Robust Trend Estimation for AR(1) Disturbances

Roland Fried and Ursula Gather
University Carlos III, Madrid, and University of Dortmund

Abstract: We discuss the robust estimation of a linear trend if the noise follows an autoregressive process of first order. We find the ordinary repeated median to perform well except for negative correlations. In this case it can be improved by a Prais-Winsten transformation using a robust autocorrelation estimator.

Zusammenfassung: Wir behandeln die robuste Schätzung eines linearen Trends bei autoregressiven Fehlern erster Ordnung. Die Repeated Median Regression zeigt ein gutes Verhalten bei positiven Korrelationen. Bei negativen Korrelationen ist eine Verbesserung durch eine Prais-Winsten Transformation mittels eines robusten Korrelationsschätzers möglich.

Keywords: Robust Regression, Autocorrelations, Detrending, Cochrane-Orcutt Estimator, Prais-Winsten Estimator.

1 Introduction

We discuss the robust estimation of a linear trend from a noisy time series $Y_1, \ldots, Y_n$ of moderate size n in the presence of autoregressive disturbances of first order, AR(1), and irrelevant measurement artifacts (outliers). The model reads

$$Y_t = \mu + \beta(t - m - 1) + \epsilon_t, \quad t = 1, \ldots, n = 2m + 1, \tag{1}$$

$$\epsilon_t = \phi\epsilon_{t-1} + \delta_t, \tag{2}$$

where δ_t are innovations from a white noise process with zero mean and variance $\sigma^2 > 0$. We center time to interpret μ as the central level and assume stationary errors, $|\phi| < 1$.

The estimation of linear trends in the presence of AR(1) errors has received considerable attention in econometrics. A number of papers compares the efficiencies of the ordinary least squares (OLS), the generalized least squares (GLS), the first differences (FD), the Cochrane-Orcutt (CO) and the Prais-Winsten (PW) estimators among others for estimation of the slope β, often under the idealized assumption that ϕ is known (see Krämer, 1980, 1982, Steman and Trenkler, 2000, and the references cited therein). However, methods based on least squares are highly vulnerable to contamination by outliers. This makes simple robust alternatives interesting.

Robust fitting of linear trends to data within a moving time window of moderate length n has been investigated recently by Davies, Fried and Gather (2004). In this context, the central level μ is of primary interest. Based on a comparison of robust regression techniques, they find Siegel's (1982) repeated median (RM) to be very suitable for the estimation of trends because of its robustness, stability and computational tractability. However, application of the ordinary repeated median means treating the data as independent, although autocorrelations can cause monotonic data patterns similar to time-varying trends.

We investigate here whether the repeated median can be improved in the presence of AR(1) noise. One possibility is simultaneous estimation of the autoregressive and the trend parameters by robust regression. Robust regression techniques for fitting AR models to data with a constant level have been suggested by Rousseeuw and Leroy (1987) and Meintanis and Donatos (1999). Another approach is preliminary estimation of ϕ, followed by trend estimation from transformed data.

We compare here repeated median approaches for robust estimation of linear trends. Section 2 reviews least squares techniques and transfers them to repeated median regression. Section 3 presents a comparison for small samples. Section 4 gives some concluding remarks.

2 Methods for linear trend estimation

We first review transformations for including ϕ in the estimation of μ and β. Setting $\mathbf{Y} = (Y_1, \ldots, Y_n)'$ and $\boldsymbol{\varepsilon} = (\epsilon_1, \ldots, \epsilon_n)'$, the model equations (1) and (2) can be expressed in vector form as

$$\begin{aligned} \mathbf{Y} &= \mu\mathbf{1} + \beta\mathbf{J} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim N\left(\mathbf{0}, \frac{\sigma^2}{1-\phi^2}\boldsymbol{\Sigma}\right) \\ \boldsymbol{\Sigma} &= \begin{pmatrix} 1 & \phi & \ldots & \phi^{2m} \\ \phi & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \phi \\ \phi^{2m} & \ldots & \phi & 1 \end{pmatrix}, \end{aligned}$$

where $\mathbf{1} = (1, \ldots, 1)'$ is a vector of ones and $\mathbf{J} = (-m, \ldots, m)$ are the centered time points. The inverse correlation matrix is $\boldsymbol{\Sigma}^{-1} = \boldsymbol{P}'\boldsymbol{P}/(1-\phi^2)$,

$$\boldsymbol{P} = \begin{pmatrix} \sqrt{1-\phi^2} & 0 & \ldots & \ldots & 0 \\ -\phi & 1 & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & -\phi & 1 & 0 \\ 0 & \ldots & 0 & -\phi & 1 \end{pmatrix}.$$

The PW estimator of μ and β corresponds to OLS applied to the whitened data

$$\boldsymbol{P}\mathbf{Y} = \mu\boldsymbol{P}\mathbf{1} + \beta\boldsymbol{P}\mathbf{J} + \boldsymbol{P}\boldsymbol{\varepsilon}\,. \tag{3}$$

The CO estimator does not make use of the first line of $\boldsymbol{P}$, i.e. it uses only $Y_2 - \phi Y_1, \ldots, Y_n - \phi Y_{n-1}$, conditioning on Y_1. The difference between the two estimators can be supposed to be negligible for large n, but it can be relevant for small n. If we do not know ϕ we first estimate it and use a two-stage procedure with ϕ being substituted by the preliminary estimate in a second step.

