On Local Linear Surveillance of Trend Stability and Asymptotic Results for the Stopped Detectors

Ansgar Steland[1]

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

Abstract. The problem of detecting a change in a deterministic trend disturbed by stationary noise is a classic problem of sequential analysis, which generalizes the related detection problem considering i.i.d. error terms. In some areas, methods which can cope with error terms that behave as a random walk with possibly dependent increments are also of importance. For instance, many econometric time series are random walks, and in engineering and computer science they appear as models for damage processes in reliablity as well as workload processes of computing devices as network routers. Focusing on random walk errors, this note studies detectors based on local linear estimates under non-standard conditions, discusses recent results and contributes accompanying asymptotics for the stopped processes and stopping times, respectively. It turns out that the limit processes can be represented as functionals of Brownian motion and certain Itô integrals, respectively. We also discuss simulation experiments investigating the distribution of the stopped processes and apply the procedure to simulated series.
Keywords. Brownian motion, change-point, control chart, FCLT, invariance principle, nonparametric estimation, time series.

1 Introduction

Suppose we observe a sequence of random variables $\{Y_n\} = \{Y_n : n \in \mathbb{N}_0\}$ where Y_n is observed at the nth time instant and observations arrive sequentially. Often the aim is to detect a change in the mean indicating that the process is no longer in a state of statistical control. This can be achieved by surveillance (monitoring) procedures. Usually, the detection performance is improved when basing a detector on a statistic which performs some preliminary averaging procedure in order to reduce the noise component instead of using the observations themselves which corresponds to the Shewhart chart. Methods based on moving sums (MOSUM charts), cumulated sums (CUSUM charts), exponentially weighted averages (EWMA charts), and kernel averages are well studied from a nonparametric viewpont. For recent results and references to the literature dealing with dependent but stationary error terms we refer to Aue et al. (2008a, 2008b) and Gut and Steinebach (2004). To detect stationarity of the error terms when the null hypothesis (in-control process) states that the errors behave as a random walk, surveillance procedures have been proposed by Steland (2008).

When aiming at nonparametric estimation of the process mean, local estimation techniques have proven to be superior, particularly local linear estimation. Consistency can be ensured provided that the time instants where observations are available get dense, asymptotically. We refer to Fan and Gijbels (1996) for the general methodology. Masry and Fan (2002) provide asymptotic theory for mixing processes. Grégoire and Hamrouni (2002) used this technique to define a random objective function for a posteriori (off line) estimation of a jump point in a smooth curve. Having these results in mind, using sequential versions of these estimators seems to be promising to develop monitoring (surveillance) procedures where the aim is to detect quickly a rapid change of the process mean, particularly, since often it is not possible to specify the mean after the change. Thus, basing a detector on a control statistic which is known to provide good estimates in the non-sequential framework seems promising. Monte Carlo experiments have demonstrated that such procedures perform very well in many cases. However, the sequential asymptotic distribution theory has been an open problem.

The rest of the paper is organized as follows. In Section 2 we introduce the (sequential) local linear estimation principle and explain the proposed surveillance procedure. Relevant asymptotic results are reviewed in Section 3. Section 4 provides the main result of the present article whose proof is postponed to an appendix. Finally, we discuss results from a simulation study and illustrate the application of the method.

2 Sequential Local Linear Estimation

Let us briefly review the basic idea of local linear estimation. Given a sequence $Y_1, Y_2, \dots$ of real-valued random variables denote the corresponding (marginal) means by $m(t) = E(Y_t)$. Suppose that locally at the current time instant $t_n = n \in \mathbb{N}$ the approximation

$$m(s) = \beta_{0n}(t_n) + \beta_{1n}(t_n)(s - t_n) + o(1)$$

with unknown local intercept $\beta_{0n} = \beta_{0n}(t_n)$ and slope $\beta_{1n} = \beta_{1n}(t_n)$ holds true. Notice that these local parameters can be very informative in an analysis, since they measure locally 'level' and 'derivative' of the trend. Thus, statistical intuition suggests to use sequential estimates of them to construct detection procedures for process surveillance.

