On the Online Detection of Monotonic Trends in Time Series

R. FRIED[1]* and M. IMHOFF[2]

[1] Department of Statistics, University of Dortmund, 44221 Dortmund, Germany

[2] Surgical Department, Hospital of Dortmund, 44137 Dortmund, Germany

Abstract

The online detection of a slow monotonic trend in a time series with a time-varying mean is an important task in biomedical applications like intensive care monitoring, that is rendered difficult by autocorrelations. Statistical control charts designed for controlling industrial processes are not adequate as they typically rely on a fixed target value and since they do not consider the mean to be time-varying. Moreover, many detection rules for slow monotonic trends assume a trend to be linear. We develop an approach for online detection of slow, possibly nonlinear monotonic trends. For estimation of the model parameters we use a moving time window in order to incorporate time-varying autocorrelations. The performance of some variations of this approach and of the resulting detection rules is investigated in a simulation study, and an application to real data is given. We find that shrinkage estimation of a time-varying mean enhances the discriminatory power of the test statistic.

Key words: statistical process control; online monitoring; change-point detection; shrinkage estimation.

1 Introduction

In many biomedical and environmental applications of statistical data analysis subsequently measured observations need to be analyzed online. The fast and reliable detection of patterns of change in such time series data is important since these patterns may point at some change in the data generating mechanism. In intensive care for instance, early detection of a monotonic trend in a time series representing heart rate or blood pressure allows the

*Corresponding author: e-mail: fried@statistik.uni-dortmund.de

physician to take some therapeutical intervention before a critical threshold is exceeded. Positive autocorrelations render this task difficult since they cause monotonic sequences in the data, just like deterministic trends. Moreover, deterministic trends do not only influence the process mean, but they also strongly influence the ordinary sample autocorrelations, which are typically needed to standardize test statistics for trend detection. In retrospective analysis often a simple linear trend is fitted to the data and the autocorrelations are approximated by an autoregressive model for the noise (Cochrane and Orcutt, 1949, Bloomfield and Nychka, 1992). The reason is that autoregressive models constitute a quite flexible model class describing a wide variety of autocorrelation functions and that simple algorithms for model fitting exist.

For online trend detection, control charts based on exponential smoothing are frequently recommended (Trigg, 1964, Cembrowski *et al.*, 1975, Montgomery and Mastrangelo, 1991, Schack and Grieszbach, 1994). However, for any choice of the weighting parameter there are scenarios where the resulting EWMA chart shows poor performance. Monte Carlo comparisons of EWMA and CUSUM charts reveal that none of these procedures has overall optimal performance for the detection of slow trends in independent measurements (Chang and Fricker, 1999). Moreover, these charts are designed for industrial processes where we may assume stationarity in the steady state, i.e. unique model parameters for the whole process, as well as the existence of a fixed target value. In applications such as monitoring pollution in environmental sciences or controlling vital signs in intensive care stationarity can often not be assumed, and it is not possible to specify a target value in advance because of natural fluctuations within the data generating mechanism itself like seasonality or biorhythms (Högel, 2000, Schmid and Steland, 2000, Gather, Imhoff and Fried, 2002, Gather *et al.*, 2003).

We adapt Brillinger's (1989) approach for retrospective detection of a monotonic trend to the online-monitoring context by applying a test statistic to the data in a moving time window. Moving window techniques are useful to estimate model parameters which are time-varying and to construct adaptive control limits assuming stationarity to hold only locally (Dahlhaus, 1997). W.r.t. the length of the time window we must look for a compromise since a long time window results in a small variance for the expense of a large bias. We compare several strategies for automatic preliminary detrending of the data as we need to reduce the influence of trend patterns on the estimated time series model.

We proceed as follows. Section 2 describes the basic underlying model and the proposed procedure. In Section 3 the reliability of this procedure and of the parameter estimates is checked by simulations assuming an AR(1) model, which is frequently used in practice. In Section 4, the procedure is then applied to online-monitoring data observed in intensive care, before we finish with a discussion of the results.

2 The model

We assume that at each time point $t \in \mathbb{Z}$ the measurement Y_t of a deterministic signal μ_t is disturbed by additive autocorrelated random noise E_t,

$$Y_t \quad = \quad \mu_t + E_t\,, \quad t \in \mathbb{Z}.$$

For retrospective trend detection often a linear model such as $\mu_t = \beta_1 t + \beta_0$ is assumed for the mean (Cochrane and Orcutt, 1949, Bloomfield and Nychka, 1992, Sun and Pantula, 1999) and it is tested whether an estimate $\hat{\beta}_1$ of β_1 is significantly different from zero. However, this means to specify a fixed form of the mean. Trends which are not linear may not be detected this way. This problem becomes even more serious in online monitoring since a procedure needs to function automatically and reliably in a wide variety of situations.

ABELSON and TUKEY (1963) suggest to use a weighted sum $\sum_{t=1}^{N} c_t Y_t$ to test for a monotonic increase of μ_t during a time interval $t = 1, \ldots, N$, i.e. $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_N$ with $\mu_t < \mu_{t+1}$ for at least one $t \in \{1, \ldots, N-1\}$. They restrict the weights $c_1, \ldots, c_N$ to fulfill $\overline{c} = \sum_{t=1}^{N} c_t/N = 0$, such that $\sum c_t \mu_t$ also equals zero in case of a constant mean $\mu_1 = \mu_2 = \ldots = \mu_N$. Then the weights are determined to solve

$$\max_c \min_\mu \frac{|\sum(c_t - \overline{c})(\mu_t - \overline{\mu})|^2}{\sum(c_t - \overline{c})^2 \sum(\mu_t - \overline{\mu})^2}$$

where $\overline{\mu} = \sum \mu_t/N$, i.e. the weights are chosen to have worst case discriminatory power for an extremely unfavorable trend which is as high as possible. This results in

$$c_t = \left[(t-1)\left(1 - \frac{t-1}{N}\right)\right]^{1/2} - \left[t\left(1 - \frac{t}{N}\right)\right]^{1/2}$$

and the corresponding worst case is a single step change. Thus, the hypothesis of a constant mean should be rejected in favor of a monotonic increasing (decreasing) mean if

$$T = \frac{\sum_{t=1}^{N} c_t Y_t}{\hat{\tau}}$$

takes large positive (negative) values, where $\hat{\tau}$ is an estimate of the standard deviation τ of the weighted sum. ABELSON and TUKEY (1963) estimate τ and calculate critical values for T under the assumption that the noise process consists of independent identically distributed variables. BRILLINGER (1989) generalizes this approach to stationary noise processes using a running mean to estimate the mean μ_t at time t and a nonparametric approach to estimate τ. The test statistic turns out to be asymptotically normally distributed. However, Monte Carlo experiments show that the resulting rule is too sensitive even in retrospective applications to long time series (WOODWARD, BOTTONE and GRAY, 1997).

