Robust Shift Detection in Time-Varying Autoregressive Processes

Roland Fried
Technische Universität Dortmund, Germany

Abstract: Tests for shift detection in locally-stationary autoregressive time series are constructed which resist contamination by a substantial amount of outliers. Tests based on a comparison of local medians standardized by a highly robust estimate of the variability show reliable performance in a broad variety of situations if the thresholds are adjusted for possible autocorrelations.

Zusammenfassung: Robuste Tests zur Erkennung von Sprüngen in lokal-stationären autoregressiven Zeitreihen werden konstruiert mit dem Ziel, auch bei erheblicher Ausreißerkontamination zuverlässig zu arbeiten. Ein Vergleich lokaler Mediane standardisiert mit einem robusten Skalenschätzer erweist sich bei Verwendung ausreichend großer kritischer Werte als besonders zuverlässig in relevanten Szenarien.

Keywords: Jumps, Outliers, Test Resistance, Time Series.

1 Introduction

A basic objective of time series analysis is the estimation of the time-varying level (the signal) underlying the series. Sudden changes of this level are of particular interest since they point at changes in the data generating mechanism and since they can severely mislead a statistical analysis if they are not taken into account. Running medians (Tukey, 1977) are popular for signal extraction since they almost preserve sudden shifts and resist outliers, as opposed to e.g. moving averages. However, sometimes one wants a filtering procedure even to indicate the presence of a shift automatically, i.e. to give an alarm.

Similar to the superiority of running medians with respect to moving averages, robust functionals lead to much more reliable tests for shift detection in the presence of outliers than linear functionals (Fried & Gather, 2007, Fried, 2007a). Particularly, tests based on the differences of the medians in separate time windows standardized by a robust scale estimate achieve high robustness and considerable power. Such simple tests implicitly treat the data as independent, although subsequent observations are often positively autocorrelated. Positive autocorrelations cause monotonic patterns in the data resembling the occurrence of shifts in short windows. We should hence adjust the tests for autocorrelations because a shift is detected too often otherwise.

In this paper we discuss median based tests for level shift detection in time series with time-varying dependencies. The median mostly preserves its nice properties in the case of positive autocorrelations (Fried, 2007b). We assume that the autocorrelations can be described locally by a low order autoregressive (AR) model and use highly robust estimators for model fitting. Then we can either adjust the critical values of simple test statistics for the autocorrelations, or we can perform the test on the AR residuals.

Section 2 introduces tests for level shift detection in AR models based on linear statistics or median comparisons. Section 3 critically evaluates the arising detection rules. Section 4 concludes.

2 Level Shift Detection in Autoregressive Models

Let $(Y_t)_{t\in\mathbb{Z}}$ be an autoregressive (AR) stochastic process of order p with parameters $\phi_{1,t},\ldots,\phi_{p,t}$, which vary slowly over time,

$$(1-\phi_{1,t}B-\ldots-\phi_{p,t}B^p)(Y_t-\mu_t)=\epsilon_t\ ,\ t\in\mathbb{Z}\ . \tag{1}$$

Here, B denotes the backshift operator, $BY_t=Y_{t-1}$. We assume the zeros of the characteristic polynomials $\Phi_t(B)=(1-\phi_{1,t}B-\phi_{2,t}B^2-\ldots-\phi_{p,t}B^p)$ to be larger than one in absolute value for all $t\in\mathbb{Z}$ to guarantee local stationarity. We further assume the innovations ϵ_t to be independently normally distributed with common mean zero and time-varying variance σ_t^2. This implies that μ_t is the level of Y_t then. We assume the model parameters to be almost constant in sufficiently short windows of subsequent variables $Y_{t-n+1},\ldots,Y_t$, e.g. approximately $\mu_{t-n+1}=\mu_{t-n+2}=\ldots=\mu_t$, so that the local model parameters can be approximated from a moving time window.

