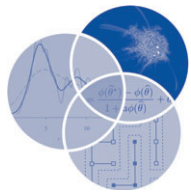
Advanced Review



Robust estimation of (partial) autocorrelation

Alexander Dürre, Roland Fried* and Tobias Liboschik

The autocorrelation function (acf) and the partial autocorrelation function (pacf) are elementary tools of linear time series analysis. The sensitivity of the conventional sample acf and pacf to outliers is well known. We review robust estimators and evaluate their performances in different data situations considering Gaussian scenarios with and without outliers as well as times series with heavy tails in a simulation study. © 2015 Wiley Periodicals, Inc.

How to cite this article:
WIREs Comput Stat 2015, 7:205–222. doi: 10.1002/wics.1351

Keywords: autocovariance; outliers; time series; correlogram

INTRODUCTION

The autocorrelation function (acf) and the partial autocorrelation function (pacf) are elementary tools of linear time series analysis. The graphical presentation as a correlogram gives an idea of the linear dependencies between pairs of observations in different time lags. A sinusoidal shape indicates a seasonality, whereas a slow decay suggests possible long range dependence or nonstationarity. Besides descriptive purposes, the autocorrelation and the autocovariance function[1] can be used for model identification (see Ref 2, p. 184–188), for fitting autoregressive models using the Yule-Walker equations, for determining the periodogram (see, e.g., Refs 3, p. 234–238, 4, p. 265–267), for detecting periodicities,[5] for clustering or classifying time series,[6] and for predicting future values of the time series.[1]

The sensitivity of the conventional estimators, the sample acf and pacf, to outliers is well known (see Refs 7, 8 or 9, p. 247–257). A single outlier can drive the sample autocorrelation at every time lag h toward zero, whereas $h+1$ or more successive outliers can make it arbitrarily close to one, both making the estimation worthless. Several robust alternatives have been proposed in the literature to overcome this problem. We review such approaches and evaluate their performances to provide some guidance on which estimator to apply in which data situation.

The first part introduces some notation and background which will be used to describe the robust procedures for estimating the acf and pacf thereafter. It is followed by a simulation study to assess the accuracy and robustness of these estimators. In the final part we draw some conclusions.

BACKGROUND AND NOTATION

Let $(X_t)_{t\in\mathbb{Z}}$ denote a real-valued time series. We assume $(X_t)_{t\in\mathbb{Z}}$ to be second order stationary, meaning that the mean and the variance are constant and do not depend on the observation time t, i.e., $\mathrm{E}(X_t)=\mu$ and $\mathrm{Var}(X_t)=\sigma^2<\infty$ for all $t\in\mathbb{Z}$, while the autocovariance and hence the autocorrelation depend on the time lag only, i.e., $\mathrm{Cov}(X_{t+h},X_t)=\gamma(h)$ and $\mathrm{Cor}(X_{t+h},X_t)=\rho(h)$ for all $t,h\in\mathbb{Z}$. Because $\rho(h)=\rho(-h)$, only positive time lags $h\in\mathbb{N}_0$ need to be considered. Both the autocovariance and the acfs of a stationary process are always positive-semidefinite, i.e., for every $k\in\mathbb{N}$ the matrix

$$\Gamma^{(k)}=\left(\Gamma^{(k)}_{i,j}\right)_{i,j=1,\ldots,k+1} \quad \text{with} \quad \Gamma^{(k)}_{i,j}=\gamma(i-j), \quad (1)$$

*Correspondence to: roland.fried@tu-dortmund.de

Department of Statistics, TU Dortmund University, Dortmund, Germany

Conflict of interest: The authors have declared no conflicts of interest for this article.

is positive-semidefinite. For a stationary time series the usual relation

$$\mathrm{Cor}\left(X_{t+h}, X_t\right) = \frac{\mathrm{Cov}\left(X_{t+h}, X_t\right)}{\sqrt{\mathrm{Var}\left(X_{t+h}\right) \cdot \mathrm{Var}\left(X_t\right)}}$$

$$\text{implies} \quad \rho(h) = \frac{\gamma(h)}{\gamma(0)}. \tag{2}$$

This allows us to translate estimators of the autocovariances into estimators of the autocorrelations and vice versa, if an estimate of the variance $\gamma(0)$ is available.

For a vector of observations $\mathbf{X} = (X_1, \ldots, X_n)$, let $\overline{X}$ be the arithmetic mean, $X_{(1)}, \ldots, X_{(n)}$ denote the ordered sample in ascending order and R_t the rank of X_t, $t = 1, \ldots, n$.

The sample analogues of $\gamma(h)$ and $\rho(h)$ are the empirical or sample autocovariances and autocorrelations $\hat{\gamma}(h)$ and $\hat{\rho}(h)$ (in the simulation study abbreviated as: **Emp. acf**), which are given by

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-h}\left(X_t - \overline{X}\right)\left(X_{t+h} - \overline{X}\right), \tag{3}$$

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, \quad h \in \mathbb{N}\ .$$

The denominator n is used in the formula for $\hat{\gamma}(h)$ instead of the more intuitive number of cross-products $n-h$, since this guarantees positive-semidefiniteness of the resulting functions $\hat{\gamma}$ and $\hat{\rho}$ for the price of a larger bias. In Ref 10, p. 244, an asymptotic formula for the bias of the sample acf of Gaussian processes is derived:

$$\mathrm{Bias}\left(\hat{\rho}(h)\right) = -\frac{1}{n}\left(h\rho(h) + \left(1 - \rho(h)\right)\sum_{i=-\infty}^{\infty}\rho(i) + 2\zeta(h) - 2\rho(h)\zeta(0)\right) + O\left(n^{-2}\right), \tag{4}$$

where $\zeta(h) = \sum_{i=-\infty}^{\infty}\rho(i)\rho(i+h)$. Equation (4) indicates a large negative bias in case of a small n and a large positive, slowly decaying acf. The estimator is asymptotically unbiased for fixed h as n goes to infinity. The asymptotic distribution of the sample autocorrelation can be found for example in Ref 3, Theorems 7.2.1–7.2.2. Calculation of the empirical acf is recommended only for $n \geq 50$ and $h \leq n/4$ (Ref 2, p. 32).