In online monitoring of a locally stationary process with a possibly slowly time-varying mean and autocorrelations we may test whether a monotonic trend has occurred during a time window of the last n observations. Large sample asymptotics are not useful if the window width n is not very large. The choice of n will often be guided by the application. In our case, we use $n = 60$ observations corresponding to one hour of measurements. Brillinger's approach for standardization should not be applied here since the number of observations used for a running mean must not be very small to reduce the impact of the noise, but it must be small relatively to the window width n. Therefore we assume in the following that the noise can be approximated locally within each time window by an AR(p) process

$$E_t \quad = \quad \phi_{1,N} E_{t-1} + \ldots + \phi_{p,N} E_{t-p} + U_t \,, \quad t = 1, \ldots, n,$$

where we denote the observed $N \geq n$ values by $y_{n-N+1}, \ldots, y_0, y_1, \ldots, y_n$ for notational simplicity, i.e. $y_1, \ldots, y_n$ correspond to the current time window. Here, $\phi_{1,N}, \ldots, \phi_{p,N}$ are unknown autoregressive coefficients and $\{U_t, t \in \mathbb{Z}\}$ denotes an unobservable white noise process with uncorrelated, identically distributed disturbances having mean zero and unknown variance σ_N^2. We note that all parameters may vary in time, i.e. depend on the current time window, and suppress indices representing the time window further on for simplicity.

Two types of control charts have been suggested for considering autocorrelations. One possibility is to fit a time series model like an AR(p) model to the data and to apply a standard control chart for independent data to the residuals. Alternatively one can use the untransformed observations and adjust the control limits for the autocorrelations. We prefer the latter approach here since experience shows that it is often better to use the original observations than residuals from a possibly misspecified time series model (see Lu and Reynolds, 1999, and the references cited therein). Nevertheless, comparing the procedure worked out in the following to a residuals based analogue might be interesting.

In order to standardize the weighted sum $\sum_{t=1}^{n} c_t Y_t$ we need to estimate its variance

$$\tau^2 = Var\left(\sum_{t=1}^{n} c_t Y_t\right) = \sum_{t=1}^{n}\sum_{s=1}^{n} c_t c_s \gamma(t-s)\,. \tag{1}$$

Hence, we need reliable estimates of the autocovariances $\gamma(0), \gamma(1), \ldots, \gamma(n-1)$ at time lags

$0, \ldots, n-1$, or, equivalently, of the AR model parameters. A deterministic trend seriously affects the ordinary sample autocovariances as these decay to zero very slowly, irrespective of the true values. Similarly, standard estimates of the AR parameters are seriously mislead by a deterministic trend if we assume the mean to be constant. Thus we have to eliminate or at least reduce the impact of a trend in a simple way before we estimate the model parameters. In the remainder of this section we present some variations of a model based approach for detrending the data.

The basic idea is to approximate a deterministic trend within the current time window by a linear model $\mu_t = \mathbf{f}(t)'\boldsymbol{\beta}$ and to estimate the autocovariances from the residuals. As mentioned above, in retrospective applications often a simple linear trend $\mu_t = \beta_0 + \beta_1 t$ is assumed. In the following we suggest some modifications of this approach that may be particularly useful in an automatic online application.

Using a simple linear trend for detrending is not completely satisfactory as we want reliable estimates for all time windows, also for those where only some of the observations are influenced by a trend, e.g. the beginning of a trend period. Therefore we suggest fitting a higher dimensional trend function $\mathbf{f}(t) = (f_0(t), \ldots, f_k(t))'$ with $k+1 \geq 2$ to achieve more flexibility. The components $f_i(t)$ could be chosen as low order polynomials for instance, but then extrapolation of a trend is difficult because of the strong curvature of polynomials of order higher than one. Instead we suggest using ramp functions such as

$$g_s(t) = \begin{cases} 0, & t < s \\ t - s, & t \geq s \end{cases},$$

$s \in \{0, \ldots, n-1\}$, describing a linear trend starting at time point s. By constructing linear combinations $\mu_t = \mathbf{f}(t)'\boldsymbol{\beta}$ we get piecewise linear functions. For $n = 60$ we take $k = 3$ with $f_1 = g_0$, $f_2 = g_{20}$, $f_3 = g_{40}$, and f_0 constant, i.e. $f_0(t) = 1$, $t = 1, \ldots, n$. This allows a rough approximation of any trend. We can estimate $\boldsymbol{\beta}$ using

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{\boldsymbol{V}} &= \left(\mathbf{X}'\boldsymbol{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{V}^{-1}\mathbf{y}, \\
\text{where } \mathbf{X} &= \begin{pmatrix} f_0(1) & \dots & f_k(1) \\ f_0(2) & \dots & f_k(2) \\ \vdots & & \vdots \\ f_0(n) & \dots & f_k(n) \end{pmatrix}
\end{aligned}
$$

and $\mathbf{y} = (y_1, \dots, y_n)'$. For $\boldsymbol{V} = \boldsymbol{I}$, the $(n \times n)$-identity matrix, we get the ordinary least squares estimate $\tilde{\boldsymbol{\beta}}_k = \hat{\boldsymbol{\beta}}_{\boldsymbol{I}}$, while for $\boldsymbol{V} = \boldsymbol{\Sigma} = (\gamma(i-j))_{1 \le i,j \le n}$, the $(n \times n)$ covariance matrix of $E_1, \dots, E_n$, we get the generalized least squares estimate. While the latter is not applicable as $\boldsymbol{\Sigma}$ is unknown, the former assumes independence of the observations when estimating $\boldsymbol{\beta}$. For any reasonable choice of $\boldsymbol{V}$ we can reduce the impact of a trend by subtracting $\mathbf{f}(t)'\hat{\boldsymbol{\beta}}_{\boldsymbol{V}}$ from y_t obtaining estimates $\hat{\gamma}_{\boldsymbol{V}}(h)$, $h = 0, \dots, n-1$, of the autocovariances from the residuals $\hat{Z}_t = Y_t - \mathbf{f}(t)'\hat{\boldsymbol{\beta}}_{\boldsymbol{V}}$, $t = 1, \dots, n$.

In the following we restrict attention to AR(1) models. These models are frequently used in practice for adjusting a trend estimate when correlation is present, particularly in case of short time series. Some simplifications are possible then. Since there is only one autoregressive parameter we simply call it ϕ. Then we have $\gamma(0) = \sigma^2/(1-\phi^2)$ and $\gamma(h) = \phi^h\gamma(0), h \ge 1$. We get estimates of $\gamma(h), h = 0, \dots, n-1$, by inserting estimates of σ^2 and ϕ into these equations. We can estimate ϕ and σ consistently (for $n \to \infty$) using

$$
\tilde{\phi}_k = \frac{\sum_{t=2}^{n}\left(y_t - \mathbf{f}(t)'\tilde{\boldsymbol{\beta}}_k\right)\left(y_{t-1} - \mathbf{f}(t-1)'\tilde{\boldsymbol{\beta}}_k\right)}{\sum_{t=1}^{n-1}\left(y_t - \mathbf{f}(t)'\tilde{\boldsymbol{\beta}}_k\right)^2} \tag{2}
$$

$$
\tilde{\sigma}_k^2 = \frac{1}{n-k-1}\sum_{t=2}^{n}\left[\left(y_t - \mathbf{f}(t)'\tilde{\boldsymbol{\beta}}_k\right) - \tilde{\phi}_k\left(y_{t-1} - \mathbf{f}(t-1)'\tilde{\boldsymbol{\beta}}_k\right)\right]^2 \tag{3}
$$

assuming that the model $\boldsymbol{\mu}_t = \mathbf{f}(t)'\boldsymbol{\beta}$ holds exactly (Nickerson and Basawa, 1992). We denote the number of regressor variables by subscripts here. In the denominator of $\tilde{\phi}_k$ we sum from $t = 1$ to $t = n-1$ instead of $t = n-1$ for reducing bias. Since $\tilde{\phi}_k$ may turn out to be larger than one we restrict it to be at most 0.99 in the following. In order to improve these estimates we can use a two-step approach. First we estimate $\boldsymbol{\beta}$ by ordinary least squares to get an estimate $\tilde{\boldsymbol{\Sigma}}$ of the covariance matrix $\boldsymbol{\Sigma}$ inserting $\tilde{\phi}_k$ and $\tilde{\sigma}_k$. Then we calculate a feasible generalized least squares estimate $\hat{\boldsymbol{\beta}}_k = \hat{\boldsymbol{\beta}}_{\tilde{\boldsymbol{\Sigma}}}$. Inserting $\hat{\boldsymbol{\beta}}_k$ into (2) and (3) instead of $\tilde{\boldsymbol{\beta}}_k$ we get two-step weighted least squares estimates $\hat{\phi}_k$ and $\hat{\sigma}_k$.