To estimate the local parameters we fit a straight line to the data by weighted least squares, i.e., given the data $Y_1, \dots, Y_n$ at the current time instant t_n, we minimize the objective

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

with respect to $(\beta_0, \beta_1) \in \mathbb{R}^2$. $\{w_{ni}\}$ are nonnegative weights defined via a kernel function K and a bandwidth $h > 0$, namely

$$w_{ni} = K([i-n]/h) / \sum_{j=1}^{n} K([j-n]/h).$$

Denote the minimizers by $\widehat{\beta}_{0n}$ and $\widehat{\beta}_{1n}$. Notice that these estimates are $\mathcal{F}_n = \sigma(Y_s : s \leq n)$-measureable by construction, thus providing a reasonable basis for the construction of a stopping time, both from an intuitive and a mathematical viewpoint. Specifically, one may consider the stopping rules

$$L_T^{(i)} = \inf\{k \leq n \leq T : T^{-1/2}\widehat{\beta}_{in} > c_i\}, \qquad i = 0, 1, \tag{1}$$

for control limits (critical value) c_0, c_1. k denotes the first time point where surveillance starts. To ensure that any decision depends on a minimal number of observations, it is assumed that

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

for some constant $\kappa \in (0, 1)$. Notice that monitoring stops when the time horizon T is reached. Such procedures are also called *closed end* stopping times. We propose to select the control limits to ensure that the type I error does not exceed a pre-specified nominal value $\alpha \in (0, 1)$ as $T \to \infty$.

To the $\mathcal{F}_T$-adapted sequences $\{\widehat{\beta}_{in} : 1 \leq n \leq T\}$, $i = 1, 2$, we may associate cádlàg processes representing these estimates as time proceeds. Define

$$\widehat{\beta}_{iT}(s) = T^{-\gamma}\widehat{\beta}_{i,\lfloor Ts \rfloor}, \quad s \in [0, 1], \quad i = 0, 1, \tag{2}$$

where $\lfloor x \rfloor$ denotes the integer part of a real number x. The rate constant γ depends on whether the error terms $\epsilon_t = Y_t - m(t)$ behave as a random walk or as a stationary process.

3 A Review of the Asymptotic Theory for Random Walk Errors

In Steland (2009a) a functional central limit theorem (FCTL, invariance principle) has been established under a class of local change-point models for the mean, disturbed by error terms which behave asymptotically as a random walk with possibly dependent increments. The assumptions on the increments are weak and allow for many sets of assumptions encountered in time series models and applied work. Particularly, it is not required that the error terms belong to a parametric class or form a linear process. Further, the class of local change-point models covers various settings of practical interest. The asymptotic results clarify to a large extent the stochastic behavior of any surveillance procedure, which is a functional of the process (2).

We shall now review some of the results obtained in Steland (2009a). Suppose that $Y_{T1}, \dots, Y_{TT}$ arrive sequentially. Assume

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

where the array of constants $m_{Tt} = E(Y_{Tt})$ satisfies

$$m_{Tt} = T^{1/2} \int_0^{t/T} \mu(r)\, dr + o(T^{1/2}), \qquad 0 \le t \le T,\ T \ge 1, \tag{3}$$

for some function $\mu : [0,1] \to \mathbb{R}$ with $\int_0^1 |\mu(r)|\, dr < \infty$. The model function μ is used to define the change-point model. Further, it is assumed that the errors fulfill a FCLT, i.e., $\{\epsilon_{Tt} : 1 \le t \le T, T \ge 1\}$ is an array of zero mean random variables with

$$T^{-1/2} \epsilon_{\lfloor Ts \rfloor} \Rightarrow \eta B(s), \tag{4}$$

as $T \to \infty$, for some constant $\eta \in (0, \infty)$. Here B denotes a standard Brownian motion (Wiener process).