The sample acf can also be derived from a multivariate covariance estimation. This approach has some desirable features when carried out robustly, as will be seen later on. The matrix $\Gamma^{(k)}$ of the first autocovariances (see (1)) can be estimated by building a data matrix from the lagged observations. Let $\widetilde{X}_t$, $t \in \mathbb{Z}$, denote the centered observations. We use the sample mean $\overline{X}$ for centering, if not stated otherwise. Defining

$$\mathbf{Z}_k' = \begin{bmatrix} \widetilde{X}_1 & \widetilde{X}_2 & \cdots & \widetilde{X}_{k+1} & \cdots & \widetilde{X}_n & 0 & \cdots & 0 \\ 0 & \widetilde{X}_1 & \cdots & \widetilde{X}_k & \cdots & \widetilde{X}_{n-1} & \widetilde{X}_n & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots & & \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & \widetilde{X}_1 & \cdots & \widetilde{X}_{n-k} & \widetilde{X}_{n-k+1} & \cdots & \widetilde{X}_n \end{bmatrix} \in \mathbb{R}^{(k+1)\times(n+k)}, \tag{5}$$

the ordinary positive-semidefinite sample autocovariance matrix is obtained from Pearson's product moment covariance estimator

$$\hat{\Gamma}^{(k)} = Z_k' Z_k / n. \tag{6}$$

Application of the well known identity for correlation matrices,

$$\Xi_{i,j}^{(k)} = \Gamma_{i,j}^{(k)} \Big/ \sqrt{\Gamma_{i,i}^{(k)} \cdot \Gamma_{j,j}^{(k)}}, \tag{7}$$

yields the estimation $\hat{\Xi}^{(k)}$.

A model for stationary acfs is the autoregressive moving average (ARMA) process, which is defined by

$$X_t = \phi_0 + \sum_{i=1}^{p}\phi_i X_{t-i} + \sum_{i=1}^{q}\theta_i a_{t-i} + a_t, \tag{8}$$

with parameters $\phi_0, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q \in \mathbb{R}$, and innovations $(a_t)_{t\in\mathbb{Z}}$ forming white noise, that is a stationary sequence of uncorrelated random variables with mean zero and variance σ^2.

Of special interest are AR processes where $q = 0$,[11] since from them another identity for ρ can be derived. If all solutions z of $1 - \phi_1 z - \ldots - \phi_p z^p = 0$ are outside the complex unit circle, then (8) models a stationary process with marginal mean

$$\mu = \frac{\phi_0}{1 - \phi_1 - \ldots - \phi_p}. \quad (9)$$

The Yule-Walker equations relate the coefficients $\phi_1, \ldots, \phi_p$ of an AR(p) model to the first p autocorrelations $\rho(1), \ldots, \rho(p)$. To shorten notation we assume $\phi_0 = 0$. Then the equations are obtained by multiplying (8) with X_{t-h}, $h = 1, \ldots, p$, taking expectations and dividing by $\gamma(0)$,

$$\begin{aligned}
\rho(1) &= \phi_1 + \phi_2\rho(1) + \ldots + \phi_p\rho(p-1)\\
\rho(2) &= \phi_1\rho(1) + \phi_2 + \ldots + \phi_p\rho(p-2)\\
&\vdots\\
\rho(p) &= \phi_1\rho(p-1) + \phi_2\rho(p-2) + \ldots + \phi_p. \quad (10)
\end{aligned}$$

Autocorrelations of higher order can be extrapolated using the recursion

$$\rho(h) = \phi_1\rho(h-1) + \phi_2\rho(h-2) + \ldots + \phi_p\rho(h-p), \quad h = p+1, p+2, \ldots \quad (11)$$

Even if (X_t) is not an AR process of order p, fitting such a model can still be beneficial. Let $\pi_{p,0} + \sum_{i=1}^{p} \pi_{p,i}X_{t-i}$ denote the best approximation of X_t by an AR (p) model in the sense of mean squared error for any $p \in \mathbb{N}$. Then

$$\hat{X}_t = \pi_{h-1,0} + \sum_{i=1}^{h-1} \pi_{h-1,i}X_{t-i}, \quad (12)$$

is the best linear prediction of X_t given the previous h-1 observations and analogously

$$\hat{X}_{t-h} = \pi_{h-1,0} + \sum_{i=1}^{h-1} \pi_{h-1,i}X_{t-h+i}, \quad (13)$$

is the best linear prediction of X_{t-h} given the future up to time t. The resulting residuals

$$U_{h,t} = X_t - \hat{X}_t \quad \text{and} \quad V_{h,t} = X_{t-h} - \hat{X}_{t-h}, \quad (14)$$

are called forward respectively backward residuals. They define the pacf

$$\pi(h) = \pi_{h,h} = \begin{cases} \mathrm{Cor}(X_{t+1}, X_t), & h = 1 \\ \mathrm{Cor}(U_{h,t}, V_{h,t}), & h \geq 2 \end{cases}, \quad (15)$$

which is another important tool for model building. It measures the correlation of X_t and X_{t+h} after eliminating the linear effects of all intermediate observations $X_{t+1}, \ldots, X_{t+h-1}$. Unlike the acf, the pacf only needs to be bounded between −1 and 1 to be valid.[12]

A connection between the acf and pacf is given by the Durbin-Levinson algorithm. For a stationary process with $\mu = 0$, $\gamma(0) > 0$ and $\gamma(h) \to 0$ for $h \to \infty$ it reads

$$\pi(h) = \left(\rho(h) - \sum_{i=1}^{h-1} \pi_{h-1,i}\rho(h-i)\right) v_{h-1}^{-1}, \quad h \geq 2,$$

$$\text{where} \begin{pmatrix} \pi_{h,1} \\ \vdots \\ \pi_{h,h-1} \end{pmatrix} = \begin{pmatrix} \pi_{h-1,1} \\ \vdots \\ \pi_{h-1,h-1} \end{pmatrix} - \pi(h) \begin{pmatrix} \pi_{h-1,h-1} \\ \vdots \\ \pi_{h-1,1} \end{pmatrix}$$

$$\text{and } v_h = v_{h-1}\left(1 - \pi(h)^2\right), \quad (16)$$

with $\pi_{h,h} = \pi(h)$. The recursion starts with $\pi(1) = \rho(1)$ and $v_0 = 1$. Conversely, the acf can be derived from the pacf using the relationship given by Ref 13

$$\rho(h) = \sum_{i=1}^{h-1} \pi_{h-1,i}\rho(h-i) + \pi(h)\left(1 - \sum_{i=1}^{h-1} \pi_{h-1,i}\rho(i)\right). \quad (17)$$

Instead of estimating the partial autocorrelations (15) from the sample acf, Burg proposed an alternative estimator (see Ref 14) for $\pi(h)$ as

$$\hat{\pi}(h) = 2\frac{\sum_{t=h+1}^{n} U_{h,t}V_{h,t}}{\sum_{t=h+1}^{n} \left[U_{h,t}^2 + V_{h,t}^2\right]}. \quad (18)$$

It can be interpreted as a correlation estimator for the forward and backward residuals as the denominator estimates the sum of their variances.