A higher dimensional parameterization $\boldsymbol{\mu}_t = \mathbf{f}(t)'\boldsymbol{\beta}$, $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_k)'$, provides flexibility for time windows with non-linear trend. In a steady state, however, it may result in instability of the mean estimates as we fit an overparameterized model. A possibility to overcome this problem is data-driven shrinkage of the parameter estimate in the full model towards the estimate in a reduced model corresponding to a steady state. Particularly, we consider a convex combination $\hat{\boldsymbol{\beta}}_S$ of $\hat{\boldsymbol{\beta}}_k$ and $\hat{\boldsymbol{\beta}}_0 = (\hat{\mu}, 0, \ldots, 0)'$ corresponding to fitting a $(k+1)$-dimensional trend function and a constant mean respectively to the current time window

$$\hat{\boldsymbol{\beta}}_S = \hat{\boldsymbol{\beta}}_k - \frac{c\hat{\sigma}_k^2}{\left(\hat{\boldsymbol{\beta}}_k - \hat{\boldsymbol{\beta}}_0\right)' \left(\mathbf{X}'\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{X}\right) \left(\hat{\boldsymbol{\beta}}_k - \hat{\boldsymbol{\beta}}_0\right)} \left(\hat{\boldsymbol{\beta}}_k - \hat{\boldsymbol{\beta}}_0\right) . \tag{4}$$

The amount of shrinkage is controlled by a factor which is just c times the inverse of a χ^2-statistic which compares the fit of the reduced model to the fit of the full model. We restrict this factor to be not larger than one. Then shrinkage estimators $\hat{\phi}_S$ and $\hat{\sigma}_S$ can be obtained by inserting $\hat{\boldsymbol{\beta}}_S$ into formulas (2) and (3). Shrinkage estimation of the parameters of a regression model with correlated errors has been treated by NICKERSON and BASAWA (1992) and by CHATURVEDI and WAN (2000). In the former paper it is shown that the shrinkage estimator of the mean has smaller expected weighted mean square error than the weighted least squares estimator if the covariance matrix is known up to a scale parameter, $n \geq k$ and $0 \leq c \leq 2(k-1)(n-k)/(n-k+3)$. For unknown covariances, it is shown there that the shrinkage estimator of the mean has better asymptotic efficiency than the weighted least squares estimator if the true mean is approximately constant. In the latter paper it is shown that the shrinkage estimator of the mean dominates the feasible generalized least squares estimator in case of known variances and unknown correlations. Thus, shrinkage may well improve the estimation of the mean; see SARGAN (2001) for a discussion of this topic. We set $c = 4$, that is at the upper end of the interval suggested by Nickerson and Basawa since we have $k = 3$.

If it was known in advance that all changes of the mean in the course of time are monotonic we could use isotonic regression (WU, WOODROOFE and MENTZ, 2001) to restrict the estimate. However, in our application the mean can show smooth non-monotonic behavior even in the steady state and then this restriction may be unduly severe.

We are not primarily interested in the mean estimates but we use them to detrend the data and to estimate the AR parameters thereafter. Therefore we check in Section 3 whether these improved estimates of the regression function also result in better estimates of the autocovariances. We compare the performance of $(\tilde{\phi}_1, \tilde{\sigma}_1^2)$, $(\tilde{\phi}_3, \tilde{\sigma}_3^2)$, $(\hat{\phi}_3, \hat{\sigma}_3^2)$ and $(\hat{\phi}_S, \hat{\sigma}_S^2)$, where the index denotes the number of regressors in the trend function, while S denotes shrinkage. The first estimator is computationally very cheap as we detrend the data by just fitting a simple linear trend, the second needs fitting a four dimensional trend function, the third is a two-step estimator, that needs iteration to obtain feasible generalized least squares in a second step, while shrinkage estimation only affords to additionally evaluate equation (4) once in each step.

3 A simulation study

In the following we perform a Monte Carlo study to check the reliability of the parameter estimators, to derive critical values for the test statistic and to check the power of the proposed procedure.

3.1 The simulation design

We simulate time series $Y_1, \ldots, Y_{300}$ of length $N = 300$ corresponding to five hours of monitoring of a process with observations measured every minute from

$$\begin{aligned} Y_t &= \mu_t + E_t \\ E_t &= \phi E_{t-1} + U_t, \end{aligned}$$

where U_t denotes $N(0,1)$-distributed disturbances. We consider $\phi \in \{0.0, 0.1, \ldots, 0.9\}$ and several deterministic mean functions μ_t. The latter either include no trend, a linear trend $\mu_t^{(1)} = a(t-100)1_{100<t<200} + 100a1_{t\geq 200}$ starting at time point $t = 101$ with a duration of 100 time points and slope $a \in \{0, 0.05, 0.1\}$, or a non-linear trend $\mu_t^{(2)}$ having the shape of the ascending part of the sinus-function, i.e. zero derivatives at both endings, and causing the same total change in level as a linear trend with slope $a \in \{0.05, 0.1\}$ during the same time interval. For monitoring, we move a time window of length n=60 observations corresponding to one hour of measurements through the series.

First we illustrate the problems resulting from positive autocorrelations and the corresponding behavior of shrinkage estimates as proposed above. Figure 1 shows simulated time series with an inserted linear trend having slope $a = 0.05$. The autoregressive parameter is chosen to be $\phi = 0.0$ (independent errors), $\phi = 0.3$ (small autocorrelations), $\phi = 0.6$ (moderate autocorrelations) and $\phi = 0.9$ (large autocorrelations) respectively. The realizations of the test statistic T with shrinkage-based standardization as described above are also shown. For $\phi = 0.9$ the underlying deterministic trend is barely visible as there are monotonic increasing as well as monotonic decreasing patterns. Nevertheless, T increases during the trend period even in case of such high autocorrelations, but it takes more time until it becomes large. The highest value of the test statistic even occurs some time after the trend period, when only some of the observations in the time window are influenced by the trend. The reason is that there is a long monotone pattern in the data caused by the strong positive correlations. In case of small to moderate autocorrelations T increases strongly briefly after the start of the trend.

Figure 2 shows the shrinkage factors obtained for the previous four time series as a function of time. For $\phi \in \{0.0, 0.3, 0.6\}$, i.e. for small to moderately large autocorrelations, this factor is close to one in a steady state, while it is close to zero in a trend period. In case of large autocorrelations ($\phi = 0.9$), the resulting monotonic patterns cause the shrinkage factor to be small most of the time.

A comparison of the mean estimates reveals that a four dimensional regression function fits the data most of the time very well, but during the steady state the data are overfitted as could be expected. Shrinkage reduces this overfitting. On the other hand, fitting a simple linear trend means a crude approximation particularly at the beginning and the end of a trend period.