Simultaneous estimation of μ, β and ϕ is possible by substituting ϵ_t in equation (1) by its expression (2), and ϵ_{t-1} in turn by (1). This results in

$$Y_t = \mu(1-\phi) + \beta[1 + (t-m-2)(1-\phi)] + \phi Y_{t-1} + \delta_t\,, \quad t = 2, \ldots, n.$$

Setting $\nu = \mu(1-\phi) + \beta$ and $\alpha = \beta(1-\phi)$ we get a linear regression model with autoregressive term conditioning on Y_1

$$Y_t = \nu + \alpha(t-m-2) + \phi Y_{t-1} + \delta_t\,, \quad t = 2, \ldots, n. \tag{4}$$

The original parameters are $\beta = \alpha/(1-\phi)$ and $\mu = (\nu - \beta)/(1-\phi)$, i.e. there is a one-to-one correspondence if we restrict ϕ to $(-1, 1)$.

The FD estimator is based on the common idea to remove non-stationarities by taking differences. For the model described in (1) and (2), the first differences $Z_t = Y_t - Y_{t-1}$ follow a non invertible ARMA(1,1) process

$$Z_t = \alpha + \phi Z_{t-1} + \delta_t - \delta_{t-1}\,, \quad t = 2, \ldots, n.$$

Estimation of α and ϕ using $Z_2, \ldots, Z_n$ allows to clean the data from the trend and the autocorrelations. The mean of the differences is rather efficient for estimating α if ϕ is large and n is small (Krämer, 1982). However, the robust analogue, the median of the first differences, can be improved substantially by the repeated median applied to $Y_1, \ldots, Y_n$ (Fried, 2004). We therefore neglect estimators based on first differences.

2.1 Regression estimators of ϕ

The multiple linear regression model (4) can be fitted to data $y_2, \ldots, y_n$ for simultaneous estimation of μ, β and ϕ, conditioning on y_1. Conditional least squares (CLS) minimizes

$$\sum_{t=2}^{n} (y_t - \nu - \alpha(t-m-2) - \phi y_{t-1})^2\,.$$

Solving the normal equations we get simultaneous CLS estimates (SCLS)

$$\begin{aligned}
\hat{\nu}_{SCLS} &= \overline{y}_n - \hat{\phi}_{SCLS} \cdot \overline{y}_1 + \hat{\alpha}_{SCLS}/2 \\
\hat{\alpha}_{SCLS} &= \frac{\displaystyle\sum_{t=2}^{n} (t-m-2)(y_t - \overline{y}_n - \hat{\phi}_{SCLS}(y_{t-1} - \overline{y}_1))}{\displaystyle\sum_{t=2}^{n} (t-m-2)(t-m-1.5)},
\end{aligned}$$

where $\overline{y}_n$ and $\overline{y}_1$ are the arithmetic mean of $y_2, \ldots, y_n$, and $y_1, \ldots, y_{n-1}$, respectively. The estimate of ϕ reads

$$\hat{\phi}_{SCLS} = \frac{\sum_{j=2}^{n} y_{j-1} \left(y_j - \overline{y}_n - \frac{\sum_{t=2}^{n}(t-m-2)(y_t - \overline{y}_n)}{\sum_{t=2}^{n}(t-m-2)(t-m-1.5)}(j-m-1.5) \right)}{\sum_{j=2}^{n} y_{j-1} \left(y_{j-1} - \overline{y}_1 - \frac{\sum_{t=2}^{n}(t-m-2)(y_{t-1} - \overline{y}_1)}{\sum_{t=2}^{n}(t-m-2)(t-m-1.5)}(j-m-1.5) \right)},$$

where the second terms in the numerator and the denominator are the separately OLS detrended observations $y_2, \ldots, y_n$ and $y_1, \ldots, y_{n-1}$, respectively. This trend correction would be missing if we restricted α to be zero. We note that the resulting estimates $\hat{\mu}_{SCLS}$ and $\hat{\beta}_{SCLS}$ correspond to the CO approach based on $\hat{\phi}_{SCLS}$. This suggests a construction of PW type estimators $\hat{\mu}_{PW-SCLS}$ and $\hat{\beta}_{PW-SCLS}$ using $\hat{\phi}_{SCLS}$.

A related approach is the joint initial detrending of $y_1, \ldots, y_n$ by OLS, followed by the estimation of ϕ from the residuals and the construction of OLS estimates from the transformed data (3). We denote the resulting estimators by CO-DLS and PW-DLS.