Assuming (3), (4) and some additional regularity conditions on the kernel K and the bandwidth h which are given in the main result (Theorem 1) of the present paper, it is shown in Steland (2009a) that

$$\left(T^{-1/2} \widehat{\beta}_{0T}(s), T^{-1/2} \widehat{\beta}_{1T}(s) \right) \Rightarrow \left(\mathcal{B}(s), C^{-1}(s) \mathcal{A}(s) \right), \tag{5}$$

as $T \to \infty$, where

$$\mathcal{B}(s) = (\mathcal{Z}(s) - C^{-1}(s)\mathcal{A}(s)D(s), C^{-1}(s)\mathcal{A}(s)), \qquad s \in [0,1],$$

with

$$D(s) = \xi \int_0^s K_s(\xi(r-s)) r\, dr, \tag{6}$$

$$C(s) = \xi \int_0^s K_s(\xi(r-s)) W_\xi(r,s)\, dr, \tag{7}$$

$$W_\xi(r,s) = \xi(r-s) - \xi \int_0^s K_s(\xi(z-s)) z\, dz, \tag{8}$$

$$\mathcal{A}(s) = \xi \int_0^s K_s(\xi(r-s)) \left[\mathcal{Y}(r) - \xi \int_0^s K_s(\xi(z-s)) \mathcal{Y}(z)\, dz \right] W_\xi(r,s)\, dr, \tag{9}$$

$$\mathcal{Z}(s) = \xi \int_0^s K_s(\xi(r-s)) \mathcal{Y}(r)\, dr, \tag{10}$$

$$\mathcal{Y}(s) = \int_0^s \mu(r)\, dr + \eta B(s), \tag{11}$$

$$K_s(z) = K(z) / \int_0^s \xi K(\xi(r-s))\, dr. \tag{12}$$

The weak convergence result (5) also yields a central limit theorem (CLT) for the stopping time. We have under the above conditions

$$L_T^{(i)}/T \xrightarrow{d} \mathcal{L}^{(i)} = \inf\{ s \in [\kappa, 1] : \mathcal{B}(s) > c_i \}, \quad i = 0, 1, \tag{13}$$

as $T \to \infty$.

The asymptotic distribution theory for the case when the error terms are stationary and satisfy certain additional (weak) regularity conditions will be discussed in detail in Steland (2009b). In this case the limit process can be represented via certain stochastic Itô process.

4 A Central Limit Theorem for the Stopped Control Statistics

Recall the definition of the stopping time $L_T^{(0)}$.

$$L_T^{(0)} = \inf\{k \le n \le T : T^{-1/2}\widehat{\beta}_{0T}(n/T) > c_0\}.$$

Having in mind that we stop in any case at time T, put $L_T^{(0)} = T$ if $\{\cdots\} = \emptyset$. Analogously, we put $\mathcal{L}^{(0)} = 1$, if $\mathcal{B}(s)$ does not reach the level c_0. When the surveillance procedure $L_T^{(0)}$ gives a signal at time $n(< T)$, we know that $T^{-1/2}\widehat{\beta}_{0T}(n/T) > c_0$. There may be some overshot, and the question arises what can be said about the overshot asymptotically. When a signal is given, one may also look at the value of the estimated slope. Thus, this section is devoted to the asymptotics of the stopped bivariate sequential local linear processes.

The stopped control statistics are given by

$$B_{0T}^{st} = T^{-1/2}\widehat{\beta}_{0T}(L_T^{(0)}/T), \quad B_{1T}^{st} = T^{-1/2}\widehat{\beta}_{1T}(L_T^{(0)}/T).$$

We are now in a position to establish the following CLT.

Theorem 1. *Assume (3) and (4). Suppose that the bandwidth h is chosen as a function of T such that*

$$|T/h - \xi| = O(1/T)$$

for some known constant $\xi > 0$, and let K be a bounded and Lipschitz continuous kernel which is positive on $(-\xi, \xi)$. Then, given the event that a signal is observed,

$$B_{0T}^{st} \xrightarrow{P} c_L \quad \text{and} \quad B_{1T}^{st} \xrightarrow{d} C^{-1}(\mathcal{L}^{(0)})\mathcal{A}(\mathcal{L}^{(0)}),$$

as $T \to \infty$. The random variable $\mathcal{L}^{(0)}$ is defined in (13).