A level shift at a time point $\tau\in\mathbb{Z}$ means a sudden change of the central location of the time series from one level to another one. The observed data stem from a disturbed process (Z_t) then (Peña, 2000)

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

where ω is the size of the shift. In the following we derive a linear regression equation from (2) with uncorrelated error terms, which can be readily used for estimation of ω and testing. Using a step function $S_t^{(\tau)}$, which takes the value one for $t\geq\tau$ and zero otherwise, equation (2) can be written as

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

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

$$Z_t - \mu_t = \Phi_t^{-1}(B)\epsilon_t + \omega(1-B)^{-1} I_t^{(\tau)} .$$

Multiplication by $\Phi_t(B)$ gives a regression equation for the observable residuals $a_t = \Phi_t(B)(Z_t - \mu_t)$, which is linear in the shift size ω and has the uncorrelated errors ϵ_t:

$$a_t = \epsilon_t + \omega\Phi_t(B)(1-B)^{-1}\, I_t^{(\tau)} .$$

Denoting $\ell_t(B) = \Phi_t(B)(1-B)^{-1} = 1 - \ell_{1,t}B - \ell_{2,t}B^2 - \ldots$ we get

$$a_t = \begin{cases} \epsilon_t,\ t < \tau \\ \omega + \epsilon_\tau,\ t = \tau \\ -\omega\ell_{t-\tau,t} + \epsilon_t,\ t > \tau . \end{cases} \tag{4}$$

Estimates of ω and tests of the null hypothesis $H_0 : \omega = 0$ can directly be deduced from this equation.

In the following let $Z_1, \ldots, Z_n$ be a single window of n subsequent observations of a time series. We want to test whether there is a level shift at a given time point $\tau \in \{2, 3, \ldots, n\}$. We assume that the model parameters are (almost) constant throughout the window and drop the index t from all parameters therefore.

2.1 Detection Based on Least Squares

The least squares (LS) estimator $\hat{\omega}_{LS}$ of ω can be derived from (4) straightforwardly,

$$\hat{\omega}_{LS} = \frac{a_\tau - \ell_1 a_{\tau+1} - \ell_2 a_{\tau+2} - \ldots - \ell_{n-\tau} a_n}{1 + \ell_1^2 + \ell_2^2 + \ldots + \ell_{n-\tau}^2} .$$

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

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

We specialize to an AR(1) model in the following, that is $\Phi(B) = 1 - \phi B$. Then we have $\ell(B) = 1 + (1-\phi)B + (1-\phi)B^2 + \ldots$, leading to the normally distributed estimate

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

Standardization leads to a test statistic T_τ for level shift detection at time τ, i.e. for testing the null hypothesis $H_0 : \omega = 0$ of no shift against the alternative $H_1 : \omega \neq 0$:

$$T_\tau = \frac{\sum_{t=\tau}^{n} a_t - \phi \sum_{t=\tau+1}^{n} a_t}{\sigma\sqrt{1 + (n-\tau)(1-\phi)^2}} . \tag{5}$$

For calculation of T_τ we need estimates of ϕ, σ and μ. Simple estimates of the central location μ of the observations and the standard deviation σ of the innovations (ϵ_t) are the sample mean $\overline{z}$ of the z_t and the empirical standard deviation of the observable residuals a_t, respectively. The conditional least squares (CLS) estimator of $\phi = \phi_1$ after centering by the sample mean $\overline{z}$ is

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

We artificially restrict $\hat{\phi}_{CLS}$ not to become larger than 0.99 in absolute value so that the stationarity assumption is guaranteed locally for the fitted model. A variant is to use data from the pre-shift period $t = 1, \ldots, \tau - 1$ only when calculating the estimates of μ, ϕ and σ since the estimates from the full window $t = 1, \ldots, n$ are severely biased under the alternative $H_1 : \omega \neq 0$. Both variants will be discussed in Section 3.

The distribution of the test statistic T_τ with the estimates plugged in can be approximated by a standard normal under the null hypothesis of no shift if n respectively τ is large. Otherwise we can derive critical values for T_τ by simulation under H_0.

2.2 Detection Based on Robust Estimates

For construction of an outlier-resistent test for shift detection from (4) we replace least squares by median regression, leading to the estimate

$$\hat{\omega}_M = med\{a_\tau, -a_{\tau+1}/\ell_1, \ldots, -a_n/\ell_{n-\tau}\} .$$

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

$$\tilde{T}_\tau = \frac{\sqrt{2(n-\tau+1)(1-\phi)^2}\hat{\omega}_M}{\sigma\sqrt{\pi}} \tag{6}$$

can be approximated by a standard normal distribution under H_0 if $n - \tau$ is large.