To derive formulae for simultaneous RM estimators of ν, α and ϕ we note that each triple (i, y_{i-1}, y_i), (j, y_{j-1}, y_j), (k, y_{k-1}, y_k), $i \neq j \neq k \neq i$, corresponds to a unique solution $(\nu_{ijk}, \alpha_{ijk}, \phi_{ijk})$ of (4) with

$$\phi_{ijk} = \frac{y_k - y_i - \frac{k-i}{j-i}(y_j - y_i)}{y_{k-1} - y_{i-1} - \frac{k-i}{j-i}(y_{j-1} - y_{i-1})}$$

whenever the data are in general position. General position here means that no triple of observations lies on a straight line. This condition avoids zero denominators, and is almost surely fulfilled if the data come from a continuous distribution. We can drop all triples for which it is not fulfilled from the calculations otherwise. Again, the second terms in the numerator and the denominator mean a trend correction and would be missing if the slope was set to zero. Siegel's (1982) repeated median in our situation now reads

$$\hat{\phi}_{SRM} = med_{i=2,\ldots,n} med_{j \neq i} med_{k \neq i,j} \frac{y_k - y_i - \frac{k-i}{j-i}(y_j - y_i)}{y_{k-1} - y_{i-1} - \frac{k-i}{j-i}(y_{j-1} - y_{i-1})} \tag{5}$$

$$\hat{\alpha}_{SRM} = med_{i=2,\ldots,n} med_{j \neq i} \frac{y_i - y_j - \hat{\phi}_{SRM}(y_{i-1} - y_{j-1})}{i-j}$$

$$\hat{\nu}_{SRM} = med_{i=2,\ldots,n} \left(y_i - \hat{\alpha}_{SRM}(i-m-2) - \hat{\phi}_{SRM} \cdot y_{i-1} \right) .$$

A drawback of the simultaneous RM is the $O(n^3)$ computation time needed by a straightforward implementation even if a routine for calculation of the median in linear time is used.

Since SRM in fact estimates $\hat{\phi}$ using a trend correction first, it is near at hand to detrend all observations applying the ordinary RM, and then to estimate ϕ from the residuals r_i, $i = 1, \ldots, n$, using

$$\hat{\phi}_{DRM} = med_{i=2,\ldots,n} med_{j \neq i} \frac{r_i - r_j}{r_{i-1} - r_{j-1}} \, .$$

Thereafter we can insert $\hat{\phi}_{DRM}$ into the equations (5) instead of $\hat{\phi}_{SRM}$ to calculate estimates $\hat{\mu}_{DRM}$ and $\hat{\beta}_{DRM}$. This reduces the computation time to $O(n^2)$ or less when using the algorithm by Matousek, Mount and Netanyahu (1998).

We note that estimators based on (5) are of CO type. We can construct PW analogues including y_1 in the estimation of μ and β setting $(i^\star, y_i^\star)$ to $(i - \hat{\phi}(i-1), y_i - \hat{\phi} y_{i-1})$ for $i \in \{2, \ldots, n\}$, and to $\sqrt{1 - \hat{\phi}^2}(1, y_1)$ for $i = 1$:

$$\begin{aligned} \hat{\beta}_{PW} &= med_{i=1,\ldots,n} med_{j \neq i} \frac{y_i^\star - y_j^\star}{i^\star - j^\star} \\ \hat{\mu}_{PW} &= med_{i=1,\ldots,n} (y_i^\star - \hat{\beta}_{PW} \cdot i^\star) \, , \end{aligned} \tag{6}$$

where $\hat{\phi}$ can be either chosen as $\hat{\phi}_{SRM}$ or $\hat{\phi}_{DRM}$.

A difficulty of regression methods is that ϕ may be estimated to be larger than one, corresponding to an explosive behavior. This causes further problems in the subsequent estimation of μ and β. The estimate of ϕ needs thus to be restricted artificially by an upper and lower bound to prevent non-stationary estimates.

2.2 Correlation estimators of ϕ

A possible remedy to overcome non-stationary estimates of ϕ is to use the fact that $\phi = \gamma(1)/\gamma(0)$ is the lag-one autocorrelation coefficient in the AR(1) model, where $\gamma(h)$ is the lag-h autocovariance. Correlation estimates are usually guaranteed to be at most one in absolute value. We may then correct the observations for the dependencies and estimate μ and β from the transformed data using the CO or the PW approach.