Theorem 1 can be used to construct asymptotic level α tests to draw inference on the stopped control statistics. It is straightforward to establish similar results for the control statistics stopped at the random stopping time $L_T^{(1)}$. For brevity of presentation we omit these results.

Remark 1. Notice that $\{T^{-1/2}\widehat{\beta}_{0T}(s) : s \in [\kappa, 1]\}$ and $L_T^{(0)}$ are dependent as well as $\{\mathcal{B}(s) : s \in [\kappa, 1]\}$ and $\mathcal{L}^{(0)}$. Thus, the result does not automatically follow from (5) and (13). A more subtle argument is required which is given in the appendix.

5 Simulations and Application

Figure 1 illustrates the application of the proposed method to two simulated time series of length $T = 250$. Both series as well as the series for the simulation study were generated by the change-point model

$$Y_n = \sum_{i=1}^{n}(\Delta \mathbf{1}_{\{i \ge 125\}} + \epsilon_i), \qquad n = 1, \ldots, 250,$$

where ϵ_i are i.i.d. standard normal random variables and $\Delta = 0.4$. That is, there is change-point at $q = 125$ and after the change the random walk is affected by deterministic trend with slope Δ. Before the change the series is a pure random walk with standard normal increments. Notice that 125 pre- and post-change observations define a difficult small sample situation.The random walks errors produce spurious trends which are hard to distinguish from deterministic trends. We applied the sequential local linear estimator with a Gaussian kernel and two bandwidth choices, namely $h = 250/10$ (dashed) and $h = 250/20$ (solid), i.e., ξ equals 10 and 20, respectively. Both choices correspond to high data-fidelity as often done in practice.The asymptotic control limits were calculated to ensure a nominal type I error rate of 5% and are added to the plot. They were obtained by simulating trajectories of the limit process yielding simulated replications of the stopping time $L_T^{(0)}$. According to the simulation study in Steland (2009a), the accuracy of this approximation is quite good for the parameter choices used in this example.

Figure 1 illustrates the procedure using two simulated series. For the series of the left panel the signal is given at time 150 if $\xi = 10$ and at time 145, if $\xi = 20$. For the other series the change is detected later, namely at time 241, if $\xi = 10$, and at time 245, when $\xi = 20$. Figure 2 provides the simulated distribution of the stopped statistic B_{1T}^{st} for $\xi = 10$ and also shows density estimates when a signal is given or not. By the convention of this section, the latter case corresponds to $T^{-1/2}\widehat{\beta}_{1T}(1)$. The non-normality is clearly visible. The probability of a signal is estimated by 0.583.

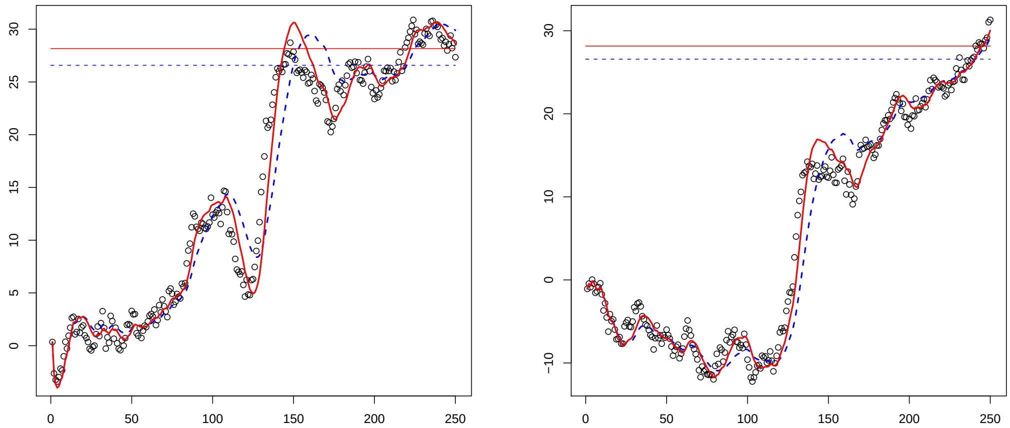

Fig. 1. Random walks with change in the drift. Depicted are the sequential local linear estimates for $h = 250/10$ (dashed) and $h = 250/20$ (solid), respectively.