In summary, the above equations allow construction of (robust) autocorrelation estimators by estimating ρ either directly, or by estimating the pacf π and using (17), or by fitting an AR model of sufficiently large order p and applying (10) and (11).

ROBUST AUTOCORRELATION ESTIMATORS

Different proposals for robust estimation of autocorrelations and partial autocorrelations have been derived using different ideas. We review such approaches in the following.

Estimation Based on Univariate Transformations

An intuitive idea of limiting the influence of outliers is rejecting or at least downweighting very large and small values of the time series, where outlyingness will be relative to the marginal distribution of X_t, ignoring the serial dependence. Such transformations reduce the effects of outliers on the sample acf, but produce a bias which does not vanish asymptotically. An exact bias correction is often not available, so we need to rely on asymptotic approximations or simulations for this. For more details see the section on implementation.

A robust estimator of autocovariances and autocorrelations can be constructed using univariate trimming (abbr.: **Trim**), that is omitting terms in the sum in (3) which correspond to the most extreme observations,

$$\hat{\gamma}^{(\alpha)}(h) = \frac{1}{\sum_{t=1}^{n-h} L_t^{(\alpha)} L_{t+h}^{(\alpha)}} \left\{ \sum_{t=1}^{n-h} \left(X_t - \overline{X}^{(\alpha)} \right) \left(X_{t+h} - \overline{X}^{(\alpha)} \right) L_t^{(\alpha)} L_{t+h}^{(\alpha)} \right\},$$

$$\text{where } \overline{X}^{(\alpha)} = \frac{1}{\sum_{t=1}^{n} L_t^{(\alpha)}} \sum_{t=1}^{n} X_t L_t^{(\alpha)}$$

$$\text{and } L_t^{(\alpha)} = \begin{cases} 1, & X_{(g)} < X_t < X_{(n-g+1)} \\ 0, & \text{otherwise} \end{cases}$$

$$\text{with } g = \lfloor \alpha \cdot n \rfloor \text{ for } 0 \le \alpha < 0.5.$$

Chan and Wei[15] propose trimming constants α between 0.01 and 0.1, depending on the suspected percentage of outliers. As usually, larger fractions α increase robustness but decrease the efficiency of the estimator at clean samples without outliers. The acf is estimated by dividing the trimmed autocovariance through the trimmed variance $\hat{\gamma}^{(\alpha)}(0)$. Simulations indicate that without a bias correction the estimator is significantly biased for $n = 50$ (see also Ref 15), and it may be easily seen that the bias does not vanish asymptotically if α is fixed.

To obtain high robustness, Chakhchoukh[16] suggests substituting the sum in the sample acf by the median, calculating

$$\hat{\rho}(h) = \frac{\text{med}\left(\widetilde{X}_1 \widetilde{X}_{1+h}, \ldots, \widetilde{X}_{n-h} \widetilde{X}_n \right)}{\text{med}\left(\widetilde{X}_1^2, \ldots, \widetilde{X}_n^2 \right)},$$

where $\widetilde{X}_t$ is the centered time series, for example using the median. This estimator (abbr.: **Mediancor**) can be seen as a limiting case of the above trimming based estimator, with $\alpha = 0.5$. For an asymptotically consistent estimation of $\rho(h)$ a nonlinear transformation of $\hat{\rho}(h)$ is necessary, which needs to be determined numerically.

With the aim of robustly fitting time series models, Bustos and Yohai[17] introduces the so called residual autocovariances (RA-estimators), which can also be used to estimate the acf. Albeit being defined more generally, this approach boils down to a more sophisticated transformation of the time series (see Ref 17, for the general definition). Instead of trimming a constant amount of the largest and smallest observations, observations are downweighted only if being unusually large or small. Note that the amount of rejected observations depends on the sample itself. For the transformed time series Y_t, $t = 1, \ldots, n$, one gets

$$Y_t = \psi\left(\frac{X_t - m}{s} \right), \tag{19}$$

where m and s are suitable estimators for μ and $\gamma(0)$. The median and the median absolute deviation about the median (MAD) are common robust choices for these quantities (see Ref 18, for definitions and properties). Conventional choices of the transformation function ψ with tuning parameters c_j are the Huber function

$$\psi(x) = \psi_{c_1}(x) = \text{sign}(x) \min\left(|x|, c_1 \right) \tag{20}$$

and Tukey's bisquare function

$$\psi(x) = \psi_{c_2}(x) = \begin{cases} x\left(1 - x^2/c_2^2\right)^2, & 0 \le |x| \le c_2 \\ 0, & |x| > c_2. \end{cases} \tag{21}$$

The resulting estimators are abbreviated by **RA-Huber** and **RA-Tukey**. The objective of Ref 17 was not estimation of the acf, so a bias correction was not proposed. However, common choices of the tuning parameters like $c_1 = 1.37$ for the Huber function and $c_2 = 4.68$ for Tukey's function modify a Gaussian time series only slightly in the absence of outliers, so that the resulting bias is small.

Estimation Based on Signs and Ranks

For the purpose of model selection Garel and Hallin[19] introduces rank-based statistics, which can also be applied for acf estimation. Construction of ranks means a special data transformation

as treated in the previous subsection. Nevertheless we present this approach separately together with sign based estimators since both are popular utilities from nonparametric statistics and often mentioned together. Additionally, bias corrections are known explicitly at least for Gaussian processes.