The traditional Yule-Walker (YW) estimator of ϕ is the lag-one sample autocorrelation based on a constant level. The difference to the CLS estimate is that the overall mean $\overline{y}$ is generally used for centering and that the variability in the denominator is estimated from all observations. This causes an increased bias toward zero in small samples. To cope with the trend in the data we can again detrend the data by OLS before calculating a Yule-Walker estimate from the residuals $r_1, \ldots, r_n$.

$$\hat{\phi}_{DYW} = \frac{\sum_{t=2}^{n} r_{t-1} \cdot r_t}{\sum_{t=1}^{n} r_t^2} \, .$$

Instead of $\hat{\phi}_{DYW}$ we can use the ordinary sample correlation for $(r_1, r_2), \ldots, (r_{n-1}, r_n)$

$$\hat{\phi}_{DSC} = \frac{\sum_{t=2}^{n}(r_{t-1} - \overline{r}_1)(r_t - \overline{r}_n)}{\sqrt{\sum_{t=2}^{n}(r_{t-1} - \overline{r}_1)^2 \sum_{t=2}^{n}(r_t - \overline{r}_n)^2}}.$$

This estimate neglects the fact that σ is the same for $r_1, \ldots, r_{n-1}$ and $r_2, \ldots, r_n$. It standardizes by two estimates of the variability based on the first and the last $n-1$ residuals and also guarantees $|\hat{\phi}_{DSC}| \leq 1$ because of the Cauchy-Schwarz inequality.

A highly robust estimation method for the autocovariances and autocorrelations has been proposed by Ma and Genton (2000). Their SSD method uses

$$\gamma(h) = \frac{1}{4}\left[Var(Y_t + Y_{t-h}) - Var(Y_t - Y_{t-h})\right] \tag{7}$$

and estimates these variances by Q_n of Rousseeuw and Croux (1993). The Q_n-scale estimate is based on an order statistic of all pairwise differences,

$$Q_n(x_1, \ldots, x_n) = c_n \cdot \{|x_i - x_j| : 1 \leq i < j \leq n\}_{(\ell)}\,,\ \ell = \binom{\lfloor n/2 \rfloor + 1}{2}.$$

Q_n has a high asymptotic efficiency of 82% and can be computed in $O(n \log n)$ time. The finite sample correction c_n cancels out in the estimation of ϕ. We found it superior to scale $\hat{\gamma}_{SSD}(1)$ using the sum of the variances in (7) instead of $\hat{\gamma}_{SSD}(0)$. This was also suggested by Ma and Genton (2000). For robust detrending we apply the ordinary RM and estimate ϕ from the residuals. The resulting estimate

$$\hat{\phi}_{DSSD} = \frac{Q^2_{n-1}(r_2 + r_1, \ldots, r_n + r_{n-1}) - Q^2_{n-1}(r_2 - r_1, \ldots, r_n - r_{n-1})}{Q^2_{n-1}(r_2 + r_1, \ldots, r_n + r_{n-1}) + Q^2_{n-1}(r_2 - r_1, \ldots, r_n - r_{n-1})}$$

is guaranteed to lie within $[-1, 1]$. If less than 25% of the observations in general position are replaced by outliers, the terms in the numerator and the denominator will neither become zero nor infinity, guaranteeing an asymptotic breakdown point of 25%. Using $\hat{\phi}_{DSSD}$ we can obtain CO and PW type estimates of μ and β via (5) and (6), respectively.

3 Simulations

In the following we perform Monte Carlo experiments to compare the methods introduced above for a small sample size n. It is straightforward to check that the estimators are well-behaved w.r.t. changes of location, scale, or a constant trend, i.e. they possess desirable equivariance and invariance properties due to the initial detrending in the estimation of ϕ. Hence we restrict the analysis w.l.o.g. to $\mu = \beta = 0$ and $\sigma^2 = 1$. To prevent problems because of non-stationary estimates we set the estimate of ϕ to ± 0.99 whenever these bounds are exceeded. We concentrate on the PW type estimators since we found them to

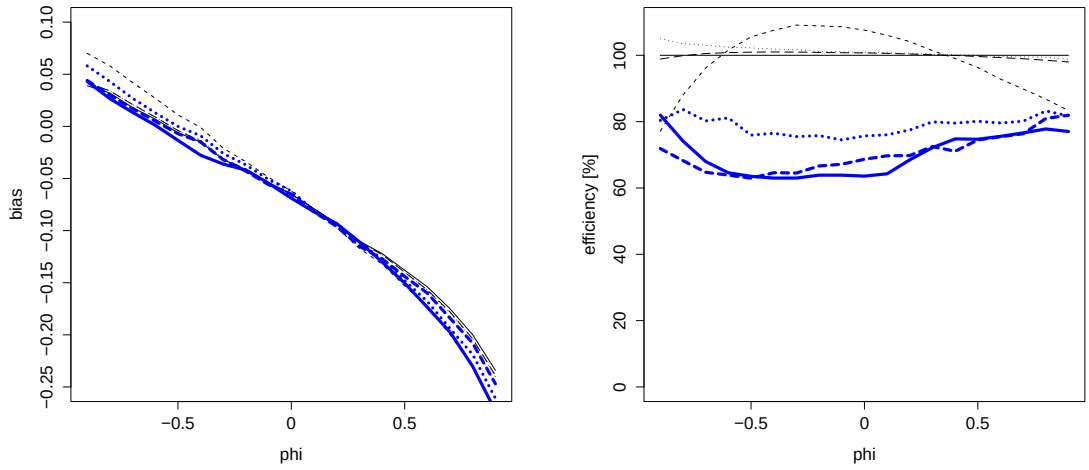

Figure 1: Bias (left) and finite sample efficiency (right) for ϕ relative to SCLS (solid): DCLS (wide-dashed), DYW (narrow–dashed), DSC (dotted), SRM (bold solid), DRM (bold dashed), DSSD (bold dotted).

be superior to the CO versions, in particular for μ in case of large positive ϕ. This verifies the postulation of Beach and MacKinnon (1978) also for the methods based on the RM. The least squares (LS) methods are included just for comparison.

