discussion on the selection of the control limit to achieve a given nominal type I error rate applies.

The results of the following section will show that L_T is suited to detect changes in terms of $\int_\vartheta^t \Delta(s)\,ds$, whereas S_T addresses $\Delta(t)$ itself.

4. ASYMPTOTIC THEORY

In this section, we provide an asymptotic justification of the assumption of a locally linear trend, which underlies the local linear estimation approach. Further, we provide an invariance principle for the local linear processes, which will imply central limit theorems for the proposed detectors. Our main result holds under Assumption (M) on the deterministic drift present in the time series $\{Y_{Tt}\}$, which is induced by the function μ. Having in mind the change-point models given above, the choice $\mu = 0$ yields the FCLT under the in-control model, i.e., when there is no change in the series. If $\mu \neq 0$, the FCLT provides the asymptotic properties under that local alternative.

4.1. *Asymptotic linear representation*

Let us start with the following result yielding an asymptotic linear representation of the array $\{m_{Tt}\}$ of process means, which justifies the local linear estimation approach and identifies the (asymptotic) parameters. Given the framework of the present paper, it serves as a substitute of the usual assumption that the mean to be estimated can be locally approximated by a linear function. Recall that n is interpreted as the current time instant.

LEMMA 4.1 *Given Assumption (M), the array $\{m_{Tt}\}$ has the following property. For $T \ge 1$ and $1 \le n \le T$*

$$m_{Tt} = \beta_{0T}(n) + \beta_{1Tt}(n)\left(\frac{t}{T} - \frac{n}{T}\right) + o(T^{1/2}), \tag{4.1}$$

where

$$\beta_{0T}(n) = \sqrt{T}\int_0^{n/T} \mu(r)\,dr,$$
$$\beta_{1Tt}(n) = \sqrt{T}\mu(\xi^*_{Tnt}),$$

for some ξ^*_{Tnt} between n/T and t/T.

REMARK 4.1 *(i) Notice that $\beta_{0T}(n)$ recovers the higher order term in the trend m_{Tn}, and $\beta_{1T}(n)$ its 'derivative'.*
(ii) If $\mu \in C^1([0,1])$, one may put $\beta_{1Tt}(n) = \sqrt{T}\mu(n/T)$.

As a corollary, we obtain the following result, which identifies the asymptotic behavior of β_{0T} and β_{1Tt}, locally near the current time instant, i.e., for $t \approx n$, with $n \sim T \to \infty$. It also shows that (4.1) indeed yields an asymptotic linear representation of the process mean.

COROLLARY 4.1 *Given Assumption (M), we have uniformly for $s \in [0,1]$*

$$T^{-1/2}\beta_{0T}\left(\frac{\lfloor Ts \rfloor}{T}\right) \to \int_0^s \mu(r)\,dr,$$

as $T \to \infty$, and

$$T^{-1/2}\beta_{1Tt}\left(\frac{\lfloor Ts \rfloor}{T}\right) \to \mu(s),$$

provided $t/T \to s$, as $T \to \infty$. The latter convergence is uniform in $s \in [0,1]$ if μ is continuous

4.2. *Asymptotic distributions*

The main theoretical result of this paper is the following FCLT which provides the joint asymptotic law of the local linear estimates, studied as a pair of (sequential) càdlàg processes. Establishing the joint asymptotic law particularly allows us to construct detection procedures depending on both the intercept and slope estimates as control statistics.