Since we are more interested in estimation than in testing, we use a definition slightly different from Ref 19, namely

$$\hat{\rho}(h) = c\sum_{i=1}^{n-h} J\left(R_i/(n+1)\right)\cdot J\left(R_{i+h}/(n+1)\right), \quad (22)$$

with $c = 1/\sum_{i=1}^{n} J\left(R_i/(n+1)\right)^2$ and J a score function. Van der Waerden or normal scores are obtained by

$$J(x) = \Phi^{-1}(x), \quad x \in (0,1),$$

where $\Phi(x)$ is the cumulative distribution function of a standard normal, and lead to asymptotically optimal tests under normality.[19] The asymptotical Gaussian efficiency of the resulting estimator is higher than those of other rank-based estimators.[20] However, the related Gaussian rank correlation (abbr.: **GRCor**) is not very robust against outliers.[21]

More widely used is the Spearman score function $J(x) = x - (n+1)/2$, which result in an autocorrelation estimator based on the popular Spearman's ρ (abbr.: **Spearman**). Whereas van der Waerden scores yield an asymptotically unbiased estimation in the normal case, Spearman's ρ needs to be transformed by $g(\rho) = 2\sin(\pi\rho/6)$, see for example Ref 22.

Further popular nonparametric correlation estimators are Kendall's τ (abbr.: **Kendall**)

$$\hat{\rho}(h) = \frac{1}{(n-h)(n-h-1)}\sum_{i>j}\operatorname{sign}\left(\left(X_i - X_j\right)\left(X_{i+h} - X_{j+h}\right)\right)$$

and the quadrant correlation (abbr.: **Quadrant**)

$$\hat{\rho}(h) = \frac{1}{n-h}\sum_{i=1}^{n-h}\operatorname{sign}\left(\left(X_i - \hat{\mu}\right)\left(X_{i+h} - \hat{\mu}\right)\right),$$

where the center $\hat{\mu}$ is estimated by the median of the timeseries. For both estimators transformation by $g(\rho) = \sin(\pi\rho/2)$ yields unbiasedness under the bivariate normal distribution, and also for a wider range of distributions.[23] A disadvantage of such transformations is that they can destroy the positive-semidefiniteness of the estimators.

Estimation Based on Partial Autocorrelation

Autocorrelation estimators constructed from pairwise correlation estimators possibly lack positive-semidefiniteness as mentioned before. Valid estimation of the pacf is easier, since one only needs to ensure estimates within −1 and 1. Using relation (17) between pacf and acf with initialization $\hat{\pi}(1) = \hat{\rho}(1)$ then yields a positive-semidefinite autocorrelation estimation. This motivates estimating the pacf first, as suggested by Refs 13, 23. Both approaches differ in the choice of the correlation estimator for $\pi(h)$ based on the identity (15).

An M-estimator as a variant of the Burg estimator (18) is proposed by Masarotto[13]:

$$\hat{\pi}(h) = 2\frac{\sum_{t=1}^{n-h} W_{h,t}\left(X_t - \hat{X}_t\right)\left(X_{t+h} - \hat{X}_{t+h}\right)}{\sum_{t=1}^{n-h} W_{h,t}\left[\left(X_t - \hat{X}_t\right)^2 + \left(X_{t+h} - \hat{X}_{t+h}\right)^2\right]}, \quad (23)$$

where $W_{h,t} = w\left(d_{ht}(b)/s_{ht}^2\right)$ with weight function $w(x) = 3/(1+x)$, $d_{ht}(b) = U_{ht}^2 + V_{ht}^2 - 2bU_{ht}V_{ht}$ and s_{ht} satisfying

$$\sum_{t=h+1}^{n} w\left(d_{ht}(b)/s_{ht}^2\right) d_{ht}(b) = 2(n-h)s_{ht}^2.$$

Masarotto[13] argues that the resulting acf estimator (abbr.: **PA-M**) is consistent and asymptotically normal at least under normality. Asymptotical confidence bands can be constructed numerically. As an alternative, Mottonen et al.[23] proposes sign and rank-based correlation estimators, e.g., Spearman's ρ (abbr.: **PA-Spearman**), Kendall's τ (abbr.: **PA-Kendall**) and quadrant correlation (abbr.: **PA-Quadrant**), as described in the previous subsection. Generally, every robust bivariate correlation estimator can be applied.

Estimation Based on Multivariate Correlation

Approaches based on univariate transformations ignore the serial dependence of the data, possibly downweight good observations and overlook outliers. The left panel of Figure 1 depicts a realization of an AR process with $\phi_0 = 0$ and $\phi_1 = 0.9$. Prediction bounds based on the univariate marginal distribution simply identify the most extreme observations as possible outliers, although these observations might be due to the dynamics of the underlying process. A bivariate or even multivariate analysis based on the marginal distribution of subsequent observations[24]

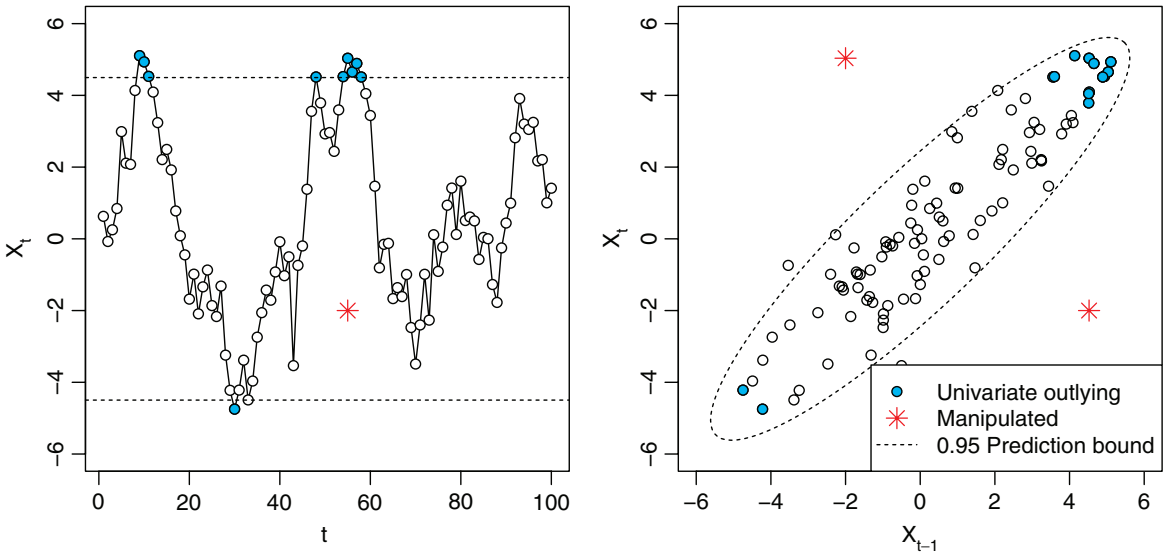


FIGURE 1 | Time series with 95% prediction bounds based on the univariate (left) and the bivariate (right) marginal distribution corresponding to subsequent observations. Univariate margins identify the most extreme observations as outliers, while multivariate inspection takes the dependencies between subsequent observations into account and identifies the true outliers.