3.1 Gaussian innovations

We compare the finite sample efficiencies of the methods measured by the mean square errors for the estimators of μ, β and ϕ in the case of Gaussian innovations. We consider time series of moderate length $n = 31$ and approximate the MSEs as a function of ϕ from 10000 time series for each $\phi \in \{-0.9, -0.8, \ldots, 0.9\}$, see Figures 1 and 2.

Figure 1 compares the biases and relative efficiencies for the estimators of ϕ. All methods are biased towards zero, and the biases are increasing in $|\phi|$. DYW is the most biased, followed by the DSSD. For positive ϕ, the SRM is similarly biased as the DYW. For ϕ close to zero, DYW is the most efficient method, while SCLS and DCLS are more efficient for $|\phi| \geq 0.5$. DSC is the best for negative ϕ. Among the robust methods, DSSD is the most efficient, while SRM and DRM are close to each other.

For μ and β, all methods are unbiased because of the underlying symmetry. Therefore we only compare the efficiencies relative to the optimal GLS estimator knowing ϕ in Figure 2. The PW type LS estimators perform about equally well for negative and moderately large positive ϕ. For high positive ϕ, the DSC is the most efficient one, followed by the DCLS. OLS is more efficient than the CO type estimators then, but it is less efficient than the PW methods for large $|\phi|$. See Maeshiro (1976), Chipman (1979), Krämer (1980, 1982) or Steman and Trenkler (2000) for more information. As opposed to this, the ordinary RM is more efficient than the RM based PW type estimators for positive ϕ

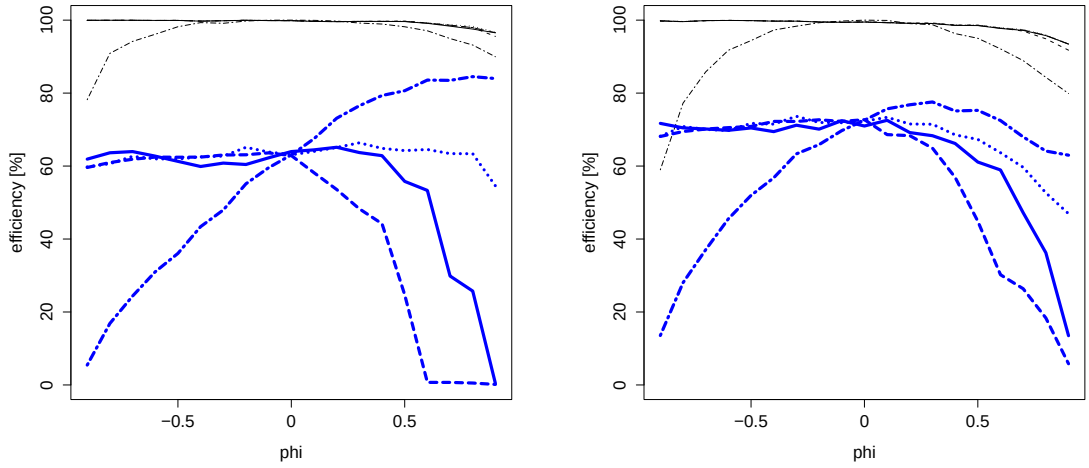

Figure 2: Finite sample efficiency for μ (left) and β (right) relative to GLS: OLS (dash-dot), SCLS (solid), DCLS (wide-dashed), DYW (narrow-dashed), DSC (dotted), SRM (bold solid), DRM (bold dashed), DSSD (bold dotted) and RM (bold dash-dot).

throughout. For negative ϕ, the adjusted RM based estimators of both PW and CO type perform about equally well, but for large positive ϕ the SRM and the DRM deteriorate. This is even more so for the CO versions, and then also happens for the SCLS and, to a smaller extent, the DCLS. These problems are due to a few estimates of ϕ which are very close to one, resulting in very large errors in the estimation of μ and β. See Park and Mitchell (1980) for a discussion of this phenomenon.

Such problems diminish with increasing sample size. Increasing n to 101 e.g., we found the CO-DRM to offer more than 60% efficiency for μ if $\phi = 0.8$. Nevertheless, the ordinary RM stays considerably more efficient than the other RM approaches in case of positive ϕ also for a larger n.

3.2 Heavy tailed innovations

We compare the performance of the methods in case of heavy tails, which may generate so called innovative outliers. We repeat the previous analysis but now generate the innovations from Student's t-distribution with 3 degrees of freedom.