3.2 Comparison of the parameter estimates

A crucial point for the proper standardization of the weighted sum is the estimation of the AR(1) parameters ϕ and σ^2. In order to explore the properties of the distinct estimators proposed above we simulate 500 time series for each of $\phi = 0.0, 0.3, 0.6, 0.9$ and a linear time trend with slope 0.05 between the time points 101 and 200. We calculate the averages of these parameter estimates across all the 500 time series corresponding to each model for each time point to assess the bias of the distinct methods for such a trend.

Figures 3 and 4 depict the averaged estimates of the autoregressive coefficient ϕ and of the standard deviation τ of the weighted sum, respectively. We discuss them jointly as the results are essentially the same. When all observations within the current time window arise from a steady state, the estimators based on simple linear detrending usually show the

Figure 1: Simulated time series (solid), underlying mean M (bold solid) and test statistic T (dashed) for zero ($\phi = 0.0$, top left), small ($\phi = 0.3$, top right), moderate ($\phi = 0.6$, bottom left) and large ($\phi = 0.9$, bottom right) autocorrelations.

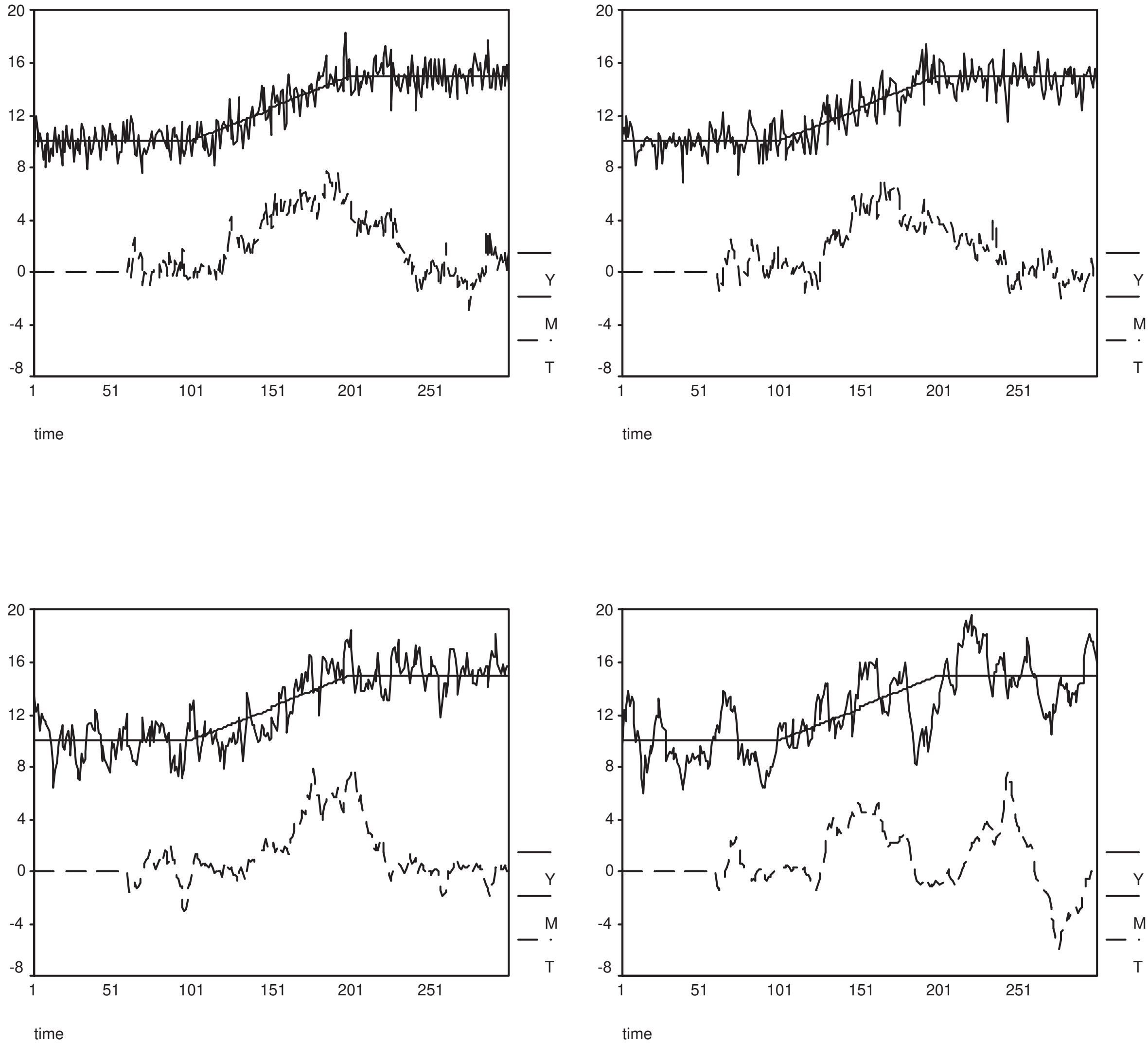
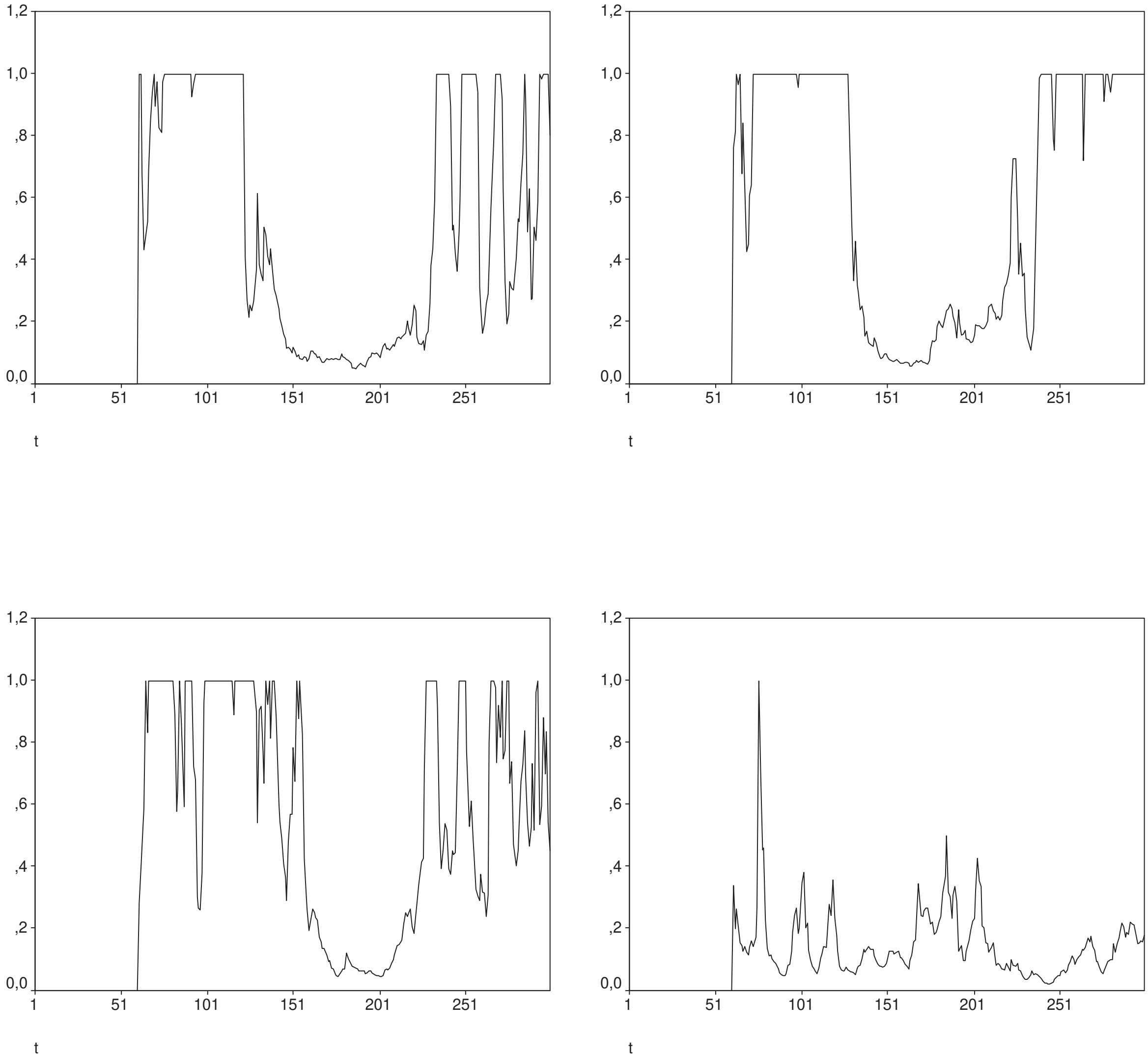