allows us to take the dependencies among subsequent observations into account, and can achieve better downweighting of spurious observations in the subsequent analysis than a simple univariate consideration. Estimation of the acf from a robust estimate of the multivariate covariance matrix is thus promising. Such estimators can be based, e.g., on multivariate trimming or weighting. Moreover, some multivariate robust correlation estimators even gain efficiency with increasing dimension.[25]

Multivariate methods can be formulated in terms of the data matrix $\mathbf{Z}_k$ in (5). Note that centering is unnecessary, since the described approaches estimate a robust center. The computing time of many robust procedures increases exponentially in the dimension,[26] so one should choose k rather small. To simplify notation, we denote the i-th row of $\mathbf{Z}_k$ as $\mathbf{M}_i'$, so that we are in the usual multivariate case. The estimation result will always be a valid covariance matrix but it does not have the Toeplitz structure with constant off-diagonals, albeit by definition all values of the h-th off-diagonal estimate $\rho(h)$. An intuitive solution is averaging the values across each off-diagonal, i.e.,

$$\hat{\rho}(h) = \frac{1}{k-h+1} \sum_{i=1}^{k-h+1} \hat{\Xi}^{(k)}_{i,i+h}, \tag{24}$$

though positive-semidefiniteness gets possibly lost then.

There is a large literature on robust multivariate correlation estimation. We will concentrate on the most common proposals, but of course others could be employed as well.

An M-estimator of scatter (abbr.: **Multi-M**), which can be represented as a weighted least squares estimate, is introduced in Ref 27, see also Ref 9. Given an initial estimator $\left(\hat{\mu}, \hat{\Sigma}\right)$ for expectation and covariance, robust weights are obtained from the outlyingness of the observations as measured by the Mahalanobis distance

$$d_i^2 = \left(\mathbf{M}_i - \hat{\mu}\right)' \hat{\Sigma}^{-1} \left(\mathbf{M}_i - \hat{\mu}\right), \quad i = 1, \ldots, n+k. \tag{25}$$

After that, the estimation is sequentially updated by

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n+k} v\left(d_i\right) \left(\mathbf{M}_i - \hat{\mu}\right) \left(\mathbf{M}_i - \hat{\mu}\right)'$$

$$\text{and} \quad \hat{\mu} = \frac{\sum_{i=1}^{n+k} w\left(d_i\right) \mathbf{M}_i}{\sum_{i=1}^{n+k} w\left(d_i\right)}, \tag{26}$$

where $w(d_i)$ and $v(d_i)$ are suitable weights. Popular are Huber weights: $w(d_i) = \min(1, c_0/|d_i|)$ and $v(d_i) = w(d_i)^2/r$, where c_0 determines robustness and efficiency and r depending on c_0 and the probability model ensures consistency. Using the weight function $v(d) = (k+1)/d^2$ results in Tyler's M-estimator (abbr.:

Multi-TylerM), which is a kind of minimax estimator within the elliptical model.[28]

The breakdown point[29] of M-estimators cannot exceed an upper bound which decreases with increasing dimension.[9] Since the effective amount of outlying pairs in the estimation of the acf can be twice the number of outlying observations, other estimators might be preferred if one is interested in larger time lags.

The disadvantage of the decreasing breakdown point does not apply to multivariate S-estimators[30,31] (abbr.: **Multi-S**). They are defined as

$$\hat{\Sigma} = \arg\min_{\mu,\ \Sigma} \left\{ \det(\Sigma) : \frac{1}{n+k}\sum_{i=1}^{n+k} w \left(\left(\mathbf{M}_i - \mu\right)' \Sigma^{-1} \left(\mathbf{M}_i - \mu\right) \right) = b_0 \right\},$$

where w is a bounded smooth and nondecreasing function, e.g.

$$w(y) = \min\left(\frac{y^2}{2} - \frac{y^4}{2c^2} + \frac{y^6}{6c^4}, \frac{c^2}{6} \right).$$

The constant c determines the breakdown point, whereas b_0 depends on the probability model; see Ref 32 for more details. An algorithm for computing this implicitly defined estimator can be found in Ref 33. Although the breakdown point does not depend on the number of dimensions, single outliers can cause a larger bias in higher dimensions.[9]

A popular robust covariance estimator is the minimum covariance determinant (abbr.: **Multi-MCD**).[34,35] For a given constant α between 0 and 0.5 the usual product moment covariance is calculated for the subset of proportion $1-\alpha$ which leads to the matrix with the smallest determinant. An approximate procedure was proposed by Ref 36, since finding this subset is very time consuming for large n. Larger trimming constants α lead to more robust but less efficient estimators, with the efficiency for large α being rather low.[37] To combine high robustness and large efficiency, often an additional reweighting step is added (abbr.: **Multi-wMCD**): Robust Mahalanobis distances are obtained based on an initial MCD fit, and then the ordinary covariance matrix is calculated from all observations with Mahalanobis distances not exceeding a certain quantile of the χ^2-distribution with $k+1$ degrees of freedom. The 0.975-quantile has been recommended for this cut-off.[36] An asymptotically fully efficient reweighting step (abbr.: **Multi-effMCD**) with a data-adaptive choice of the quantile was suggested in Ref 38.

