ROBUST SHIFT DETECTION IN AUTOREGRESSIVE PROCESSES

R. Fried
Department of Statistics, University of Dortmund
44221 Dortmund, GERMANY, e-mail: fried@statistik.uni-dortmund.de

Abstract

We use a local approach and highly robust estimators for the detection of level shifts in autocorrelated data. The arising detection rules identify shifts with a short delay, resist some outliers and adapt to time-varying model parameters. The performance of the rules is investigated using analytical resistances and via simulations.

1 Introduction

The detection of level shifts in time series has attracted considerable attention. Iterative procedures based on maximum likelihood have been suggested which try to detect and distinguish patterns like shifts and several types of outliers sequentially starting from a model not including any pattern. Such procedures are designed for retrospective analysis of the whole series, they possibly confuse different patterns, and they can fail because of several patterns masking each other [5], [7].

Aiming at rules for automatic level shift detection which are not mislead by other patterns, we use a local approach and highly robust estimators. Median filtering [9] is popular for approximating the level underlying a time series because of its simplicity and its robustness against outliers. The median preserves its good properties under positive autocorrelations [2]. Tests based on differences between medians from subsequent time windows standardized by a robust scale estimate provide high robustness and good detection power [3]. They outperform other rules for shift detection in these respects. However, they implicitly treat the data as independent, while often we find subsequent observations to be positively autocorrelated.

Continuing previous work on procedures for highly robust online level estimation with short delays [1], [4], we investigate the effects of autocorrelations on level shift detection. We assume that the observed time series can be locally modeled by an autoregressive (AR) process as described in Section 2. The mentioned median comparisons are adapted to AR noise for incorporating autocorrelations in Section 3. Besides the level and the variance we need to estimate the local AR parameters then, adding an additional source of variability in the reasoning process. The arising detection rules are compared to those treating the observations as independent in Section 4.

2 Autoregressive Processes with Level Shifts

Let $(X_t)_{t\in\mathbb{N}}$ be an autoregressive process of order p (AR(p)),

$$\Phi(B)(X_t - \mu) = a_t \; ; \; t \in \mathbb{Z} \; . \tag{1}$$

Here, μ is the level of the time series and B is the backshift operator, $Bx_t = x_{t-1}$. The zeros of the characteristic polynomial $\Phi(B) = (1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p)$ are assumed to lie outside the complex unit circle to guarantee stationarity. The a_t are independent identically distributed random variables with mean zero and variance σ_a^2. Often a Gaussian distribution is assumed for the a_t. We assume that the process can be locally modeled by an AR(p) process within all time windows of fixed width n. We denote the variables in a given time window simply by $X_1, \ldots, X_n$.

A level shift is an abrupt change of the central location of the time series from one level to another. If a level shift occurs at a time point $\tau \in \{2, \ldots, n\}$, we observe a disturbed process (Z_t) instead of the stationary process (X_t),

$$Z_t = \begin{cases} X_t, & t < \tau \\ X_t + \omega, & t \geq \tau \end{cases} , \tag{2}$$

where ω is the size of the shift. Let $S_t^{(\tau)}$ be a step function, which takes the value one for $t \geq \tau$ and is zero otherwise. Equation (2) can be written as

$$Z_t = X_t + \omega S_t^{(\tau)} . \tag{3}$$

Differencing $S_t^{(\tau)}$ gives the indicator function $I_t^{(\tau)}$, which takes the value one at $t = \tau$ and is zero otherwise, $(1-B)S_t^{(\tau)} = I_t^{(\tau)}$. Inserting this equation and the definition (1) into (3) leads to

$$Z_t - \mu = \Phi^{-1}(B)a_t + \omega(1-B)^{-1}I_t^{(\tau)} . \tag{4}$$

Multiplication by $\Phi(B)$ gives an equation for the observable residuals $e_t = \Phi(B)(Z_t - \mu)$:

$$e_t \quad = \quad a_t + \omega\Phi(B)(1-B)^{-1}\, I_t^{(\tau)} . \tag{5}$$

Denoting $\ell(B) = \Phi(B)(1-B)^{-1} = 1 - \ell_1 B - \ell_2 B^2 - \ldots$ we get

$$e_t = \begin{cases} a_t, & t < \tau \\ \omega + a_\tau, & t = \tau \\ -\omega\ell_{t-\tau} + a_t, & t > \tau \end{cases} . \tag{6}$$