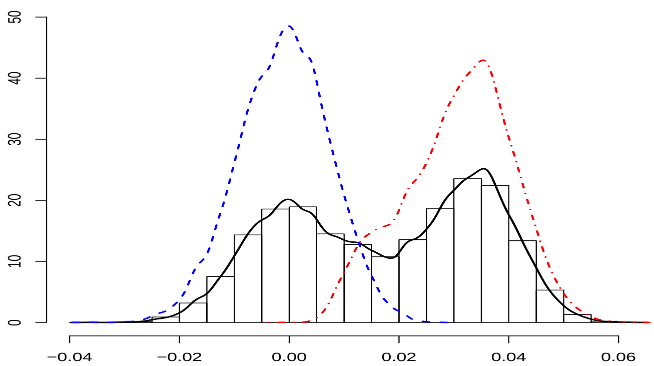

Fig. 2. Simulated distribution (histogram and density esimate) of stopped statistics for $\xi = 10$ (solid). Density esimates when a signal is given (dashed-dotted, red) or not (dashed, blue).

A Proof of Theorem 1

Proof. Denote by $(\Omega, \mathcal{F}, P)$ the underlying probability space. By virtue of the Skorohod/Dudley/Wichura representation theorem in general metric spaces, we may and will assume that

$$\sup_{s \in [0,1]} |T^{-1/2} \widehat{\beta}_{0T}(s) - [\mathcal{Z}(s) - C^{-1}\mathcal{A}(s)D(s)]| \stackrel{a.s.}{\to} 0,$$

$$\sup_{s \in [0,1]} |T^{-1/2} \widehat{\beta}_{1T}(s) - C^{-1}\mathcal{A}(s)| \stackrel{a.s.}{\to} 0,$$

as well as $L_T^{(0)} \stackrel{a.s.}{\to} \mathcal{L}^{(0)}$, if $T \to \infty$. Notice that $s \mapsto \mathcal{B}(s)$ is continuous a.s. Thus, there exists a measureable set $A \subset \Omega$ with $P(A) = 1$ such that

$$\mathcal{B} \in C[0,1] \quad \text{and} \quad L_T^{(0)}/T \to \mathcal{L}^{(0)},\ T \to \infty, \quad \text{on } A,$$

which implies

$$\mathcal{B}(L_T^{(0)}/T) \to \mathcal{B}(\mathcal{L}^{(0)}),\ T \to \infty, \qquad \text{on } A.$$

Thus, we also have

$$\mathcal{B}(L_T^{(0)}/T) \stackrel{a.s.}{\to} \mathcal{B}(\mathcal{L}^{(0)}), \qquad T \to \infty.$$

Now we may argue as follows.

$$\begin{aligned} |B_{0T}^{st} - \mathcal{B} \circ \mathcal{L}^{(0)}| &\leq |T^{-1/2}\widehat{\beta}_{0T}(L_T^{(0)}/T) - \mathcal{B}(L_T^{(0)}/T)| + |\mathcal{B}(L_T^{(0)}/T) - \mathcal{B}(\mathcal{L}^{(0)})| \\ &\leq \sup_{s \in [\kappa,1]} |T^{-1/2}\widehat{\beta}_{0T}(s) - \mathcal{B}(s)| + o_P(1). \end{aligned}$$

Thus, the second half of the Skorohod/Dudley/Wichura representation theorem yields

$$B_{0T}^{st} \stackrel{d}{\to} \mathcal{B} \circ \mathcal{L}^{(0)},$$

as $T \to \infty$. Thus, we obtain for any measureable set A,

$$\begin{aligned} P(B_{0T}^{st} \in A) &= P(T^{-1/2}\widehat{\beta}_{0T}(L_T^{(0)}/T) \in A) \\ &\to P((\mathcal{Z} - C^{-1}\mathcal{A}D) \circ \mathcal{L}^{(0)} \in A), \end{aligned}$$