Multivariate outliers can be inconspicuous if one only looks at individual dimensions, but there is always a one-dimensional projection in which the observation is outlying.[39] Based on this idea, Donoho and Stahel[40,41] propose to use the maximal distance to the median for every possible projection to measure outlyingness, i.e.,

$$r_i = \max_{\mathbf{a}:\|\mathbf{a}\|=1} \frac{\mathbf{a}'\mathbf{M}_i - \text{Median}\left(\mathbf{a}'\mathbf{M}_1, \ldots, \mathbf{a}'\mathbf{M}_{n+k}\right)}{\text{MAD}\left(\mathbf{a}'\mathbf{M}_1, \ldots, \mathbf{a}'\mathbf{M}_{n+k}\right)},$$

with $\|\cdot\|$ being the Euclidean norm. Practical algorithms only consider a finite set of randomly chosen vectors for $\mathbf{a}$. The number of such directions needs to increase strongly for higher dimensions to ensure reliable outlier detection. The resulting Stahel-Donoho estimator (abbr.: **Multi-SD**) is defined as the weighted covariance

$$\hat{\Sigma} = \frac{1}{\sum_{i=1}^{n} w_i} \sum_{i=1}^{n} w_i \left(\mathbf{M}_i - \hat{\mu}\right)\left(\mathbf{M}_i - \hat{\mu}\right)'$$

$$\text{and } \hat{\mu} = \frac{1}{\sum_{i=1}^{n} w_i} \sum_{i=1}^{n} w_i \mathbf{M}_i.$$

A common choice of the weight function is $w_i = \min(1, (c/r_i)^2)$, and c is often chosen as the 0.975-quantile of the χ^2-distribution with $k+1$ degrees of freedom.[37]

Estimation Based on Variances

An estimation principle for covariances and correlations based on estimators of variances has been proposed in Ref 42. In the context of autocorrelation estimation for stationary time series, the underlying formula reads

$$\rho(h) = \text{Cor}\left(X_{t+h}, X_t\right) = \frac{\text{Var}\left(X_{t+h} + X_t\right) - \text{Var}\left(X_{t+h} - X_t\right)}{\text{Var}\left(X_{t+h} + X_t\right) + \text{Var}\left(X_{t+h} - X_t\right)}, \quad (27)$$

see Ref 43. The usual correction factors necessary for making robust scale estimators consistent at a certain distribution are not needed when applying them for correlation estimation, since they cancel out if $X_{t+h} + X_t$ and $X_{t+h} - X_t$ are in the same location-scale family. This is fulfilled, e.g., if X_{t+h} and X_t are jointly normal or, more generally, elliptically-symmetric distributed. Note that this approach does not necessarily yield a positive-semidefinite estimation of the acf.

For estimation of the variances on the right hand side of (27), trimmed and winsorized variances have

been suggested.[42] Since any reasonable estimator of variability can be applied, Ma and Genton[43] propose Q_n,[44] (abbr.: **GK-Qn**), because of its high asymptotic breakdown point of 0.5 and its good asymptotic efficiency of 0.82 for i.i.d. Gaussian samples. The Q_n corresponds roughly to the first quartile of all absolute pairwise distances between all pairs of observations.

In the context of ordinary correlation Maronna and Zamar[45] recommended the so called τ-scale estimator (abbr.: **GK-tau**)

$$\widehat{\sigma}^2\left(X_1, \ldots, X_n\right) = \frac{\widehat{\sigma}_0^2}{n}\sum_{i=1}^{n} d_{c_2}\left(\frac{X_i - \widehat{\mu}}{\widehat{\sigma}_0}\right), \qquad (28)$$

where $\widehat{\mu}$ is an adaptively weighted mean of the observations, $\widehat{\sigma}_0$ their MAD and $d_c(x) = \min(x^2, c^2)$. Tuning constants of $c_1 = 4.5$ (for $\widehat{\mu}$) and $c_2 = 3$ results in a good trade-off between efficiency and robustness and an asymptotic Gaussian efficiency of 0.8 in case of independent observations.[45] The good properties of this estimator in the bivariate i.i.d. case are promising for the estimation of autocorrelations.

Estimation Based on Robust Filtering

As mentioned above, clean observations can be outlying with respect to the marginal distribution and thus be unnecessarily downweighted by estimators based on univariate transformations, if the autocorrelations $\rho(h)$ are large positive and slowly decaying. The robust filtering approach overcomes this problem by taking the time series structure into account. The idea is to measure the outlyingness of the prediction residuals $U_{p,t}$ instead of X_t itself. After replacing outliers by reasonable values, one can either calculate the sample acf (abbr.: **Filter-acf**) or use the fitted AR process and translate this into the acf via the Yule-Walker equations (abbr.: **Filter-ar**). Robust filtering was already introduced by Masreliez[46], but we will stick to the filter described in Ref 9, which is a slight modification proposed by Martin[47]. Note that this algorithm is quite extensive so we will summarize only the main ideas and refer to (Ref 9, p. 272–277, 320–321) for details.

Let $\widetilde{X}_t$ be centered for example by the median and approximate the process by an AR model of order $p \in \mathbb{N}$. A kind of robust AIC criterion to determine p was proposed by Maronna et al.[9]

Let $\mathbf{Y}_t = (Y_t, \ldots, Y_{t-p+1})'$, denote the vector of robustly filtered values and

$$\Phi = \begin{pmatrix} \phi_1, \ldots \phi_{p-1} & \phi_p \\ \mathbf{I}_{p-1} & \mathbf{0}_{p-1} \end{pmatrix}, \qquad (29)$$

the so called transition matrix. From this one calculates the one step ahead predictions

$$\widehat{X}_t = \sum_{i=1}^{p} \phi_i Y_{t-i}$$

and its residuals

$$\widetilde{U}_t = X_t - \widehat{X}_t.$$

Note that this is similar to usual prediction residuals defined in (14), just replacing X_j by Y_j for $j = t-1, \ldots, t-p$ to make it more robust. Eventually one sets

$$\mathbf{Y}_t = \Phi \mathbf{Y}_{t-1} + \frac{m_t}{s_t}\psi\left(\frac{\widetilde{U}_t}{s_t}\right), \qquad (30)$$

where ψ should fulfill

$$\psi(x) = \begin{cases} x & |x| < d_1 \\ 0 & |x| > d_2, \end{cases} \qquad (31)$$

with $0 < d_1 < d_2$. Our proposal is to use a polynomial of degree three between d_1 and d_2. It is uniquely defined by forcing ψ to be continuous differentiable:

$$\psi(x) = |x|\left(a + b|x| + dx^2 + e|x|^3\right) \text{ for } d_1 \le |x| \le d_2, \qquad (32)$$

where

$$a = \frac{2d_1^2 d_2^2}{\left(d_1 - d_2\right)^2} \quad b = \frac{-d_2^3 - d_1 d_2^2 - 4d_1^2 d_2}{\left(d_1 - d_2\right)^2}$$

$$d = \frac{2d_2^2 + 2d_1 d_2 + 2d_1^2}{\left(d_1 - d_2\right)^2} \quad e = \frac{-d_1 - d_2}{\left(d_1 - d_2\right)^3}.$$