Figure 3 compares the biases and the efficiencies for the estimators of ϕ. W.r.t. bias, the main difference to the case of Gaussian innovations is that now DSSD is generally the least biased. The relative performances of the LS methods are about the same as before as was to be expected, while the robust methods gain some efficiency. The DRM and the SRM do so over the whole range of ϕ, while for the DSSD this is true only for large $|\phi|$. The traditional methods perform well in this respect. This corresponds to the fact that innovative outliers can result in super-efficient estimation of ϕ, but not of μ and β.

This becomes clear when looking at the efficiencies for estimators of μ and β compared to the GLS in Figure 4. Here, the robust approaches are superior, except for very large positive ϕ. The relative performances of the linear methods are again the same as

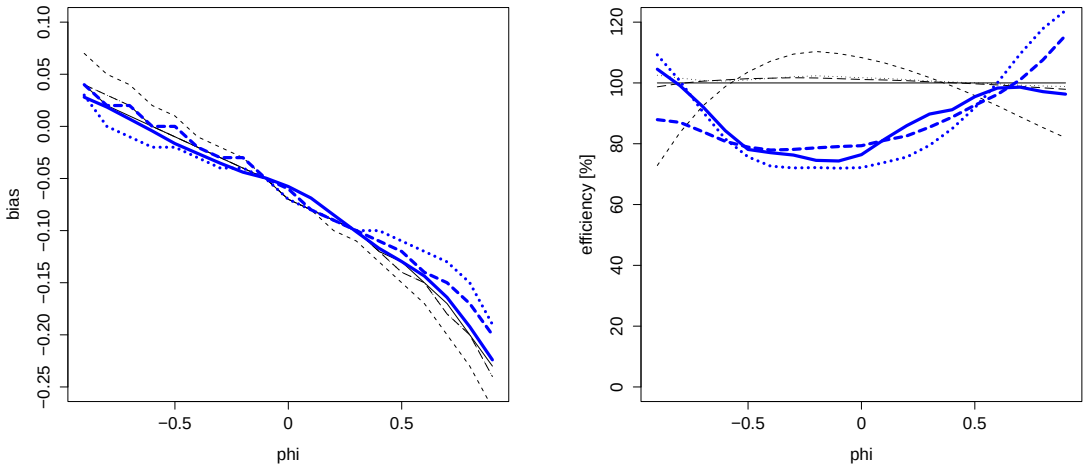

Figure 3: Bias (top) and finite sample efficiency (bottom) for ϕ relative to SCLS (solid) in case of t_3-innovations: DCLS (wide-dashed), DYW (narrow-dashed), DSC (dotted), SRM (bold solid), DRM (bold dashed), DSSD (bold dotted).

in the Gaussian case. While the OLS is inferior to the PW estimators based on LS for large $|\phi|$, the ordinary RM performs best among the robust methods in case of $\phi \geq 0$. The corrections for autocorrelations improve its performance in case of $\phi < 0$, but these methods deteriorate in case of high positive ϕ. The DSSD is somewhat better in this respect. Again, these problems are larger for the CO type estimators, and they also occur for the SCLS and to a smaller extent for the DCLS then.

3.3 Patches of additive outliers and shifts

Additive outliers (AOs) and patches of subsequent AOs cause more severe difficulties in the estimation of time series parameters than innovative outliers. A related phenomenon in the extraction of a time-varying trend is the occurrence of a level shift. A level shift can be considered as a long patch of additive outliers intruding subsequently into the moving window. To analyze the effects of a long patch of additive outliers generated e.g. by a level shift, we generate 10000 time series for $\phi = 0.5$ and each number $1, 2, \ldots, 10$ of subsequent additive outliers of same size 20 at the end of the series.

Figure 5 compares the biases and the root of the mean square errors (RMSEs). In the estimation of ϕ, the bias toward zero of the traditional methods caused by a single AO becomes obvious. In case of two or more subsequent AOs these methods become strongly positively biased, in particular the regression based estimators. The robust methods resist a few outliers well. However, they also slowly deteriorate if the number of outliers increases. In case of eight outliers they typically estimate ϕ to be close to one, which resembles the breakdown point stated above. DSSD is the most stable method, but for two or more AOs DYW mostly shows a smaller RMSE here due to its smaller variability.

W.r.t. μ and β, the classical methods show a strongly increasing bias and RMSE for an increasing number of outliers. The modified LS methods perform even worse than