To use $\tilde{T}_\tau$ as a test statistic for H_0 we estimate σ robustly by applying the Q_n method (Rousseeuw & Croux, 1993) to the residuals a_t. Q_n is based on an order statistic of all pairwise differences,

$$Q_n(a_1, \ldots, a_n) = c_n \cdot \{|a_i - a_j| : 1 \leq i < j \leq n\}_{(k)} \text{ with } k = \binom{\lfloor n/2 \rfloor + 1}{2} .$$

Here, c_n is a finite sample correction to achieve unbiasedness in a Gaussian sample of size n. Q_n possesses a large asymptotic Gaussian efficiency of 82% and can be computed in $O(n \log n)$ time. If $\phi = 0$ and the median of $z_1, \ldots, z_{\tau-1}$ is used as estimate of μ, the resulting detection rule corresponds to the median comparisons (MC) suggested in Fried (2007a).

In case of $\phi \neq 0$ we additionally need to estimate ϕ, which is also the lag-one autocorrelation $\rho(1)$. The highly robust SSD estimates (Ma & Genton, 2000) of the autocovariance and autocorrelation are based on writing the lag-one autocovariance as

$$\gamma(1) = \frac{1}{4}\left[Var(Y_t + Y_{t-1}) - Var(Y_t - Y_{t-1})\right] .$$

These variances are estimated by applying Q_n to $z_2+z_1, \ldots, z_n+z_{n-1}$ and $z_2-z_1, \ldots, z_n-z_{n-1}$, respectively. Since scaling $\hat{\gamma}_{SSD}(1)$ by the sum of the variances in (2.2) is superior to scaling by $\hat{\gamma}_{SSD}(0)$ (see Ma and Genton (2000), Fried (2007a)), 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 numerator and the denominator remain bounded and bounded away from zero.

If both τ and $n - \tau$ are large enough, the distribution of $\tilde{T}_\tau$ is still approximately standard normal under H_0 since we use consistent estimators. Otherwise, we can again derive critical values for $\tilde{T}_\tau$ from simulations.

3 Experiments

We perform some Monte Carlo experiments to investigate the performance of the detection rules in several situations. We consider time windows of different widths $n = 21$ and $n = 33$ and test for a level shift at time $\tau = 15$ and $\tau = 23$, respectively. For estimation of μ we use the observations up to time $\tau - 1$. In the LS test we use the standard deviation and the CLS estimate for σ and ϕ, both pooled from $z_1, \ldots, z_{\tau-1}$ and $z_\tau, \ldots, z_n$. In the MC test we simply apply the Q_n to the full window $z_1, \ldots, z_n$ for estimation of σ and ϕ as described in Section 2.2. Additionally we comment on the versions which use only the data $z_1, \ldots, z_{\tau-1}$ up to the time point τ at which we want to estimate a shift. The estimates are plugged into the formulae for the residuals (a_t) and the test statistics. Besides the tests based on T_τ from (5) and $\tilde{T}_\tau$ from (6) we also consider the corresponding test statistics for the original observations, which assume $\phi = 0$, and adjust their critical values for autocorrelations in the following.

First we generate 100001 time windows from each of different AR(1) models with $\phi \in \{-0.95, -0.9, \ldots, 0.95\}$ and Gaussian errors. The empirical 99.9% percentiles of the absolute test statistics will be used as critical values 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 caused by testing all time points of a long time series. Figure 1 shows that the percentiles of the test statistics increase with increasingly positive autocorrelations ϕ, particularly in case of the MC test based on the AR residuals (a_t), implying that a much larger sample size is needed for the asymptotics to become useful. We only report the results for $n = 21$ since those for $n = 33$ are qualitatively the same.

Now we check whether the tests keep their significance levels when using the simulated critical values corresponding to the estimates of ϕ after rounding. We generate 20000 Gaussian AR(1) windows without shift for each of different values of $\phi \in \{-0.9, -0.8, \ldots, 0.9\}$. The percentage false detections is calculated for measuring the sizes of the tests as a function of ϕ, see also Figure 1. Obviously, the tests become increasingly liberal with increasing value of ϕ. This can be explained by using an estimate of ϕ instead of its true value. The tests based on the original observations suffer more from this effect than those based on the AR residuals. If we use the maximum of the 99.9% percentiles derived before as critical values, the test size at least stays small until ϕ gets very close to 1, meaning that we get an approximately valid test procedure when using these thresholds for shift detection.