THEOREM 4.1 *Given Assumptions (E), (H), (K) and (M), we have*

$$\begin{pmatrix} T^{-1/2}\widehat{\beta}_{0T}(s) \\ T^{-1/2}\widehat{\beta}_{1T}(s) \end{pmatrix} \Rightarrow \begin{pmatrix} \mathcal{Z}(s) - C^{-1}(s)\mathcal{A}(s)D(s) \\ C^{-1}(s)\mathcal{A}(s) \end{pmatrix},$$

jointly as $T \to \infty$, where

$$D(s) = \xi \int_0^s K_s(\xi(r-s))r\,dr, \tag{4.2}$$

$$C(s) = \xi \int_0^s K_s(\xi(r-s))W_\xi(r,s)\,dr, \tag{4.3}$$

$$W_\xi(r,s) = \xi(r-s) - \xi \int_0^s K_s(\xi(z-s))z\,dz. \tag{4.4}$$

Further, the processes $\mathcal{A}(s)$ *and* $\mathcal{Z}(s)$ *are given by*

$$\mathcal{A}(s) = \xi \int_0^s K_s(\xi(r-s)) \left(\mathcal{Y}(r) - \xi \int_0^s K_s(\xi(z-s)) \mathcal{Y}(z)\, dz \right) W_\xi(r,s)\, dr$$

and

$$\mathcal{Z}(s) = \xi \int_0^s K_s(\xi(r-s)) \mathcal{Y}(r)\, dr,$$

respectively, with

$$\mathcal{Y}(s) = \int_0^s \mu(r)\, dr + \eta B(s),$$

for $s \in [0,1]$. *In these formulas,* $K_s(z) = K(z) / \int_0^s \xi K(\xi(r-s))\, dr$, *for* $z \in \mathbb{R}$ *and* $s \in [\kappa, 1]$.

Central limit theorems for the stopping times proposed above are now straightforward.

COROLLARY 4.2 *Given Assumptions (E), (H), (K) and (M), we have joint convergence in distribution of*

$$L_T/T \xrightarrow{d} \inf\{ s \in [\kappa, 1] : C^{-1}(s) \mathcal{A}_0(s) D(s) > c_L \},$$

and

$$S_T/T \xrightarrow{d} \inf\{ s \in [\kappa, 1] : C^{-1}(s) \mathcal{A}_0(s) > c_S \}$$

as $T \to \infty$, *where* $\mathcal{A}_0$ *denotes the process defined in Theorem 4.1 with* $\mu = 0$.

These results can be used to determine control limits ensuring, e.g., that the procedure attains a specific type I error rate, asymptotically. We study the accuracy of the resulting approximations to some extent in the next section.

5. SIMULATIONS

We conducted a simulation study to examine the accuracy of the distributional approximation obtained by central limit for the stopping time L_T as well as the resulting detection performance. Our focus is on the influence of the bandwidth selection on the statistical properties of the method, which corresponds to the choice of ξ, the limit of T/h. Further, we compared the detector based on the local linear approach with a classic CUSUM procedure which is known to be powerful to detect changes in the mean and often used in quality control.

Time series were sampled according to the model

$$Y_n = \sum_{i=1}^{n} (\Delta \mathbf{1}_{\{i \geq q\}} + \epsilon_i), \qquad n = k, \ldots, T,$$

with a shift $\Delta \in \{0.2, 0.4, 0.6\}$ after the change-point $q = \lfloor T\vartheta \rfloor$ with $\vartheta \in \{\infty, 0.25, 0.5, 0.75\}$ and i.i.d. standard normal error terms ϵ_t. Thus, we examined early changes as well as late changes of various magnitude. We were interested in analysing the small sample case and selected $T = 100$ and $T = 200$, respectively, for our study.

To these samples the level detector L_T was applied using a Gaussian kernel and the bandwidth $h = T/\xi$ with $\xi \in \{10, 20, 40\}$. Detection was started at the 25th observation.

We compared our proposal to a simple and commonly applied procedure which would be a reasonable candidate among the classic detection procedures, namely a CUSUM control chart. Specifically, we applied a one-sided CUSUM chart with reference value $K = 0.5$ and control limit $H = 6.5$ given by

$$\min\{k \leq n \leq T : C_n > H\} \qquad C_n = \max(0, \Delta Y_n - (\mu_0 + K) + C_{n-1}),\ C_0 = 0,$$

where $\Delta Y_n = Y_{n-1} - Y_n$. The CUSUM chart is usually designed to monitor an infinite series with a certain in-control ARL. However, the above CUSUM chart attains a comparable type I error rate for $T = 250$, thus allowing comparisons.