Furthermore $\mathbf{m}_t \in \mathbb{R}^p$ contains estimations of the variance of the prediction residual s_t^2 and the covariances between the residual $\tilde{U}_t$ and the robustly filtered values $Y_{t-1}, \ldots Y_{t-p+1}$. Note that these estimations are time-dependent instead of a simpler global estimation. The reason for this is that in case of outliers there is a chance that the algorithm otherwise might lose track of the data afterwards. In this case s_t will increase and thus provide the filtered values more variation to get back to the data more quickly. See Ref 47 for more details. Recursions for $\mathbf{m}_t$ are given as

$$\mathbf{M}_t = \Phi \mathbf{P}_{t-1} \Phi' + \mathbf{d}\mathbf{d}' \widehat{\sigma}_u^2$$

and

$$\mathbf{P}_t = \mathbf{M}_t - \frac{1}{s_t^2}\psi\left(\frac{\widetilde{U}_t}{s_t}\right)\frac{s_t}{\widetilde{U}_t}\mathbf{m}_t\mathbf{m}_t'$$

where $\mathbf{m}_t$ is the first column of $\mathbf{M}_t$, $\mathbf{d} = (1, 0, \ldots, 0)' \in \mathbb{R}^p$ and $\hat{\sigma}_u^2$ an estimator for the variance of the prediction residuals.

Looking only at the first row of Eq. (30) we get

$$Y_t = \hat{X}_t + s_t \psi\left(\frac{\tilde{U}_t}{s_t}\right), \tag{33}$$

indicating that Y_t will be close to X_t if $|\tilde{U}_t|$ is small, and close to $\hat{X}_t$ if it is large. Using the vector recursion instead of a simpler one-dimensional equation (33) offers the advantage that if X_t is an outlier, the algorithm will also use future information on X_s, $s > t$, to determine $\hat{X}_t$.

The parameters $\phi_1, \ldots, \phi_p$ can be estimated by minimizing the variance of the prediction residuals

$$\hat{\sigma}\left(\tilde{U}_{p+1}(\phi_1, \ldots, \phi_p), \ldots, \tilde{U}_n(\phi_1, \ldots, \phi_p)\right). \tag{34}$$

For $\hat{\sigma}$, Maronna et al.[9] proposed the τ-scale (28) because of its quick computation and good robustness. Since a nonconvex function needs to be optimized over p parameters, they suggested sequential minimization based on the Durbin-Levinson algorithm. This converts the problem into p one-dimensional optimizations, which each is proposed to be optimized by a line search.

Implementation

To the best of our knowledge, implementations of robust autocorrelation estimators are scarce in statistical software packages. The robust filtering approach and the GK approach (see Ref 43) have been implemented in S-Plus. Implementation of the estimation procedures is straightforward at least for the approaches which are based on two- or multidimensional correlation estimators, since these estimators are readily available in R.[48]

The multivariate S, Stahel-Donoho and MCD correlation estimators are available in the `rrcov` package.[49] For Tyler's multivariate M-estimator the `ICSNP` package[50] can be used, whereas a multivariate M-estimator is available in the `SpatialNP` package.[51] For the data-adaptively reweighted MCD we use the conventionally weighted MCD as a starting estimator, resulting therefore in a two step reweigthing procedure for the raw MCD.

For Masarotto's approach we choose the median and Q_n as initial estimators for location and spread of the time series. Robust scale estimators like the τ-scale and Q_n can be found in the R package `robustbase`.[52]

TABLE 1 | Tuning Parameters Used in Our Simulation Study

Estimators	Tuning Parameters
RA-Huber	$c_1 = 1.37$
RA-Tukey	$c_2 = 4.68$
Trim	$\alpha = 0.1$
Multi-M	$c_0 = \sqrt{F_{k+1}^{-1}(0.9)}$, $r = F_{k+3}(c_0^2) + \frac{c_0^2 \cdot 0.1}{k+1}$ with $F_k \sim \chi_k^2$
Multi-S	$c = 6.02$ (maximal lag of 7)
Multi-MCD	$\alpha = 0.5$
SD	$c = \sqrt{F_{k+1}^{-1}(0.95)}$ with $F_k \sim \chi_k^2$
GK-tau	$c_1 = 4.5$, $c_2 = 3$
Filter-acf / ar	$c_1 = 4.5$, $c_2 = 3$ for (34), $d_1 = 2$, $d_2 = 3$ for (31)

We use an implementation of the robust filtering approach of our own. To translate the estimated $\hat{\phi}_i$ into the acf we apply (17) instead of (10), fitting AR models of increasing order, since it produces more stable results. To optimize (34) we use the R function `optimize`, which is based on the ALGOL localmin procedure (see Ref 53, pp. 72–80, for details). We noticed in our simulations that this algorithm yields comparable results but needs only a fraction of the computation time of the proposed grid search.

We already mentioned the importance of bias reducing transformations for estimators based on univariate transformations. We approximate these under the assumption of i.i.d. bivariate normal data, see Ref 16. Note that this transformation is independent of the time series model. Based on 10,000 runs with $n = 1000$ observations each we calculate the average estimated correlation for correlation coefficients ρ varying on a fine grid between -1 and 1. The transformation is approximated by the inverse of the linear interpolation of these averages. Like for the transformations of Kendall's τ and Spearman's ρ the result is biased for small sample sizes n, but for increasing n and normally distributed innovations this bias will vanish.

Many of the proposed estimators require the choice of tuning parameters. For the purpose of a fair comparison these parameters are often selected by making all estimators equally efficient in simulation studies. Unfortunately, there are many estimators without tuning possibility and a lack of theoretical results so that this is not possible here. We use the tuning parameters proposed by the respective authors instead. If possible we favor more robust versions, since our simulation scenarios often contain highly contaminated data. The chosen tuning parameters can be found in Table 1.

Positive-Semidefiniteness

From the above approaches the usual sample acf, the procedures using partial autocorrelations, the acf of the robustly filtered values as well as the Gaussian rank autocorrelation are guaranteed to be positive-semidefinite.