Figure 2: Shrinkage factors for simulated time series with inserted trend and zero ($\phi = 0.0$, top left), small ($\phi = 0.3$, top right), moderate ($\phi = 0.6$, bottom left) and large ($\phi = 0.9$, bottom right) autocorrelations.



smallest (negative) bias. One-step third order detrending results in a substantially larger bias, and two-step estimation provides only minor improvement, that is almost negligible in case of small autocorrelations, cf. $\phi = 0.0$ and $\phi = 0.3$. This is in line with the results of BLOOMFIELD and NYCHKA (1992), who find the ordinary least squares and the optimal (in the sense of mean square error) unbiased estimate of a simple linear trend in case of AR(1) disturbances to be very close to each other if the autocorrelations are moderate to small. GRENANDER (1954) shows that ordinary least squares is asymptotically fully efficient in a broad range of regression models with correlated errors. Shrinkage estimation is usually in between simple linear and third-order detrending in a steady state. For σ (not shown here), it has even a smaller bias than simple linear detrending. For ϕ and τ, all methods show a negative bias, that increases with increasing autocorrelations.

In case of small to moderately large autocorrelations, simple linear detrending results in estimates that are typically larger in a trend period than in a steady state, the bias becoming even positive during a trend if the autocorrelations are close to zero. However, this means that we lessen the effect of a trend on the test statistics as compared to a steady state when using linear detrending for standardization. On the other hand, the negative bias of shrinkage estimation increases even slightly in a trend period if the autocorrelations are not very large. This is due to the fact that the shrinkage estimators are close to the ordinary, not trend corrected estimators during a steady state and close to the two-step third order regression estimators during a trend. Such behavior is advantageous for trend detection since it increases the differences between the values of the test statistic in a trend period and in a steady state even more, while we can cope with a bias in a steady state adjusting the estimator or the critical value. The not trend-corrected estimators are not depicted here since these show a very large positive bias in a trend period, i.e. including them in the figures would prohibit discerning differences between the other methods. An additional analysis for a non-linear,

Table 1: Simulated and fitted bias (multiplied by -1) of the shrinkage estimator $\hat{\phi}_S$ in case of no trend for several autocorrelations.

ϕ	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
sim.	0.0365	0.0496	0.0511	0.0668	0.0763	0.0977	0.1051	0.1254	0.1619	0.2084
fit.	0.0432	0.0436	0.0496	0.0594	0.0748	0.0922	0.1175	0.1435	0.1670	0.1887

sinusoidal trend leads to essentially the same results, while an analysis of the variances of the estimators does not reveal large differences.

In view of these results we prefer shrinkage estimation for detrending in case of observations measured every minute, i.e. in a situation which is not extremely time critical. The increasingly negative bias of these estimators in trend periods increases the power of the procedure if we use them to standardize the weighted sum. Two-step estimation without shrinkage improves the results only in case of very large autocorrelations. Although the estimates obtained from fitting a straight line usually increase in a trend period, the improvement in comparison to the ordinary, not trend corrected estimates is large.

Although there are bias corrected estimators for AR parameters (Fuller, 1996, chapter 6.2), we are not aware of such methods for shrinkage estimation. In order to get a bias-corrected version of $\hat{\phi}_S$ we estimate its bias in a Monte Carlo study. For each of $\phi = 0.0, \ldots, 0.9$, we simulate 200 time series of length 300. Then we calculate the sample mean of the resulting $200 \cdot 241 = 48200$ shrinkage estimates, cf. Table 1. The standard error is about 0.0006 for each value of ϕ. Plotting the simulated bias against ϕ reveals that a quadratic function might be appropriate. We find the linear term not to be significantly distinct from zero and the adjusted R^2 to increase from 0.949 to 0.955 when neglecting it. Fitting a pure quadratic function (cf. Table 1) we get the approximately bias-corrected estimates $\hat{\phi}_{S,bc} = \hat{\phi}_S(1 + 0.305\hat{\phi}_S) + 0.0424$.

Figure 3: Average estimates of ϕ in case of a linear time trend between $t = 100$ and $t = 200$: Zero ($\phi = 0.0$, top left), small ($\phi = 0.3$, top right), moderate ($\phi = 0.6$, bottom left) and large ($\phi = 0.9$, bottom right) autocorrelations. True parameter (bold solid), simple linear detrending (dotted), one-step / two-step third order detrending (dashed / solid) and shrinkage estimates (dashed-dotted).