Table I provides the results of 20,000 independent replications for each entry.

It can be seen that the proposed approximations by the distributional limit process based on our functional central limit theorems provide accurate approximations even for very small sample sizes as investigated here. To detect a linear trend in a random walk smaller values of ξ seem to be preferable. However, the loss in detection power due to larger values, which yield a smaller bias, is minor. Having this mind, one can select rather small bandwidths h. A discussion of more refined bandwidth selection strategies, e.g. data-dependent rules, is, however, beyond the scope of the present paper and will be addressed in greater detail in future work.

The results from the simulation study also indicate that our method based on sequential local linear estimation yields substantially higher detection rates compared to the CUSUM method except for very later changes when the shift is large ($\vartheta = 0.75$ and $\Delta = 0.6$ in our simulation study); here the CUSUM chart performs better. In all other settings, which are more important for applications, the new method substantially outperforms the CUSUM chart.

6. PROOFS

This section is devoted to rigorous proofs of the results presented in Section 4.

TABLE I

Simulated performance of the local linear detector and comparison with a CUSUM procedure.

T	ϑ	Δ	ξ			CUSUM
			10	20	40	
100	1	0.1	0.072	0.068	0.062	0.021
	0.25	0.1	0.22	0.206	0.187	0.054
		0.2	0.483	0.451	0.438	0.128
		0.4	0.914	0.91	0.909	0.514
		0.6	0.998	0.998	0.997	0.918
	0.5	0.1	0.152	0.141	0.125	0.039
		0.2	0.283	0.262	0.25	0.090
		0.4	0.644	0.619	0.604	0.368
		0.6	0.915	0.91	0.894	0.785
	0.75	0.1	0.097	0.092	0.087	0.03
		0.2	0.135	0.121	0.115	0.049
		0.4	0.263	0.243	0.24	0.184
		0.6	0.449	0.422	0.411	0.477
250	1	0.1	0.071	0.066	0.059	0.052
	0.25	0.1	0.359	0.328	0.321	0.130
		0.2	0.784	0.757	0.752	0.300
		0.4	0.999	0.998	0.998	0.850
		0.6	1	1	1	0.999
	0.5	0.1	0.227	0.2	0.186	0.108
		0.2	0.481	0.462	0.445	0.228
		0.4	0.931	0.931	0.916	0.720
		0.6	0.999	0.998	0.999	0.988
	0.75	0.1	0.126	0.104	0.097	0.081
		0.2	0.198	0.178	0.168	0.147
		0.4	0.459	0.44	0.421	0.461
		0.6	0.76	0.739	0.711	0.867

PROOF OF LEMMA 4.1 AND COROLLARY 4.1. Fix $1 \le n \le T$ and note that by Assumption (M)

$$m_{Tt} = \sqrt{T} F_\mu(t/T) + o(T^{1/2}),$$

where $F_\mu(x) = \int_0^x \mu(r)\, dr$, $x > 0$. Using the fact that there exist ξ^*_{Tnt} between t/T and n/T such that

$$F_\mu(t/T) = F_\mu(n/T) + \mu(\xi^*_{Tnt})(t/T - n/T)$$

we obtain

$$m_{Tt} = \sqrt{T} F_\mu(n/T) + \sqrt{T}\mu(\xi^*_{Tnt})(t/T - n/T) + o(T^{1/2}),$$

which proves Lemma 4.1. If we put $n = \lfloor Ts \rfloor$, $s \in [0,1]$,

$$\beta_{0T}(\lfloor Ts \rfloor) = \sqrt{T} \int_0^{\lfloor Ts \rfloor / T} \mu(r)\, dr, \qquad \beta_{1Tt}(\lfloor Ts \rfloor) = \sqrt{T}\mu(\xi^*_{T,\lfloor Ts \rfloor,t}),$$