Next we obtain the power of the tests as a function of the shift size $\omega = 0.25, 0.5, \ldots, 10$. Figure 1 depicts the percentage cases in which a shift was detected within 2000 separate windows for each ω and $\phi = 0.5$. The least squares methods of course offer higher power in the case of Gaussian noise considered here. Generally, the most powerful tests are those which have the biggest difficulties in preserving their size. Particularly, the tests using the original observations and critical values adjusted for ϕ have more power than those based on AR residuals. The tests based on estimated percentiles are of course more powerful than those using the maximal percentile, but remember that the former do not preserve their level. The MC test based on the AR residuals with the maximum threshold does not have power at all, what can be explained by the steeply increasing percentile function which leads to a very large maximum threshold. As opposed to this, the LS and the MC tests based on the original data as well as the LS test based on AR residuals still provide considerable power and preserve their significance level for most values of ϕ. We note that the tests using the observations up to time $\tau - 1$ only for estimation of the model parameters turn out to be less powerful than their counterparts using the whole window.

The LS tests loose their superiority in case of a single large outlier into the same direction as the shift before time τ, or into the opposite direction after time τ, while the ordering of the robust methods remains the same, see also Figure 1. One must totally rely on additional outlier detection rules when using traditional linear methods for shift detection, while the robust MC rules still work well even when some outliers are neglected.