3 Level Shift Detection

3.1 Rules derived from Least Squares

Determination of the least squares (LS) estimator $\hat{\omega}_{LS}$ of ω from (6) is straightforward:

$$\hat{\omega}_{LS} \quad = \quad \frac{e_\tau - \ell_1 e_{\tau+1} - \ell_2 e_{\tau+2} - \ldots - \ell_{n-\tau} e_n}{1 + \ell_1^2 + \ell_2^2 + \ldots + \ell_{n-\tau}^2} . \tag{7}$$

Defining $\rho_{LS}^2 = (1 + \ell_1^2 + \ell_2^2 + \ldots + \ell_{n-\tau}^2)^{-1}$, the variance of $\hat{\omega}_{LS}$ reads

$$Var(\hat{\omega}_{LS}) = \frac{\sigma_a^2}{1 + \ell_1^2 + \ell_2^2 + \ldots + \ell_{n-\tau}^2} = \rho_{LS}^2 \sigma_a^2 . \tag{8}$$

In case of an AR(1) model, where $\Phi(B) = 1 - \phi B$, we get $\ell(B) = 1 + (1-\phi)B + (1-\phi)B^2 + \ldots$, leading to the estimate

$$\hat{\omega}_{LS} = \frac{\sum_{t=\tau}^{n} e_t - \phi \sum_{t=\tau+1}^{n} e_t}{1 + (n-\tau)(1-\phi)^2}$$

and to the standardized test statistic T_τ for testing if there is a level shift at time τ

$$T_\tau = \frac{\sum_{t=\tau}^{n} e_t - \phi \sum_{t=\tau+1}^{n} e_t}{\sqrt{1 + (n-\tau)(1-\phi)^2}\sigma_a} \ .$$

The standard deviation σ_a of the innovations (a_t) can be estimated as the empirical standard deviation of the observable residuals (e_t). A simple estimator of ϕ is the Yule-Walker (YW) estimator, which is the lag-one sample autocorrelation with centering by the mean $\overline{z}$ of $z_1, \ldots, z_n$,

$$\hat{\phi}_{YW} = \frac{\sum_{t=2}^{n} (z_{t-1} - \overline{z}) \cdot (z_t - \overline{z})}{\sum_{t=1}^{n} (z_t - \overline{z})^2} \ .$$

The denominator is calculated from all data. Thus $\hat{\phi}_{YW}$ almost surely fulfills the stationarity restriction $\hat{\phi}_{YW} \in (-1, 1)$. The price is a bias towards zero for small n.

3.2 Rules derived from Robust Estimates

A robust test for shift detection can be constructed from (6) replacing least squares by median regression, leading to the estimate

$$\hat{\omega}_M = med\{e_\tau, -e_{\tau+1}/\ell_1, \ldots, -e_n/\ell_{n-\tau}\} \ . \quad (9)$$

In case of a stationary AR(1) process with Gaussian innovations, $\hat{\omega}_M$ is asymptotically normal with variance $2\pi\sigma_a^2/(4(n-\tau+1)(1-\phi)^2)$ under the null hypothesis $H_0 : \omega = 0$.

Again, σ_a^2 can be estimated from the residuals (e_t). We suggest the Q_n method [8] for this task, which is based on an order statistic of all pairwise differences. For data $y_1, \ldots, y_n$ it reads

$$Q_n(y_1, \ldots, y_n) = c_n \cdot \{|y_i - y_j| : 1 \le i < j \le n\}_{(k)} \text{ with } k = \binom{\lfloor n/2 \rfloor + 1}{2} ,$$

where c_n is a finite sample correction to achieve unbiasedness in a Gaussian sample of size n. Q_n possesses a high asymptotic Gaussian efficiency of 82% and can be computed in $O(n \log n)$ time. In case of $\phi = 0$, the detection rule resulting from standardizing $\hat{\omega}_M$ by $Q_n(z_1, \ldots, z_n)$ corresponds to the median comparison (MC) suggested in [3].