where $\xi_{T,\lfloor Ts \rfloor,t} \to 0$. Thus, $\sqrt{T}\beta_{0T}(\lfloor Ts \rfloor) \to \int_0^s \mu(r)\, dr$, as $T \to \infty$. Further, $\xi_{T,\lfloor Ts \rfloor,t} \to s$, if $t/T \to s$, as $T \to \infty$, yielding $T^{-1/2}\beta_{1Tt}(\lfloor Ts \rfloor) \to \mu(s)$ under these conditions. $\square$

We are now in a position to prove the invariance principle for the local linear process under the random walk change-point model.

PROOF OF THEOREM 4.1. Let us first consider $\widehat{\beta}_{1T}(s)$. Notice that

$$T^{-1/2}\widehat{\beta}_{1T}(s) = C_T^{-1}(s) A_T(s)$$

where

$$C_T(s) = \frac{1}{h}\sum_{i=1}^{\lfloor Ts \rfloor} K_{\lfloor Ts \rfloor}\left(\frac{i - \lfloor Ts \rfloor}{h}\right)\left(\frac{i - \lfloor Ts \rfloor}{T} - \frac{1}{h}\sum_{j=1}^{\lfloor Ts \rfloor} K_{\lfloor Ts \rfloor}\left(\frac{j - \lfloor Ts \rfloor}{h}\right)\frac{j}{T}\right)$$

which equals

$$\frac{T}{h}\int_0^{\frac{\lfloor Ts \rfloor}{T}} K_{\lfloor Ts \rfloor}\left(\frac{\lfloor Tr \rfloor - \lfloor Ts \rfloor}{h}\right)\left[\frac{\lfloor Tr \rfloor - \lfloor Ts \rfloor}{T} - \frac{T}{h}\int_0^{\frac{\lfloor Ts \rfloor}{T}} K_{\lfloor Ts \rfloor}\left(\frac{\lfloor Tz \rfloor - \lfloor Ts \rfloor}{h}\right)\frac{\lfloor Tz \rfloor}{T}\, dz\right] dr,$$

and

$$A_T(s) = \frac{1}{h}\sum_{i=1}^{\lfloor Ts \rfloor} K_{\lfloor Ts \rfloor}\left(\frac{i - \lfloor Ts \rfloor}{h}\right)\left(\frac{1}{\sqrt{T}}Y_i - \frac{1}{h}\sum_{j=1}^{\lfloor Ts \rfloor} K_{\lfloor Ts \rfloor}\left(\frac{j - \lfloor Ts \rfloor}{h}\right)\frac{1}{\sqrt{T}}Y_j\right) W_T(r,s)$$

with

$$W_T(r,s) = \frac{\lfloor Tr \rfloor - \lfloor Ts \rfloor}{T} - \frac{T}{h} \int_0^{\lfloor Ts \rfloor / T} K_{\lfloor Ts \rfloor} \left(\frac{\lfloor Tz \rfloor - \lfloor Ts \rfloor}{h} \right) \frac{\lfloor Tz \rfloor}{T} \, dz.$$

Here, $K_{\lfloor Ts \rfloor}(z) = K(z) \left(h^{-1} \sum_{j=1}^{\lfloor Ts \rfloor} K([j - \lfloor Ts \rfloor]/h) \right)^{-1}$, $z \in \mathbb{R}$. Using Assumption (K), it is straightforward to check the following statements which we shall use in the sequel.

(i) K_s is Lipschitz continuous with Lipschitz constant $L = L_K(\int_0^\kappa \xi K(\xi(r-s))\, dr$, where L_K denotes a Lipschitz constant for K.

(ii) $\sup_{s \in [\kappa,1]} |K_{\lfloor Ts \rfloor}(z_{T,s}) - K_s(z_s)| = O(\sup_{s \in [\kappa,1]} |z_{T,s} - z_s|)$ for functions $z_s, z_{T,s}$, $T \in \mathbb{N}$.