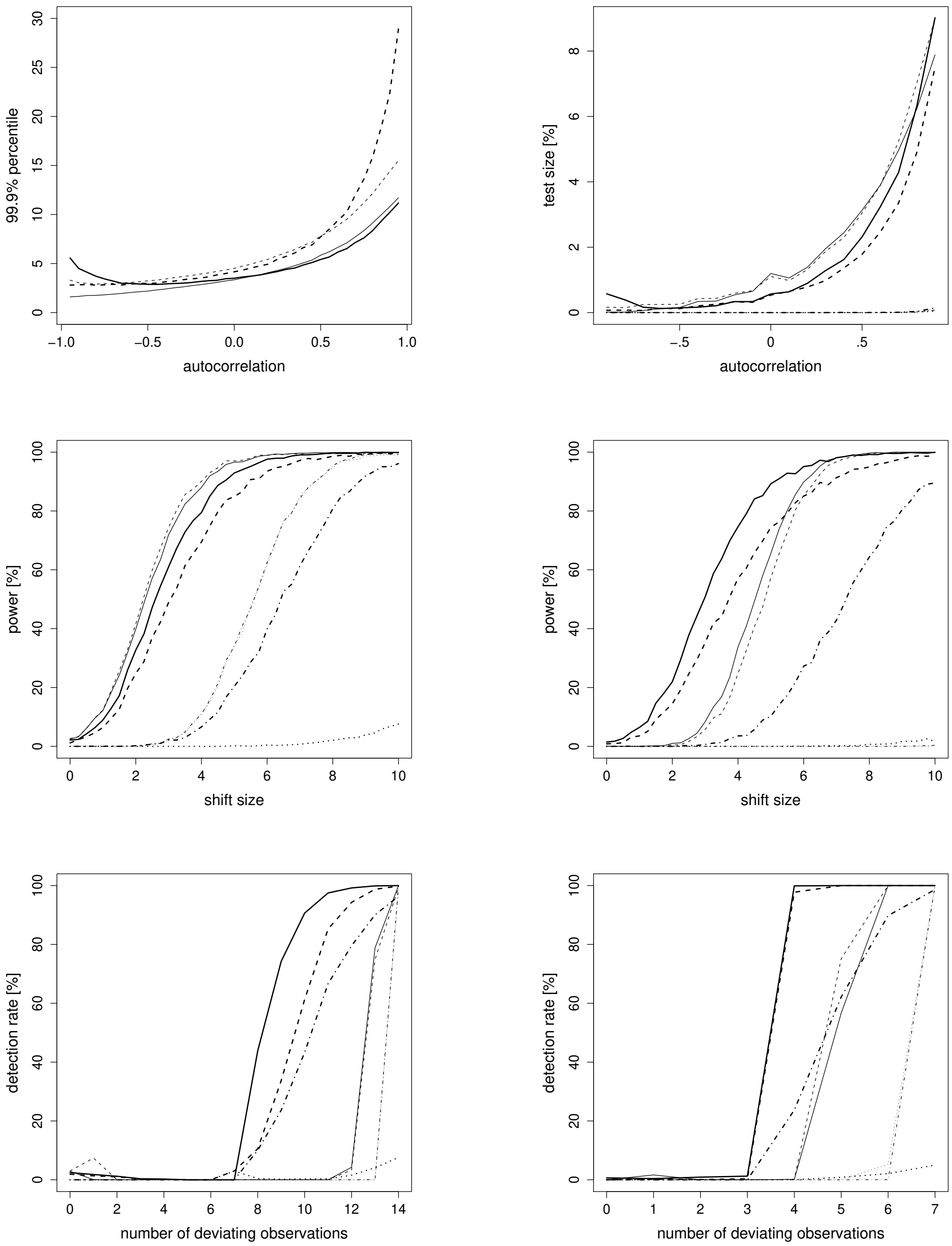

Figure 1: Results for AR(1) with $n = 21$ and $\tau = 15$: 99.9% percentiles of the absolute test statistics (top left), test sizes for different values of ϕ (top right); $\phi = 0.5$: power (center left) and power in case of an outlier of size 10 (center right), percentage detected shifts in case of an increasing number of observations shifted by 10 in the left and the right window (bottom). Estimated thresholds: LS test (thin solid), LS test on AR residuals (thin dashed), MC test (bold solid) and MC test on AR residuals (bold dashed). Maximum thresholds: LS test (thin dot-dashed), LS test on AR residuals (thin dotted), MC test (bold dot-dashed) and MC test on AR residuals (bold dotted).

For a closer investigation of the test resistances against outliers we simulate the intrusion of a level shift into the window. For this we generate windows from an AR(1) model with $\phi = 0.5$ and shift an increasing number $j = 0, 1, \ldots, 7$ of observations at times $t = 21, 20, \ldots, 15$ by $\omega = 10$. For optimal resistance we want a shift not to be detected until more than half of the data at times $\tau = 15, \ldots, n = 21$ strongly deviates from the data at $t = 1, \ldots, 14$. Figure 1 shows that the MC tests with estimated thresholds perform very well in this respect, while the MC test based on AR residuals with maximal threshold very rarely detects such a shift at all. The MC test based on the original observations with maximum thresholds performs reasonable and close to the LS tests with estimated thresholds in case of a shift of size 10, and better than these in case of a larger shift (not shown here). The LS tests with maximum thresholds do not detect a shift unless all observations are shifted. This improves if only $z_1, \ldots, z_{\tau-1}$ are used for the estimates, but neither the LS tests with maximal threshold nor the median comparison based on the original observations perform better than e.g. the MC test based on the original observations with maximum thresholds. Two large outliers might cause the LS tests with estimated thresholds to indicate a shift incorrectly with high probability then.

In a similar manner we generate windows from an AR(1)-model with $\phi = 0.5$ and an increasing number of observations at $t = 1, \ldots, \tau-1 = 14$ shifted by 10. For seven or less observations this points at a couple of outliers in a steady state, while in case of more than seven deviating observations we might regard a shift with some outliers before the shift as more likely. Figure 1 illustrates that the MC test based on the original observations with maximal threshold again performs reasonably well then, like the MC tests with estimated thresholds. A few outliers before a shift can easily mask the shift for all LS tests. The results do of course get worse in this situation if we only include the data before the shift into the estimates.

To investigate the effects of model miss-specification we also generate time series from a moving average model with parameter $\theta \in \{0.5, 1\}$, obtaining quite similar or even slightly better results in terms of power than reported above. The power of course also increases when increasing the width to $n = 33$ and testing for a shift at $\tau = 23$, with the ordering of the methods being the same. See Fried (2007c) for more results.

4 Conclusions

Median comparisons allow reliable detection of large shifts in autocorrelated time series with a locally constant level, time-varying parameters and outliers. To avoid false alarms we can compare robustly standardized medians of the original observations to sufficiently large thresholds, obtained by maximization with respect to the autocorrelations. The re-