In case of $\phi \neq 0$ we additionally need a robust estimate of ϕ. SSD is a highly robust estimation method for the autocovariances and autocorrelations [6]. It is based on writing the autocovariance at time lag h as

$$\gamma(h) = \frac{1}{4}\left[Var(X_t + X_{t-h}) - Var(X_t - X_{t-h})\right] \tag{10}$$

and estimates these variances by Q_n applied to $z_2+z_1, \ldots, z_n+z_{n-1}$ and $z_2-z_1, \ldots, z_n-z_{n-1}$, respectively. Scaling $\hat{\gamma}_{SSD}(1)$ by the sum of the variances in (10) (see [6]) turns out to be superior to the scaling by $\hat{\gamma}_{SSD}(0)$. Accordingly, we use

$$\hat{\phi}_{SSD} = \frac{Q_{n-1}^2(z_2 + z_1, \ldots, z_n + z_{n-1}) - Q_{n-1}^2(z_2 - z_1, \ldots, z_n - z_{n-1})}{Q_{n-1}^2(z_2 + z_1, \ldots, z_n + z_{n-1}) + Q_{n-1}^2(z_2 - z_1, \ldots, z_n - z_{n-1})},$$

which is guaranteed to lie within $[-1, 1]$. If less than 25% of observations in general position are replaced by outliers, the terms in the numerator and the denominator are bounded and bounded away from zero, guaranteeing an asymptotic breakdown point of 25%.

4 Experiments

To investigate the performance of the detection rules in different situations we perform some Monte Carlo experiments. Our interest is in small to moderate sample sizes since we want to use minimal assumptions, allowing the AR model parameters to vary slowly with time. Accordingly, we apply the detection rules to time windows of moderate widths. We restrict ourselves to AR(1) models in the following, considering windows of width $n = 29$ observations to test whether there is a level shift at $\tau = 21$. For estimation of μ and ϕ only the first $h = 20$ observations are used as reference period, so that a level shift at time $\tau = 21$ does not influence the estimates. The estimates of μ and ϕ are then plugged into the formulae for the residuals (e_t) and the test statistics.

First we generate 50000 time series of length $n = 29$ from each of different AR(1) models with $\phi \in \{-0.9, -0.8, \ldots, 0.9\}$ and Gaussian errors. The 99.9% quantiles of the calculated absolute test statistics are used as thresholds in the following since we aim at incorrect detection of a shift only once within 1000 time points on average. In this way we account for the multiple testing done when analyzing long time series. Figure 1 shows that the quantiles of the test statistics increase with increasingly positive autocorrelations ϕ, particularly in case of the MC test based on the AR residuals (e_t).

Using these quantiles as thresholds, we generate Gaussian AR(1) series with shifts of increasing size $\omega = 0.25, 0.5, \ldots, 10$ at time $\tau = 21$. Figure 1 depicts the percentage cases in which a shift was detected within 2000 independent time series for each ω. Of course, the least squares methods offers higher power in case of Gaussian data as considered here, and the test statistics using the original observations lead to higher power than those based on AR residuals if the data are independent ($\phi = 0$). In