as $T \to \infty$. Clearly, the event that a signal is observed is

$$S_T = \{L_T^{(0)} < T\} \in \sigma(L_T^{(0)}).$$

Therefore,

$$\begin{aligned} P(\{B_{0T}^{st} \in A\} \cap S_T) &= P(T^{-1/2}\widehat{\beta}_{0T}(L_T^{(0)}/T) \in A, L_T^{(0)} < T) \\ &\to P((\mathcal{Z} - C^{-1}\mathcal{A}D) \circ \mathcal{L}^{(0)} \in A, \mathcal{L}^{(0)} < 1) \\ &= P(c_0 \in A, \mathcal{L}^{(0)} < 1) = \mathbf{1}(c_0 \in A)P(\mathcal{L}^{(0)} < 1), \end{aligned}$$

as $T \to \infty$. Combining this fact with $P(S_T) \to P(\mathcal{L}^{(0)} < 1), T \to \infty$, we arrive at

$$P(B_{0T}^{st} \in A | S_T) \to \mathbf{1}(c_0 \in A),$$

as $T \to \infty$, for any measureable set A, which is equivalent to $B_{0T}^{st} \stackrel{P}{\to} c_0$, if $T \to \infty$. Although the weak limit of B_{1T}^{st} is not degenerate, we can argue along the lines of the above proof. We indicate the important steps. First note that the process $\{C^{-1}(s)\mathcal{A}(s) : s \in [\kappa, 1]\}$ is continuous a.s. Arguing as above we can conclude that

$$C^{-1}(L_T^{(0)}/T)\mathcal{A}(L_T^{(0)}/T) \stackrel{a.s.}{\to} C^{-1}(\mathcal{L}^{(0)})\mathcal{A}(\mathcal{L}^{(0)}),$$

as $T \to \infty$. Now use the estimate

$$\begin{aligned}|B_{1T}^{st} - C^{-1}(\mathcal{L}^{(0)})\mathcal{A}(\mathcal{L}^{(0)})| &\le |T^{-1/2}\widehat{\beta}_{1T}(L_T^{(0)}/T) - C^{-1}(L_T^{(0)}/T)\mathcal{A}(L_T^{(0)}/T)| \\ &\quad + |C^{-1}(L_T^{(0)}/T)\mathcal{A}(L_T^{(0)}/T) - C^{-1}(\mathcal{L}^{(0)})\mathcal{A}(\mathcal{L}^{(0)})| \\ &\le \sup_{s\in[\kappa,1]} |T^{-1/2}\widehat{\beta}_{1T}(s) - C^{-1}(s)\mathcal{A}(s)| + o_P(1)\end{aligned}$$

which is $o_{a.s.}(1) + o_P(1)$. This verifies the second assertion, namely,

$$B_{1T}^{st} \xrightarrow{d} C^{-1}(\mathcal{L}^{(0)})\mathcal{A}(\mathcal{L}^{(0)}),$$

as $T \to \infty$.

References

Aue, A., Horváth, L., Hušková, M. and Kokoszka, P. (2008a). Testing for changes in polynomial regression. *Bernoulli*, **14**, 3, 637-660.

Aue, A., Horváth, L., Kokoszka, P. and Steinebach J. (2008b). Monitoring shifts in mean: Asymptotic normality of stopping times. *Test*, **17**, 515-530.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*, Chapman and Hall, London.

Grégoire, G. and Hamrouni, Z. (2002). Change point estimation by local linear smoothing, *Journal of Multivariate Analysis*, **83**, 1, 56-83.

Gut, A. and Steinebach, J. (2004). EWMA charts for detecting a change-point in the drift of a stochastic process. *Sequential Analysis*, **23**, 2, 195-237.

Masry, E. and Fan, J. (1997). Local polynomial estimation of regression functions for mixing processes, *Scandinavian Journal of Statistics*, **34**, 4, 165-179.

Steland, A. (2008). Sequentially updated residuals and detection of stationary errors in polynomial regression models. *Sequential Analysis*, **27**, 3, 304-329.

Steland, A. (2009a). A surveillance procedure for random walks based on local linear estimation, *in revision*.

Steland, A. (2009b). Stability analysis of time series by sequential local linear estimation, *preprint*.