(iv) $\sup_{s \in [\kappa,1]} |K_{\lfloor Ts \rfloor}(z_1) - K_{\lfloor Ts \rfloor}(z_2)| \leq L|z_1 - z_2| + O(T^{-1})$ for large T.

Using (i) we observe that

$$\left| K_{\lfloor Ts \rfloor} \left(\frac{\lfloor Tz \rfloor - \lfloor Ts \rfloor}{h} \right) - K_s(\xi(z-s)) \right| \leq L \left| \frac{\lfloor Tz \rfloor - \lfloor Ts \rfloor}{h} - \xi(z-s) \right| = O(1/T),$$

which implies $|K_{\lfloor Ts \rfloor}(\frac{\lfloor Tz \rfloor - \lfloor Ts \rfloor}{h})\frac{\lfloor Tz \rfloor}{T} - K_s(\xi(z-s)\xi)| = O(1/T)$. We can conclude that

$$\sup_{s \in [0,1]} \left| \int_0^s K_{\lfloor Ts \rfloor} \left(\frac{\lfloor Tz \rfloor - \lfloor Ts \rfloor}{h} \right) \frac{\lfloor Tz \rfloor}{T} \, dz - \xi \int_0^s K_s(\xi(z-s)) z \, dz \right| = o(1).$$

Iterating these arguments we see that

$$\lim_{T \to \infty} \sup_{s \in [0,1]} |C_T(s) - C(s)| = 0,$$

where C is defined in (4.3). By virtue of Lemma 2.1 and the Skorohod representation theorem in general metric spaces (Shorack and Wellner [1986, Th. 4, p.47]), there exists a probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P})$ and stochastic processes $\{\widetilde{\mathcal{Y}}(s)\}$ and $\{\widetilde{Y}_{T,\lfloor T\cdot \rfloor}\}$ defined on a new probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{P})$ such that $\{\widetilde{\mathcal{Y}}\}$ is equivalent to $\{\mathcal{Y}\}$, $\{\widetilde{Y}_{T,\lfloor T\cdot \rfloor}\}$ is equivalent to $\{Y_{T,\lfloor T\cdot \rfloor}\}$, and

$$\| T^{-1/2} \widetilde{Y}_{T,\lfloor T\cdot \rfloor} - \mathcal{Y} \|_\infty \to 0$$

w.p. 1, as $T \to \infty$. Here $\mathcal{Y}$ is as defined in Theorem 4.1. However, to simplify our exposition we will denote these equivalent processes again by $Y_{T,\lfloor T\cdot \rfloor}$ and $\mathcal{Y}$. Also recall that the second half of Skorohod's theorem asserts that convergence in probability in the Skorohod metric of the equivalent processes yields weak convergence of the original processes.

Next notice that $A_T(s)$ can be represented by a sequence of statistical functionals $\psi_T : D[0,1] \to D[0,1]$, $T \in \mathbb{N}$. Indeed, we have

$$A_T(s) = \psi_T(T^{-1/2} Y_{\lfloor T\cdot \rfloor})(s), \qquad T \in \mathbb{N},$$

where the function $\psi_T(f)$, $f \in D[0,1]$, is defined by

$$\psi_T(f)(s) = \frac{T}{h} \int_0^{\lfloor Ts \rfloor / T} K_{\lfloor Ts \rfloor} \left(\frac{\lfloor Tr \rfloor - \lfloor Ts \rfloor}{h} \right) \left(f(r) - \frac{T}{h} \phi_T(f)(s) \right) W_T(r,s)\, dr,$$

for $s \in [0,1]$. The functionals $\phi_T : D[0,1] \to D[0,1]$ are given by

$$\phi_T(f)(s) = \int_0^{\lfloor Ts \rfloor / T} K_{\lfloor Ts \rfloor} \left(\frac{\lfloor Tr \rfloor - \lfloor Ts \rfloor}{h} \right) f(z)\, dz, \qquad f \in D[0,1],\ s \in [0,1].$$