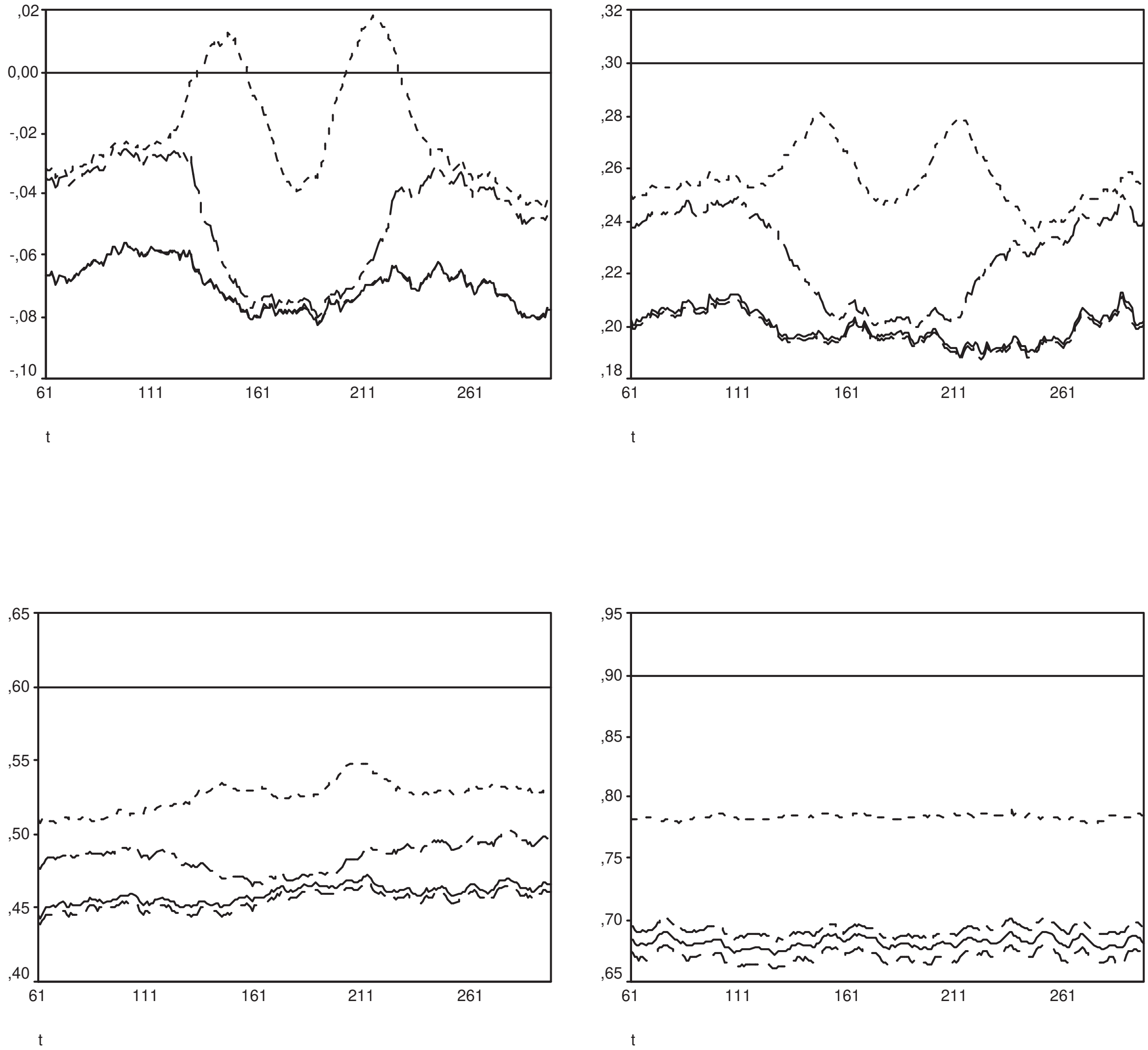
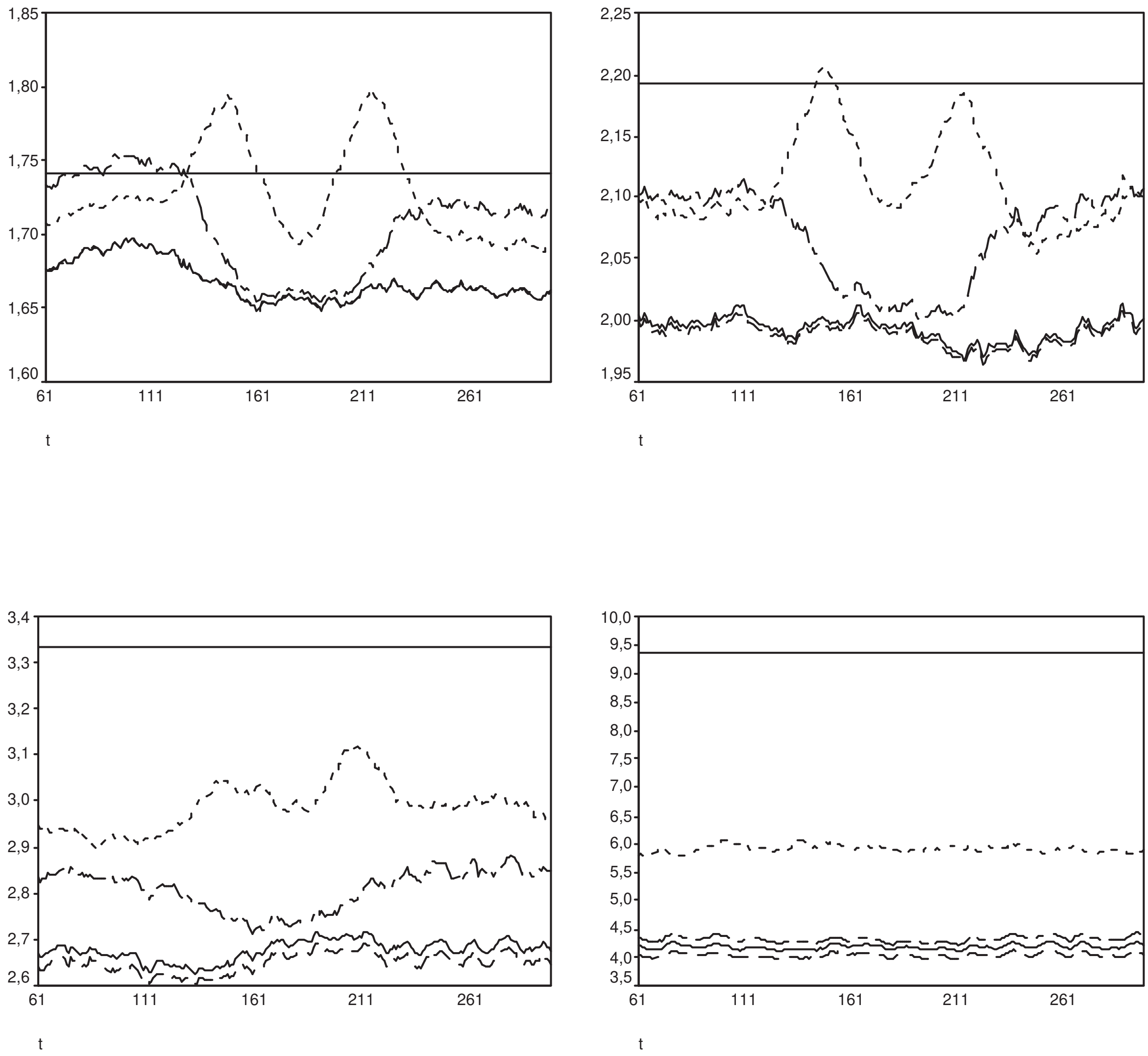


Figure 4: Average estimates of τ in dependence on ϕ in case of a linear time trend between $t = 100$ and $t = 200$: Zero ($\phi = 0.0$, top left), small ($\phi = 0.3$, top right), moderate ($\phi = 0.6$, bottom left) and large ($\phi = 0.9$, bottom right) autocorrelations. True parameter (bold solid), simple linear detrending (dotted), one-step / two-step third order detrending (dashed/solid) and shrinkage estimates (dashed-dotted).



3.3 Critical values

As stated above, Brillinger's (1989) rule for retrospective trend detection turns out to be too sensitive even for long time series. This problem may become even more serious when time series with time-varying variability are monitored online. Here, we perform multiple testing since we test at each time point whether a trend has occurred recently. To overcome this problem, we derive approximate critical values for the test statistics via simulations as these are hard to derive analytically. We simulate 5001 time series for each of $\phi = 0.0, 0.1., \ldots, 0.9$ and a constant mean. In view of the results obtained in the last subsection we use shrinkage estimation to standardize the weighted sum and apply the bias correction derived there. For each time series the maximal absolute value of the test statistic is calculated.

Table 2 provides some percentiles of the empirical distribution of these maxima for each AR parameter. These $(1-\alpha)$-percentiles are approximative 2α-significance limits for a test whether a monotonic trend occurs during 300 observations corresponding to five hours of measurement in our application. The percentiles are rather stable for small to moderately large autocorrelations, say $0 \le \phi \le 0.6$, but they increase strongly for very large autocorrelations. Thus, we should choose critical values in dependence on an estimate of ϕ. If we estimate ϕ to be less than 0.5 we regard $c = 5$ to be a conservative 5% significance bound for wrong detection of a trend within five hours of measurements. If the estimate of ϕ is substantially larger we should use a larger critical value, that can be chosen from Table 2 by interpolation.