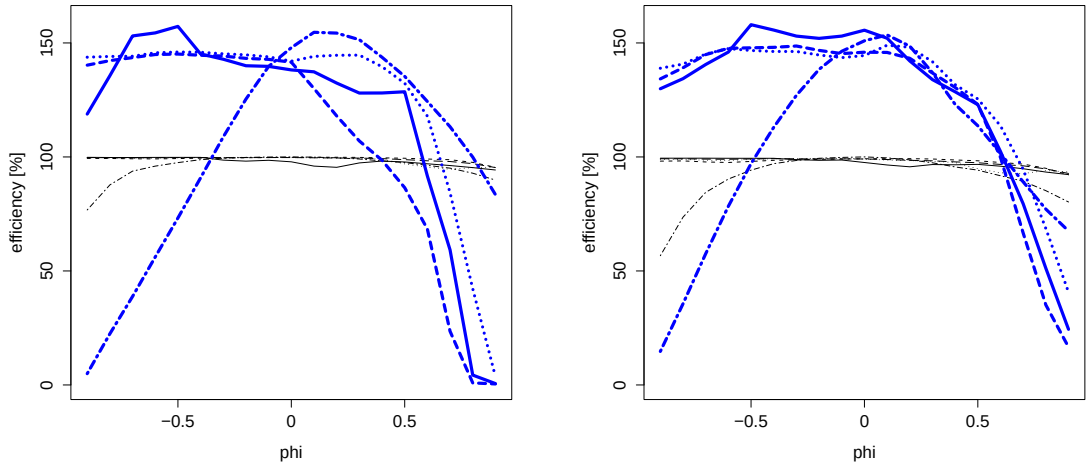

Figure 4: Finite sample efficiency relative to GLS with known ϕ in case of t_3-innovations for μ (left) and β (right): OLS (dash-dot), SCLS (solid), DCLS (wide-dashed), DYW (narrow-dashed), DSC (dotted), SRM (bold solid), DRM (bold dashed), DSSD (bold dotted), RM (bold dash-dot).

OLS. For the RM methods, the PW approach seems superior, as was to be expected. AO patches increase the risk of estimates close to non-stationarity, in particular for the regression methods. The SRM and even more the DRM deteriorate a lot for the level if there are at least five or two subsequent AOs, respectively. The DSSD performs better but it is clearly outperformed by the ordinary RM. This is in agreement with the increased sensitivity of robust functionals in case of the use of lagged versions of the time series as regressors noted by Rousseeuw and Leroy (1987) and Meintanis and Donatos (1999).

In the same analysis for $\phi = -0.5$, we found the PW transformed RM methods using a robust estimator of ϕ to be slightly superior to the ordinary RM w.r.t. both bias and variability up to about seven AOs, with the SRM being best. For eight or more subsequent AOs, however, the ordinary RM is again best, according to its optimal breakdown point.

4 Conclusions

While ordinary least squares regression can be improved considerably by a Prais-Winsten transformation if the errors have strong AR(1) autocorrelations, for the repeated median this is apparently true only for negative correlations. For positive AR-parameter it seems better not to adjust for the autocorrelations.

Regression based approaches for estimation of ϕ have difficulties with positive correlations in samples of moderate size since the estimate may become almost equal to or even larger than one. Using the Cochrane-Orcutt transformation increases further the subsequent problems in the estimation of the trend parameters due to the loss of the information about the first observation. A correlation based method offers the advantage of stationary estimates and is easier to calculate than a simultaneous robust regression approach.