Let us first consider the latter functional. We shall verify that its limit is the functional $\phi(f)$, $f \in D[0,1]$, defined by

$$\phi(f)(s) = \int_0^s K_s(\xi(r-s)) f(z)\, dz, \qquad s \in [0,1].$$

Clearly, $|\phi_T(T^{-1/2} Y_{\lfloor T \cdot \rfloor}) - \phi(\mathcal{Y})|$ can be bounded by

$$|\phi_T(T^{-1/2} Y_{\lfloor T \cdot \rfloor}) - \phi_T(\mathcal{Y})| + |\phi_T(\mathcal{Y}) - \phi(\mathcal{Y})|.$$

The first term can be estimated by $\|K\|_\infty \|T^{-1/2} Y_{\lfloor T \cdot \rfloor} - \mathcal{Y}\|_\infty$, and the second one is not larger than

$$2L \sup_{z \in [0,1]} |\lfloor Tz \rfloor / h - \xi z| \sup_{s \in [0,1]} |\mathcal{Y}(s)| = O_P(1/T).$$

It follows that

$$\|\phi_T(T^{-1/2} Y_{\lfloor T \cdot \rfloor}) - \phi(\eta B)\|_\infty \xrightarrow{P} 0,$$

as $T \to \infty$, and consequently

$$\phi_T(T^{-1/2} Y_{\lfloor T \cdot \rfloor}) \Rightarrow \phi(\eta B),$$

as $T \to \infty$. Now one can easily conclude that

$$\rho_T(T^{-1/2} Y_{\lfloor T \cdot \rfloor}) = T^{-1/2} Y_{\lfloor T \cdot \rfloor} - \frac{T}{h} \phi_T(T^{-1/2} Y_{\lfloor T \cdot \rfloor})$$

converges w.r.t. the supnorm in probability, and hence also weakly, to the process

$$\rho(\mathcal{Y})(\cdot) = \mathcal{Y}(\cdot) - \xi \phi(\mathcal{Y}(\cdot)),$$

as $T \to \infty$. Noting that

$$\psi_T(f) = \frac{T}{h} \int_0^s \phi_T(f(r)) W_T(r,s)\, dr, \qquad f \in D[0,1],$$

has the same form as ϕ_T, we obtain

$$\|\psi_T(T^{-1/2}Y_{\lfloor T\cdot\rfloor}) - \psi(\mathcal{Y})\|_\infty \to 0,$$

as $T \to \infty$, in probability. Since the limit of $C_T^{-1}(s)$ is a deterministic function, the functional version of Slutzky's theorem yields

$$\widehat{\beta}_{1T}(s) \Rightarrow C^{-1}(s)\mathcal{A}(s)$$

where for $s \in [0,1]$

$$\mathcal{A}(s) = \xi \int_0^s K_s(\xi(r-s))\left(\mathcal{Y}(r) - \xi \int_0^s K_s(\xi(z-s))\mathcal{Y}(z)\,dz\right)\left(r - s - \xi \int_0^s K_s(\xi(z-s))z\,dz\right)dr.$$

The result for $\widehat{\beta}_0(s)$, $s \in [0,1]$, follows now easily from the representation

$$T^{-1/2}\widehat{\beta}_0(s) =$$
$$\frac{1}{h}\sum_{i=1}^{\lfloor Ts\rfloor} K_{\lfloor Ts\rfloor}\left(\frac{i-\lfloor Ts\rfloor}{h}\right)\frac{1}{\sqrt{T}}Y_i - \frac{1}{\sqrt{T}}\widehat{\beta}_{1T}(s)\left(\frac{1}{h}\sum_{i=1}^{\lfloor Ts\rfloor} K_{\lfloor Ts\rfloor}\left(\frac{i-\lfloor Ts\rfloor}{h}\right)\frac{i}{T}\right). \tag{6.1}$$