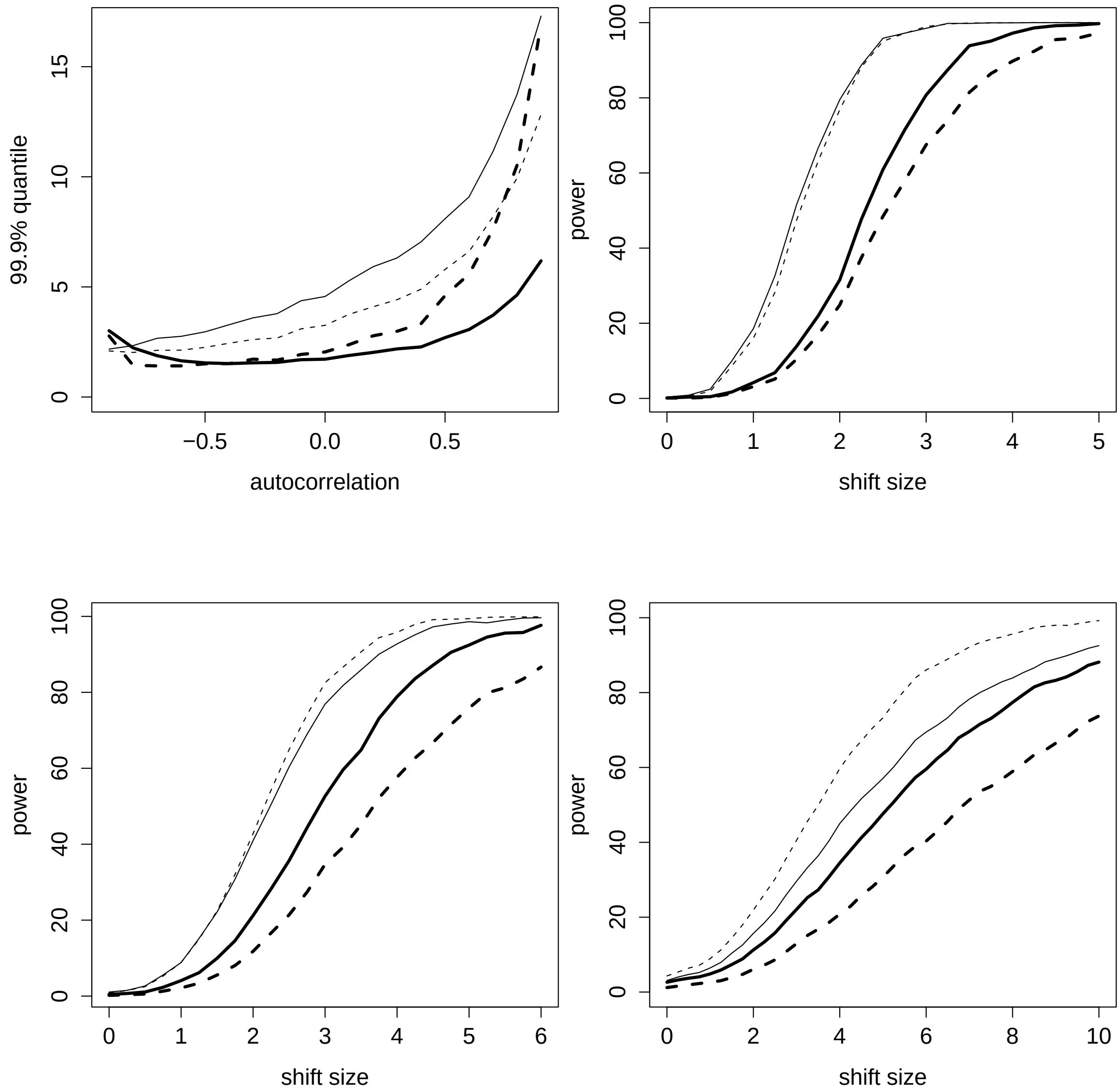

Figure 1: Results for $n = 29$ and $\tau = 21$: 99.9% quantiles of the absolute test statistics (top left) and detection power in case of $\phi = 0.0$ (top right), $\phi = 0.4$ (bottom left) and $\phi = 0.8$ (bottom right): LS test (solid), LS test on AR residuals (dashed), MC test (bold solid) and MC test on AR residuals (bold dashed).

case of moderate ($\phi = 0.4$) or large ($\phi = 0.8$) autocorrelations, the LS test on AR residuals improves the LS test on the original observations, while the MC test on the original observations still outperforms that on the residuals. We note that the tests become liberal with increasing ϕ in spite of using larger thresholds when getting larger estimates of ϕ. The size increases more for those tests showing larger power.

We also investigate the sensitivity of the methods against outliers. First we assume a shift of size 5 to occur within an AR(1) time series with $\phi = 0.5$ and replace the observation at $t = n - 2$ by an additive outlier of increasing size $1, 2, \ldots, 20$. Figure 2 depicts the detection power obtained from 2000 simulation runs each. The MC tests are little affected by a single outlier, while the LS tests loose all their power with increasing outlier size. One must completely rely on detecting outliers by additional outlier detection rules when using traditional LS methods, while the robust MC rules still work well even when some outliers are not detected.