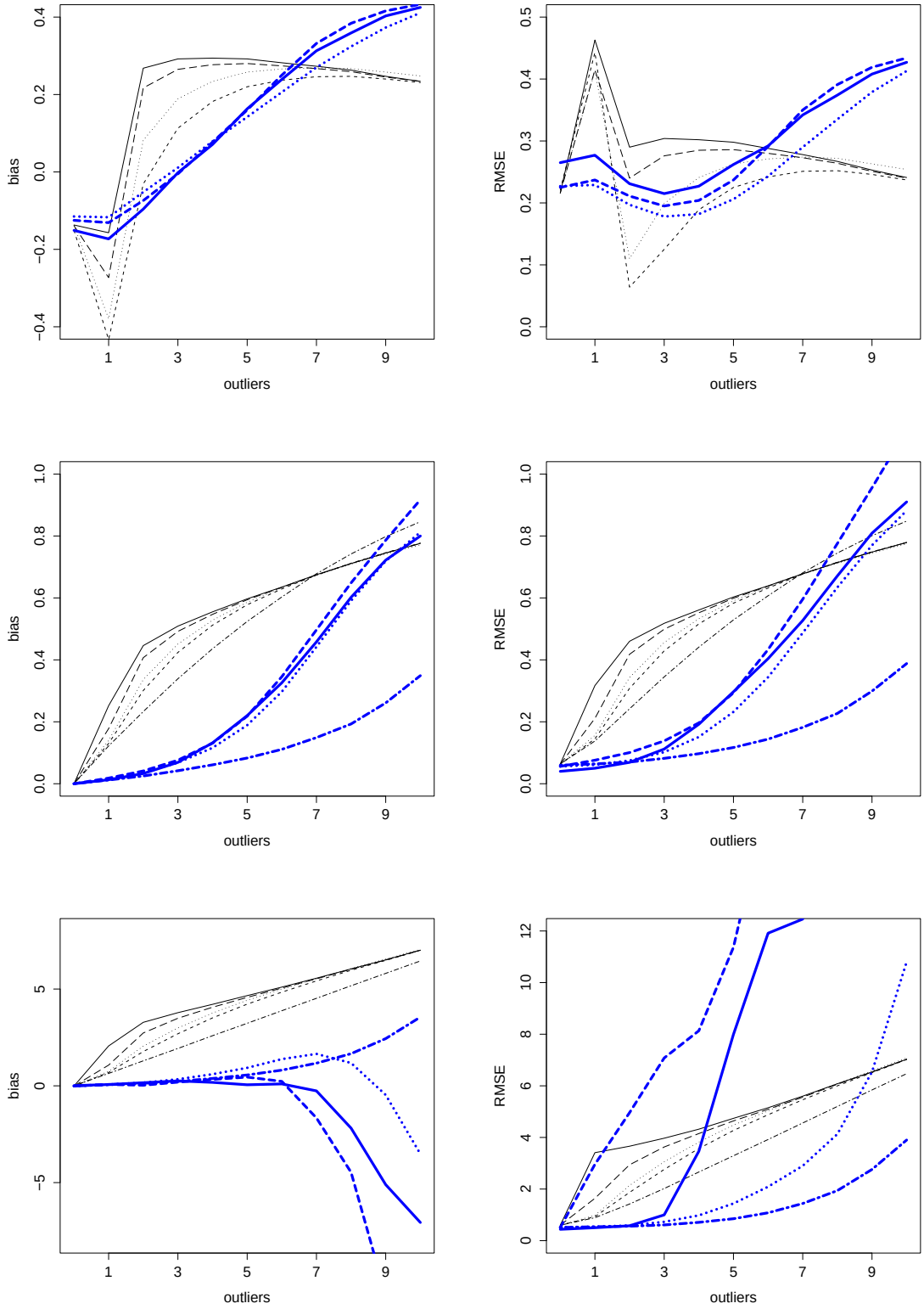

Figure 5: Bias (left) and RMSE (right) for ϕ (top), β (center) and μ (bottom) in case of an increasing number of outliers and $\phi = 0.5$: OLS (dash-dot), SCLS (solid), DCLS (wide-dashed), DYW (narrow-dashed), DSC (dotted), SRM (bold solid), DRM (bold dashed), DSSD (bold dotted), RM (bold dash-dotted).

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

References

Ch.M. Beach and J.G. MacKinnon. A maximum likelihood procedure for regression with autocorrelated errors. *Econometrica*. 46:51-58, 1978.

J.S. Chipman. Efficiency of least-squares estimation of linear trend when residuals are autocorrelated. *Econometrica*. 47:115-128, 1979.

P.L. Davies, R. Fried, and U. Gather. Robust signal extraction for on-line monitoring data. *J. Statistical Planning and Inference*. 122:65-78, 2004.

R. Fried. Robust trend extraction. In S. Aivazian, P. Filzmoser, and Y. Kharin, Y., editors, *Proceedings of the 7th International Conference Computer Data Analysis and Modeling*. Vol. I, pages 23-30. Academy of Administration at the President of the Republic of Belarus, Minsk, 2004.

W. Krämer. Finite sample efficiency of ordinary least squares in the linear regression model with autocorrelated errors. *J. Amer. Statist. Assoc.* 75:1005-1009, 1980.

W. Krämer. Note on estimating linear trend when residuals are autocorrelated. *Econometrica* 50:1065-1067, 1982.

Y. Ma and M.G. Genton. Highly robust estimation of the autocovariance function. *J. Time Series Analysis*. 21:663-684, 2000.

A. Maeshiro. Autoregressive transformations, trended independent variables and autocorrelated disturbance terms. *Review of Economics and Statistics*. 58:497-500, 1976.

J. Matousek, D.M. Mount, and N.S. Netanyahu. Efficient randomized algorithms for the repeated median line estimator. *Algorithmica*. 20:136-150, 1998.

S.G. Meintanis and G.S. Donatos. Finite-sample performance of alternative estimators for autoregressive models in the presence of outliers. *Computational Statistics & Data Analysis* 31:323-339, 1999.

R.E. Park and B.M. Mitchell. Estimating the autocorrelated error model with trended data. *J. Econometrics*. 13:185-201, 1980.

P.J. Rousseeuw and Ch. Croux. Alternatives to the Median Absolute Deviation. *J. Amer. Statist. Assoc.* 88:1273-1283, 1993.

P.J. Rousseeuw and A.M. Leroy. *Robust Regression and Outlier Detection.* Wiley, New York, 1987.

A.F. Siegel. Robust regression using repeated medians. *Biometrika*. 69:242-244, 1982.

D. Steman and G. Trenkler. Some further results on the efficiency of the Cochrane-Orcutt-estimator. *J. Statistical Planning and Inference*. 88:205-214, 2000.

Corresponding author's address:

Dr. Roland Fried
Department of Statistics
University Carlos III de Madrid
c/ Madrid 126
E-28903 Getafe
Spain

Tel. +34 91 624 5854
Fax +34 91 624 9849
Elec. Mail: rfried@est-econ.uc3m.es
http://www.est-econ.uc3m.es/