Denote the first term by $Z_T(s)$ and note that

$$Z_T(s) = \phi_T(Y_{\lfloor T\cdot\rfloor})(s) \Rightarrow \mathcal{Z}(s),$$

as $T \to \infty$. Further, for $u \in \{0,1\}$ the function

$$D_T^{(u)}(s) = \frac{1}{h}\sum_{i=1}^{\lfloor Ts\rfloor} K_{\lfloor Ts\rfloor}\left(\frac{i-\lfloor Ts\rfloor}{h}\right)\left(\frac{i}{T}\right)^u = \frac{T}{h}\int_0^s K_{\lfloor Ts\rfloor}\left(\frac{\lfloor Tr\rfloor - \lfloor Ts\rfloor}{h}\right)\left(\frac{\lfloor Tr\rfloor}{T}\right)^u dr$$

converges point-wise to the function

$$D^{(u)}(s) = \xi \int_0^s K_s(\xi(r-s))r^u\,dr,$$

as $T \to \infty$, and, by continuity of D, we even have uniform convergence in $s \in [0,1]$. Thus, since $D^{(1)} = D$,

$$(D_T^{(1)}, Z_T, T^{-1/2}\widehat{\beta}_{1T}) \Rightarrow (D, \mathcal{Z}, C^{-1}\mathcal{A})$$

in the product space $(D[0,1])^3$. Notice that the limit processes at the right-hand side are continuous a.s. Now the result follows by an application of the continuous mapping theorem using the mapping

$$(x,y,z) \mapsto y - xz, \qquad x,y,z \in D[0,1].$$

For the continuity properties of this mapping confer Whitt [1980]. The last facts also establish the joint weak convergence of $\widehat{\beta}_{0T}$ and $\widehat{\beta}_{1T}$. $\square$

Proof of Corollary 4.2. The result follows by noting that the processes are a.s. continuous, since C and D are continuous functions on $[\kappa,1]$, and the process $\mathcal{A}$ is a.s. continuous by virtue of the assumptions on K and μ. $\square$

ACKNOWLEDGEMENTS

The author thanks two anonymous referees for constructive remarks which improved the presentation of the results.

REFERENCES

Antoch, J., Gregoire, G. and Hušková, M. [2007], 'Tests for continuity of regression functions.', *J. Stat. Plann. Inference* **137**(3), 753–777.

Antoch, J., Hušková, M. and Jaruvsková, D. [2002], 'Off-line statistical process control.', Lauro, Carlo (ed.) et al., Multivariate total quality control. Foundation and recent advances. Heidelberg: Physica-Verlag. Contributions to Statistics, 1-86 (2002).

Aue, A., Horvath, L., Kuehn, M. and Steinebach, J. [2009], 'On the reaction time of moving sum detectors.', *preprint* .

Billingsley, P. [1999], *Convergence of probability measures. 2nd ed.*, Wiley Series in Probability and Statistics. Chichester: Wiley, 277 p.

Birnbaum, Z. and Saunders, S. [1969], 'A new family of life distributions.', *J. Appl. Probab.* **6**, 319–327.

Breitung, J. [2002], 'Nonparametric tests for unit roots and cointegration', *Journal of Econometrics* **108**, 343–363.

Cleveland, W. S. [1979], 'Robust locally weighted regression and smoothing scatterplots.', *J. Am. Stat. Assoc.* **74**, 829–836.

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**, 427–431.

Doksum, K. and Hoyland, A. [1992], 'Models for variable - stress accelerated life testing experiments based on Wiener processes and the inverse Gaussian distribution.', *Theory Probab. Appl.* **37**(1), 137–139.

Durrett, R. [2005], *Probability: theory and examples*, 3rd edn, Duxbury Press, Belmont, CA.

Edison, H. [1987], 'Purchasing power parity in the long run: A test of the dollar/pound exchange rate 1890-1978', *Journal of Money, Credit, and Banking* **19**, 376–387.