Next we replace an increasing number $j = 0, 1, \ldots, 9$ of observations at the end of a time series generated from an AR(1) model with $\phi = 0.5$, again measuring the percentage cases in which a shift was detected from 2000 simulation runs each, see also Figure 2. Since our interest is in detection rules which resist as many outliers as possible, the curves should stay at zero until half of the observations are shifted, and then suddenly increase to 1. Obviously, the MC tests get close to this ideal, while the LS tests often indicate a shift already in case of a few deviating observations, and they can also miss a shift even if the majority of the observations is shifted.

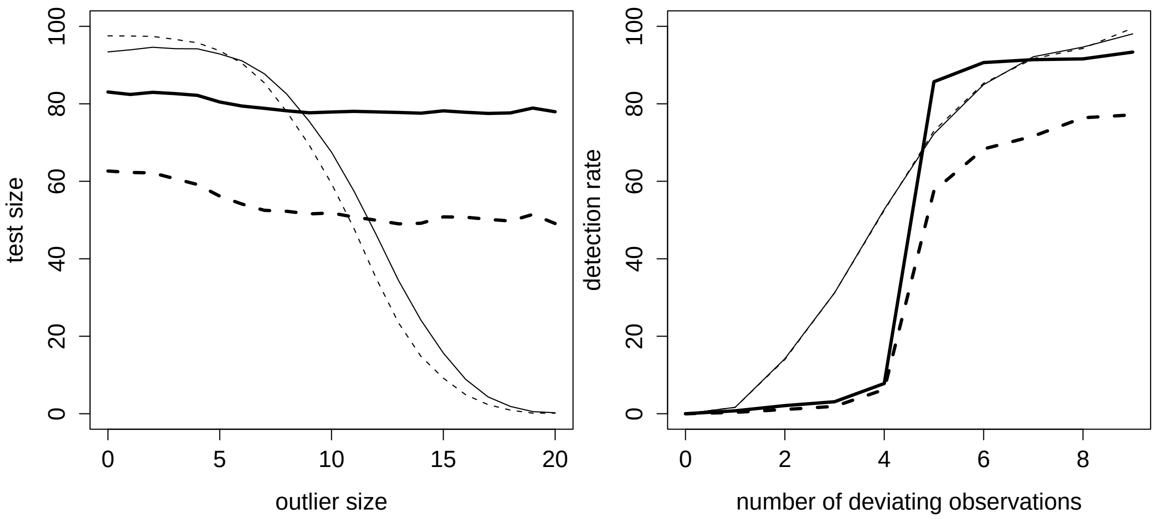

Figure 2: Results for $n = 29$ and $\tau = 21$ if $\phi = 0.5$: Power for a 5σ-shift in case of an outlier of increasing size (left) and percentage cases in which a shift was detected in case of an increasing number of outliers of size 6 (right): LS test (solid), LS test on AR residuals (dashed), MC test (bold solid) and MC test on AR residuals (bold dashed).