Table 2: Percentiles of $Max\, T_i$ in case of a constant mean for $\phi = 0.0, \ldots, 0.9$.

ϕ	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
90%	3.476	3.535	3.644	3.707	3.789	3.961	4.163	4.496	5.269	7.379
95%	3.724	3.840	3.925	4.027	4.143	4.338	4.590	5.051	6.058	8.593
97.5%	3.961	4.032	4.184	4.347	4.521	4.640	5.102	5.571	6.725	9.856
99%	4.235	4.317	4.604	4.771	4.991	5.170	5.630	6.344	7.748	11.415
99.5%	4.537	4.652	4.832	5.008	5.285	5.573	6.006	6.984	8.391	12.360

3.4 Statistical power

Now we inspect the power of the proposed procedure. We simulate 200 time series of length 300 for each of several models. Either a linear or a sinusoidal trend is inserted between $t = 101$ and $t = 200$ causing a total change of 5σ or 10σ. We use the weighted sum test statistic T calculated from $n = 60$ subsequent observations for monitoring and apply the bias-corrected shrinkage estimators for standardization. We choose the significance level $\alpha = 5\%$ and select the critical value corresponding to the estimate $\hat{\phi}_S$ for the current time window from Table 2 by interpolation. Then we count in how many time series a trend is detected between $t = 101$ and $t = 260$ for the first time as a signal outside this period means a false alarm. To check the validity of these critical values we also analyze time series without a trend. Here, we calculate the number of time series for every value of ϕ in which a trend is detected at any time point since any alarm is false then.

Table 3 provides the numbers of time series in which a trend was detected. All trends considered here can be detected reliably if the autocorrelations are small or moderate. For large autocorrelations, say $\phi \geq 0.8$, it is difficult to distinguish whether monotonic sequences are due to autocorrelations or due to a trend. For $\phi = 0.9$ the number of false alarms largely exceeds the percentage of false alarms regulated by the significance level even though

Table 3: Number of identified trends (first line) and average delay of trend detection (second line) for several trend sizes (measured in multiples of the standard deviation σ): No trend (top), linear trend (center) and sinusoidal trends (bottom)

	ϕ									
Size	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0σ	7	7	10	5	16	8	12	20	24	63
5σ	200	199	198	198	198	187	164	127	88	92
	41.3	43.7	44.5	49.0	56.2	58.4	65.6	68.5	70.7	85.4
10σ	199	199	199	196	198	200	194	185	153	132
	31.3	32.4	33.5	36.6	38.4	42.3	49.4	54.4	62.1	68.4
5σ	197	197	200	200	191	180	151	115	97	84
	34.4	36.9	40.4	45.9	50.8	57.2	65.1	70.4	77.2	76.2
10σ	200	199	198	198	198	197	194	186	176	137
	20.8	22.7	25.3	27.4	31.7	35.2	44.4	51.5	57.0	67.7

we derived critical values from simulations. This confirms the findings of Woodward and Gray (1993), who note that deterministic trends are very difficult to detect in short time series with large autocorrelations. In our case, this large percentage of false alarms may be caused by the need to estimate the lag one autocorrelation for choosing the critical value, while these values were derived assuming this autocorrelation to be fixed (and known). The results might improve if we demand that T exceeds the critical value at some subsequent time points and not a single time point only in order to reduce the impact of minor fluctuations. Obviously non-linear, sinusoidal trends are more difficult to detect than linear trends which affect the level by the same amount. This might be caused by the smooth beginning of this non-linear trend form, which has zero derivatives at its endings. However, the power for sinusoidal trends increases with increasing steepness even for very large autocorrelations.

In Table 3 the average delay of trend detection is also provided. Here, we consider only those cases where a correct alarm is given. The delay increases slightly for small to moderate autocorrelations, and significantly for large autocorrelations. While for small autocorrelations on average about 45 (34) observations are sufficient to detect a linear trend with slope 0.05 (0.1), more than 60 (50) observations are needed in case of large autocorrelations. We conclude that reliable and fast discrimination of a trend from a steady state affords a trend to be stronger then than those rather slow trends considered here when the autocorrelations are high. The average delay for a sinusoidal trend is typically smaller, but we have to keep in mind the lower detection rate for this trend form. The time delay could be reduced by using a shorter time window but the expense would be a larger number of false alarms, particularly for large autocorrelations.

The trends shown in Figure 1 are detected with a delay of 44 ($\phi = 0.0$), 44 ($\phi = 0.3$), and 72 ($\phi = 0.6$) observations respectively, while in the time series with $\phi = 0.9$ no trend is detected at all as the large positive autocorrelations cause here many monotonic patterns with durations less than 30 observations. Here, the systematic changes are small in comparison to the random variability found in the data.

4 Application to real time series

In order to judge the performance of the proposed procedure for real data we analyzed a couple of real physiological time series representing physiological variables like the heart rate or blood pressure and found the procedure to perform well. We present a single example in the following. Figure 4 shows two hundred measurements of the systolic arterial pressure of a critically ill patient taken every minute. Analyzing a large number of such time series, IMHOFF *et al.* (2002) find that the autocorrelations can be described well by AR models of low order during a steady state. This particular time series first drifts slightly around a steady state and then starts increasing slowly at about $t = 100$. Then from about $t = 164$ on it increases more strongly. An analysis of the autocorrelations shows that an AR(1) model with ϕ between 0.1 and 0.4 seems adequate. Figure 4 also shows the standardized weighted sum test statistic calculated using a moving time window of length $n = 60$. The test statistic remains well within the non-critical limits up to time point 163, then it increases and crosses the critical value $c = 5.0$ at $t = 165$.

5 Discussion

We have proposed a procedure for online detection of monotonic trends in time series with slowly time-varying parameters. Autocorrelations are modelled using a low order AR model for the noise process. We have also investigated some variations of linear detrending of the data for estimation of the autocorrelations and have found that shrinkage estimation improves the discriminatory power of the test statistic.

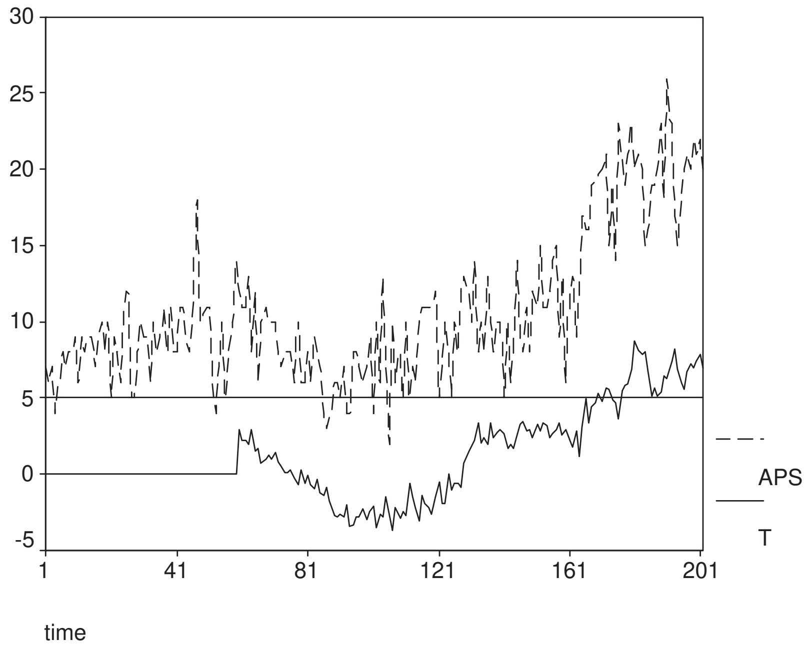


Figure 5: Time series representing arterial pressure and standardized weighted sum statistic T. We subtracted 100 from the pressure for the reason of illustration. At $t = 165$ an upward trend is detected.