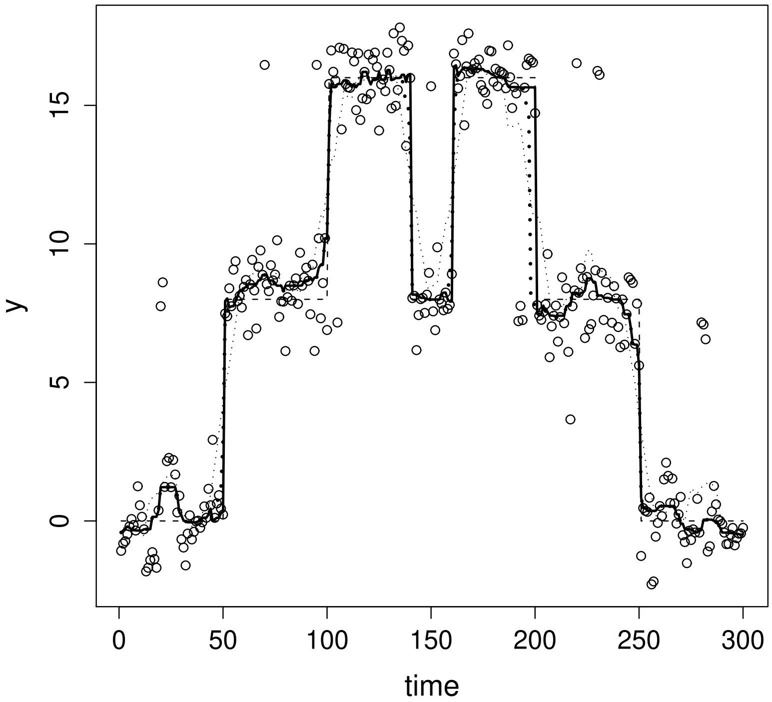

Figure 2: Step function (dashed) overlaid by time-varying AR noise (dots) and level estimates: running mean (dotted), running median (bold dotted) both with window width 15 and the same running median improved by MC test and additional rules (bold solid).

sulting tests perform reliably in a broad range of situations. For shift detection within trends we can replace the median by the repeated median (see Fried & Gather, 2005). When using Q_n we can include the whole time window in the estimation of the AR parameters and the local variability since it copes well with shifts due to being based on pairwise differences between the observations. Using an estimated threshold for shift detection, derived from robust estimates of the autocorrelations, leads to more powerful, but quite oversized tests in case of moderate to strong positive autocorrelations. This problem becomes less severe but still noteworthy when using model residuals for testing. Possibly this can be further improved under stronger assumptions than considered here. Particularly, better estimates derived from longer estimation periods might lead to large improvements, but afford the model parameters to be almost constant during longer time spans, or another more elaborated model for the dynamical dependence structure.

Figure 2 depicts a step function which is overlaid by AR(1) noise with time-varying parameters $\phi_t = 0.9 - 0.9 * sin(\pi t/300)$ and $\sigma_t^2 = 1/(1 - \phi_t^2)$ for illustration. 15 observations have been replaced by additive outliers of size 8. While the LS tests with estimated thresholds incorrectly detect a shift when there are several close-by outliers, the LS tests with maximum thresholds only detect the shifts at $t = 50$ and at $t = 140$ because of the outliers. As opposed to this, the MC test with maximum thresholds correctly detects the shifts at $t = 50, 140, 160$ and 200. This allows to improve a running median with window width 15, which we have used for comparison, at the detected shifts by using one-sided medians there. We note that the MC test with estimated thresholds even detects all shifts correctly without any false alarms in this case.

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

References

Fried, R. (2007a). On robust shift detection in time series. *Computational Statistics and Data Analysis*, *52*, 1063-1047.

Fried, R. (2007b). Robust location estimation under dependence. *Journal of Statistical Computation and Simulation*, *77*, 131-147.

Fried, R. (2007c). Robust shift detection in autoregressive models. In Y. K. S. Aivazian P. Filzmoser (Ed.), *Proceedings of the 8th international conference computer data analysis and modeling* (p. 60-67). Minsk: Publ. Center BSU.

Fried, R., & Gather, U. (2005). Robust trend estimation for ar(1) disturbances. *Austrian Journal of Statistics*, *34*, 139-151.

Fried, R., & Gather, U. (2007). On rank tests for shift detection in time series. *Computational Statistics and Data Analysis*, *52*, 221-233.

Ma, Y., & Genton, M. (2000). Highly robust estimation of the autocovariance function. *Journal of Time Series Analysis*, *21*, 663-684.

Peña, D. (2000). Intervention analysis. In D. Peña, G. Tiao, & R. Tsay (Eds.), *A course in time series analysis* (p. 415-421). New York: Wiley.

Rousseeuw, P., & Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association*, *88*, 1273-1283.

Tukey, J. (1977). *Exploratory data analysis*. Reading, Mass.: Addison-Wesley.

Author: Roland Fried, Faculty of Statistics, Dortmund University of Technology, 44221 Dortmund, Germany; tel. +49 (0)231 755 3119; mail: fried@statistics.uni-dortmund.de