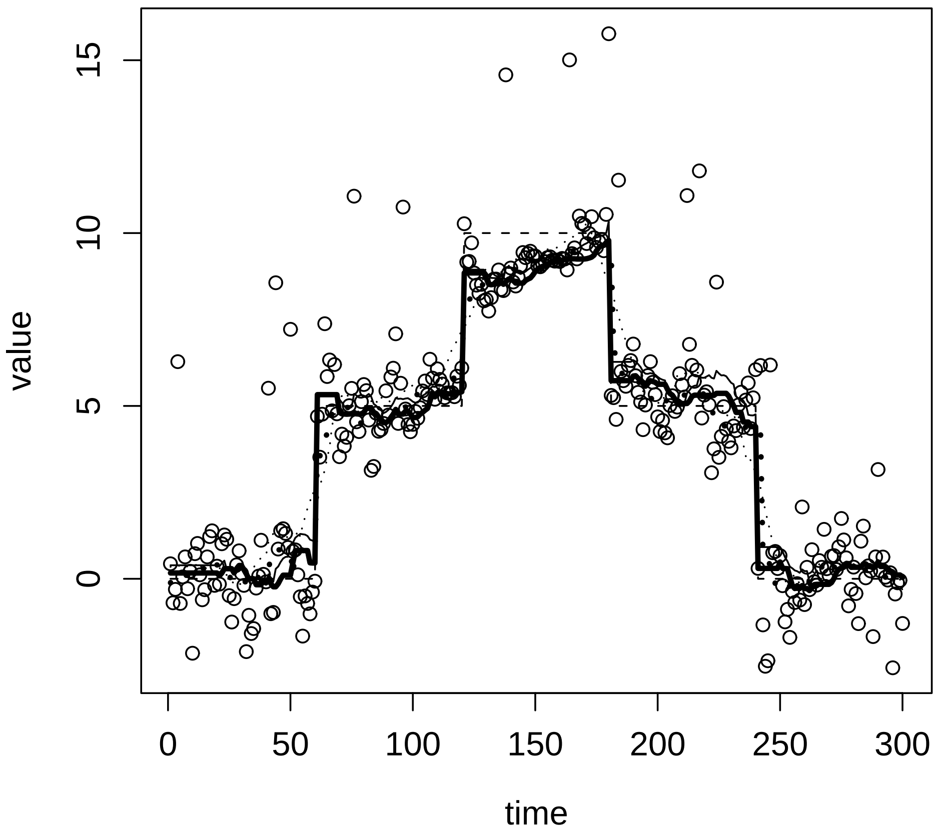

Figure 3: Step function (dashed) overlaid by time-varying AR noise (dots) and several level estimates: running mean (dotted), running median (bold dotted), running mean with LS test (solid) and running median with MC test (bold solid).

5 Conclusions

Comparison of standard medians allows reliable shift detection in the presence of outliers even in case of (possibly time-varying) autoregressive noise if we choose the threshold for shift detection using a highly robust estimate of the AR parameter. Using such tests we can improve the shift preservation of running medians and automatically indicate the presence of a shift.

Figure 3 depicts a step function which is overlaid by AR(1) noise with time-varying parameters $\phi_t = sin(\pi t/300)$ and $\sigma^2_{a,t} = \sqrt{1-\phi_t^2}$ for illustration. Each observation is generated to be an outlier of size 6 with 5% probability. Both the LS and the MC tests with $n = 29$ and $\tau = 21$ detect the shifts at $t = 60, 120, 180, 240$, but the LS tests additionally detect shifts at $t = 22$, between $t = 158$ and $t = 163$, and at $t = 291$, while the MC tests do so only at $t = 160$. Both the LS and the MC test improve the running median and even more the running mean with width $2 \cdot 9 + 1$ (to get the same delay of tracking).

The talk will also address the issue of robust shift detection within trend periods.

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

References

[1] Fried R. (2004). Robust Trend Extraction. *Proceedings of the 7th International Conference Computer Data Analysis and Modeling*, Vol. I. Aivazian S., Filzmoser P., Kharin Y. (Eds.). Academy of Administration at the President of the Republic of Belarus, Minsk, pp. 23-30.

[2] Fried R. (2007). Robust Location Estimation under Dependence. *Journal of Statistical Computation and Simulation*. Vol. **77**, pp. 131-147.

[3] Fried R. (2007). On Robust Shift Detection in Time Series. *Computational Statistics and Data Analysis*, to appear.

[4] Fried R., Gather U. (2005). Robust Trend Estimation for AR(1) Disturbances. *Austrian Journal of Statistics*. Vol. **34**, pp. 139-151.

[5] Gather U., Bauer M., Fried R. (2002). The Identification of Multiple Outliers in Online Monitoring Data. *Estadística*. Vol. **54**, pp. 289-338.

[6] Ma Y., Genton M.G. (2000). Highly Robust Estimation of the Autocovariance Function. *Journal of Time Series Analysis*. Vol. **21**, pp. 663-684.

[7] Peña D. (2000). *A Course in Time Series Analysis*. Wiley, New York.

[8] Rousseeuw P.J., Croux C. (1993). Alternatives to the Median Absolute Deviation. *Journal of the American Statistical Association*. Vol. **88**, pp. 1273-1283.

[9] Tukey J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading, Mass. (preliminary edition 1971).