Although some work remains to be done, we consider the results to be encouraging so far. The shrinkage estimators for the AR parameters are only mildly influenced by deterministic trends, and the proposed test statistic usually results in much larger absolute values in trend periods than in a steady state. Both linear and nonlinear trends were identified correctly rather soon in most of the cases. The procedure shows very satisfactory performance in case of moderately large autocorrelations. We applied the procedure to a couple of long physiologic time series observed in intensive care and found the results to agree well with the opinion of an experienced physician. Problems may arise w.r.t. the clinical relevance of the detected trend periods. Inserting a constant percentage of the current mean value instead of an estimate of the observational variance $\gamma(0)$ into formula (1) can enhance the results. Further improvements are possible by smoothing the parameter estimates resulting from subsequent time windows.

Problems may also arise in case of very large positive autocorrelations. This is also true for other procedures designed for retrospective trend detection. Considering several test statistics for the retrospective detection of a deterministic trend, Woodward and Gray (1993) found the rate of false alarms to be often larger than 50% for large positive autocorrelations. The detection of a deterministic trend in a time series with very large autocorrelations is a hard problem anyway since the behavior of the undisturbed process is already close to non-stationarity with frequently occurring monotonic sequences. Hence, using a time window of length $n = 60$ may not be sufficient to distinguish slow deterministic trends and large positive autocorrelations as both result in similar patterns.

A critical issue for the proposed procedure is the proper specification of the window width n, similarly as a suitable forgetting factor λ is needed for an EWMA chart. However, we consider the choice of a time period to be easier for an operator than the choice of a weighting factor. In our biomedical application, an experienced physician may well consider a monotonic change of a vital sign over one hour to be clinically relevant. Moreover, the power of EWMA and CUSUM charts is best for sudden shifts of the mean, whereas this is the worst case for the weighted sum statistic. This statistic mirrors the start and the end of the time window as the weights have different signs in the first and the second half of the time window. This is similar to comparing time delayed means, that has also been suggested for trend detection. Comparison of time delayed means corresponds to special weighted sums with weights $-1/m_1, \ldots, -1/m_1, 0, \ldots, 0, 1/m_2, \ldots 1/m_2$. Therefore, the resulting test statistics have lower worst-case discriminatory power than the statistic advocated here as the latter uses optimal weights.

Similarly as EWMA and CUSUM charts the procedure as presented here is not robust against outliers. A simple possibility to reduce this deficiency is to additionally apply a procedure that detects and replaces outliers online (Gather, Bauer and Fried, 2003).

Alternatively, the test statistic itself can be robustified using M-estimation for instance. This will have to be addressed in further studies.

Acknowledgments

The financial support of the Deutsche Forschungsgemeinschaft (SFB 475, "Reduction of complexity in multivariate data structures") is gratefully acknowledged.

References

ABELSON, R.P., and TUKEY J.W., 1963: Efficient utilization of non-numerical information in quantitative analysis: general theory and the case of simple order. *Annals of Mathematical Statistics* **34**, 1347–1369.

BLOOMFIELD, P., and NYCHKA, D.W., 1992: Climate spectra and detecting climate change. *Climatic Change* **21**, 275–287.

BRILLINGER, D.R., 1989: Consistent detection of a monotonic trend superposed by a stationary time series. *Biometrika* **76**, 23–30.

CEMBROWSKI, G.S., WESTGARD, J.O., EGGERL, A.A., and TOREN, E.C., 1975: Trend detection in control data: Optimization in Trigg's technique for trend analysis. *Clinical Chemistry* **21**, 1395–1405.

CHANG, J.T., and FRICKER, R.D., 1999: Detecting when a monotonically increasing mean has crossed a threshold. *Journal of Quality Technology* **31**, 217–234.

CHATURVEDI, A., and WAN, A.T.K., 2000: Exact results on the inadmissibility of the feasible generalized least squares estimator in regression models with non-spherical disturbances. *Biometrical Journal* **42**, 481–487.

Cochrane, D., and Orcutt, G.H., 1949: Application of least squares regression to relationships containing autocorrelated error terms. *Journal of the American Statistical Association* **44**, 32–61.

Dahlhaus, R., 1997: Fitting time series models to nonstationary processes. *The Annals of Statistics* **25**, 1–37.

Fuller, W.A., 1996: *Introduction to Statistical Time Series.* 2nd Edition. Wiley, New York.

Gather, U., Bauer, M., and Fried, R., 2003: The identification of multiple outliers in online monitoring data. *Estadistica*, to appear.

Gather, U., Fried, R., Lanius, V., and Imhoff, M., 2003: Online monitoring of high-dimensional physiological time series - a case-study. *Estadistica*, to appear.

Gather, U., Imhoff, M., and Fried, R., 2002: Graphical models for multivariate time series from intensive care monitoring. *Statistics in Medicine* **21**, 2685–2701.

Grenander, U., 1954: On the estimation of regression coefficients in the case of an autocorrelated disturbance. *Annals of Mathematical Statistics* **25**, 252–272.

Högel, J., 2000: Applications of statistical process control techniques in medical fields. *Allgemeines Statistisches Archiv* **84**, 337–359.

Imhoff, M., Bauer, M., Gather, U., and Fried, R., 2002: Pattern detection in intensive care monitoring time series with autoregressive models: Influence of the model order. *Biometrical Journal* **44**, 746–761.

Lu, C.-W., and Reynolds, M.R., 1999: EWMA control charts for monitoring the mean of autocorrelated processes. *Journal of Quality Technology* **31**, 166–187.

Montgomery, D.C., and Mastrangelo, C.M., 1991: Some statistical process control methods for autocorrelated data. *Journal of Quality Technology* **23**, 179–193.

NICKERSON, D.M., and BASAWA, I.V., 1992: Shrinkage estimation for linear models with correlated errors. *Shankhya* **54**, 411–424.

SARGAN, J.D., 2001: Model building and data mining. *Econometric Reviews* **20**, 159–170.

SCHACK, B., and GRIESZBACH, G., 1994: Adaptive methods for trend detection and their applications in analysing biosignals. *Biometrical Journal* **36**, 429–452.

SCHMID, W., and STELAND, A., 2000: Sequential control of non-stationary processes by nonparametric kernel control charts. *Allgemeines Statistisches Archiv* **84**, 315–336.

SUN, H., and PANTULA, S.G., 1999: Testing for trends in correlated data. *Statistics & Probability Letters* **41**, 87–95.

TRIGG, D.W., 1964: Monitoring a forecasting system. *Operational Research Quarterly* **15**, 271–274.

WOODWARD, W.A., and GRAY, H.L., 1993: Global warming and the problem of testing for trend in time series data. *Journal of Climate* **6**, 953–962.

WOODWARD, W.A., BOTTONE, S., and GRAY, H.L., 1997: Improved tests for trends in time series data. *Journal of Agricultural, Biological and Environmental Statistics* **2**, 403–416.

WU, W.B., WOODROOFE, M., and MENTZ, G., 2001: Isotonic regression: Another look at the changepoint problem. *Biometrika* **88**, 793–804.