Fan, J. and Gijbels, I. [1996], *Local polynomial modelling and its applications.*, Monographs on Statistics and Applied Probability. 66. London: Chapman & Hall xvi, 341 p.

Frankel, J. [1986], 'International capital mobility and crowding out in the U.S. economy: Imperfect integration of financial or goods markets?', R. Hafer (ed.), Lexington Books.

Grégoire, G. and Hamrouni, Z. [2002], 'Change point estimation by local linear smoothing', *J. Multivar. Anal.* **83**(1), 56–83.

Gut, A. and Steinebach, J. [2002], 'Truncated sequential change-point detection based on renewal counting processes', *Scand. J. Statist.* **29**(4), 693–719.

Horváth, L. and Kokoszka, P. [2002], 'Change-point detection with nonparametric regression.', *Statistics* **36**(1), 9–31.

Hušková, M., Kirch, C., Prášková, Z. and Steinebach, J. [2008], 'On the detection of changes in autoregressive time series. II. Resampling procedures', *J. Statist. Plann. Inference* **138**(6), 1697–1721.

Hušková, M., Prášková, Z. and Steinebach, J. [2007], 'On the detection of changes in autoregressive time series. I. Asymptotics', *J. Statist. Plann. Inference* **137**(4), 1243–1259.

Kwiatkowski, D., Phillips, P. C., Schmidt, P. and Shin, Y. [1992], 'Testing the null hypothesis of stationarity against the alternative of a unit root. (How sure are we that economic time series have a unit root?).', *J. Econom.* **54**(1-3), 159–178.

Masry, E. and Fan, J. [1997], 'Local polynomial estimation of regression functions for mixing processes.', *Scand. J. Stat.* **24**(2), 165–179.

Müller, H.-G. and Song, K.-S. [1997], 'Two-stage change-point estimators in smooth regression models.', *Stat. Probab. Lett.* **34**(4), 323–335.

Müller, H.-G. and Stadtmüller, U. [1999], 'Discontinuous versus smooth regression.', *Ann. Stat.* **27**(1), 299–337.

Schmid, W. and Steland, A. [2000], 'Sequential control of non-stationary processes by nonparametric kernel control charts.', *Allg. Stat. Arch.* **84**(3), 315–336.

Shorack, G. R. and Wellner, J. A. [1986], *Empirical processes with applications to statistics*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York.

Steland, A. [2005], 'Random walks with drift – a sequential approach.', *J. Time Ser. Anal.* **26**(6), 917–942.

Steland, A. [2007*a*], 'Monitoring procedures to detect unit roots and stationarity', *Econometric Theory* **23**(6), 1108–1135.

Steland, A. [2007*b*], 'Weighted Dickey-Fuller processes for detecting stationarity.', *J. Stat. Plann. Inference* **137**(12), 4011–4030.

Stone, C. J. [1977], 'Consistent nonparametric regression. Discussion.', *Ann. Stat.* **5**, 595–645.

Taguchi, G. [1985], 'Quality engineering in japan', *Communications in Statistics - Theory and Methods* **14**, 2785–2801.

Taguchi, G., Elsayed, E. and Hsian, T. [1989], *Quality Engineering in Production Systems*, McGraw-Hill, New York.

Tsybakov, A. [1986], 'Robust reconstruction of functions by the local-approximation method.', *Probl. Inf. Transm.* **22**, 133–146.

Whitt, W. [1980], 'Some useful functions for functional limit theorems', *Math. Oper. Res.* **5**(1), 67–85.

Whitt, W. [2002], *Stochastic-process limits*, Springer Series in Operations Research, Springer-Verlag, New York. An introduction to stochastic-process limits and their application to queues.

Wu, J. and Chu, C. [1993], 'Kernel-type estimators of jump points and values of a regression function.', *Ann. Stat.* **21**(3), 1545–1566.