Bivariate correlation estimators do not necessarily yield positive-definite correlation matrices unless they calculate the usual correlation based on transformed data. A further problem arises for multivariate correlation estimators, resulting in positive-semidefinite matrices which do not possess a Toeplitz structure, meaning that there will be different values on the off-diagonals. Enforcing this by averaging the off-diagonals, for instance, can destroy the positive-semidefiniteness. Construction of the empirical counterpart of the correlation matrix $\Xi^{(k)}$ defined in formula (7) allows to apply transformations which achieve positive-semidefiniteness, but this destroys the Toeplitz structure.

A more appealing approach is finding the best positive-semidefinite Toeplitz approximation, minimizing, e.g., the Frobenius norm.[54] In our simulations we use the simple projection method proposed there, which can be described as follows. Let A be any real symmetric matrix, in our case $\widehat{\Xi}^{(k)}$, and $A = UDU'$ denote an eigenvalue decomposition, where D is a diagonal matrix containing all eigenvalues. If A is not positive-semidefinite there will be some eigenvalues smaller than zero. Setting these to zero yields the matrix $\widetilde{D}$ and results in a projection $P_p(A) = U\widetilde{D}U'$, which is positive-semidefinite but not Toeplitz. Furthermore, denote by $P_t(A)$ the matrix which results from setting all off-diagonal elements of order j to its average for $j = 1, \ldots, k-1$. These projections can be iterated until the change in the Frobenius norm becomes negligible.

SIMULATIONS

There are only a few theoretical results available to compare the different autocorrelation estimators. We thus perform a simulation study using the statistical software R[48] with the functions and tuning constants mentioned in the section on implementation. The simulation code is available from the authors upon request. We mainly focus on first order ARMA processes (as defined in 8) because of their simplicity and popularity, considering the seven parameter settings shown in Table 2 along with their abbreviations. Figure 2 indicates that this selection covers rather different acfs. If not explicitly stated otherwise, the innovations are standard normal.

TABLE 2 | Considered Processes and Their Abbreviations

Abbr.	AR[0]	AR[0.4]	AR[0.8]	AR[−0.4]
ϕ_1	0	0.4	0.8	−0.4
θ_1	0	0	0	0

Abbr.	AR[−0.8]	MA[0.4]	ARMA[0.4,0.4]	ARMA[0.8,−0.4]
ϕ_1	−0.8	0	0.4	0.8
θ_1	0	0.4	0.4	−0.4

We calculate the acf only for the first seven lags for different reasons. Multivariate correlation estimators are time consuming for large lags and the acf of most of the processes is nearly zero for lags larger than six. So we do not expect qualitatively different behavior for higher time lags. However, simulations indicate a slight loss of efficiency of robust estimators for higher time lags.

To simplify the comparison we consider maximal bias and minimal efficiency across all lags instead of looking at all seven time lags separately. Simulations reveal that the maximal (absolute) bias

$$\max_{h=1,\ldots,7} |\text{Bias}\left(\widehat{\rho}\left(h\right)\right)|$$

is usually realized for $h = 1$, whereas the minimal efficiency compared to the sample acf $\widetilde{\rho}$

$$\min_{h=1,\ldots,7} \left(\text{MSE}\left(\widetilde{\rho}\left(h\right)\right) / \text{MSE}\left(\widehat{\rho}\left(h\right)\right)\right)$$

occurs often for $h = 1$ or the largest lag considered here, $h = 7$. In the case of contaminated data, we calculate the efficiency relative to the sample acf for clean data. This measures the amount of information lost due to outliers when using a robust estimator.

Efficiency for Uncontaminated Data

First we investigate the properties in case of clean data without outliers, starting with the AR[0.4] model. The results are based on 10,000 runs each. As mentioned before, the empirical acf is biased for small n. As can be seen in Figure 3, the small sample bias is comparable to that of the robust alternatives. Usually there is a bias toward 0, except for the PA-Quadrant and the robust filtering approach. For the latter the bias changes from negative values for small n to slightly positive values for large n. This is not surprising. It was already mentioned by Maronna et al.[9] that smoothing the time series produces a nonvanishing bias. For small n this is overruled by the natural negative bias of the autocorrelation estimation. Tyler's M-estimator and RA-estimators are less biased than the other methods

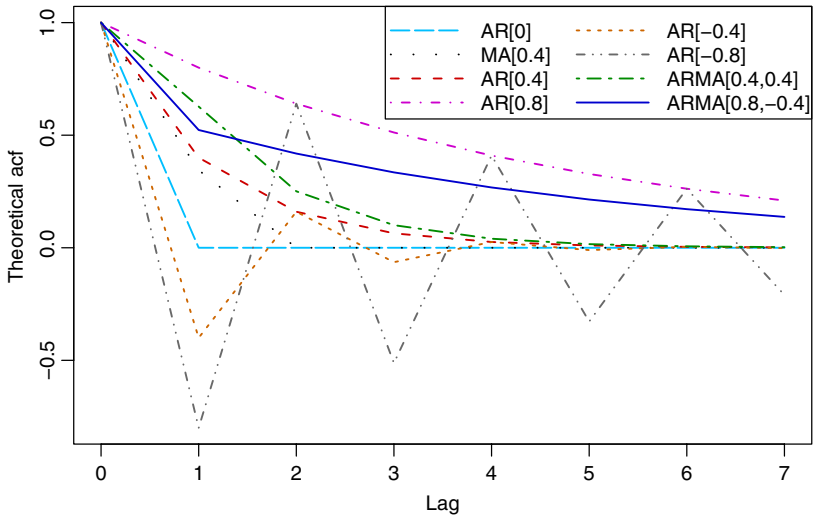


FIGURE 2 | Autocorrelation functions of the processes considered in the simulations.

for small n, resulting in a good finite sample efficiency. Generally, multivariate S- and M-estimators achieve high efficiencies. GK, rank and sign based approaches and also the reweighted MCD versions need larger samples to get a small MSE. As opposed to this, the relative efficiency of the estimation by the raw MCD, Tyler's M-estimator, the RA-approaches, the Filter-ar method and the median correlation decreases. For the MCD this is in line with simulation results in the multivariate case (see Ref 37). For the Filter-ar algorithm we notice an increase of the variance relative to the empirical acf, whereas a slower decay of the bias is the reason for the other estimators.

The findings for other models are similar. For processes with strong positive autocorrelations the maximal bias increases for all estimators just as the minimal efficiency. Nevertheless, the order of the estimators with respect to efficiency nearly stays the same. The Filter algorithms are an exception having a relatively low efficiency compared to the other approaches if the process has small absolute correlations. An explanation could be that, generally, the estimator of the fitted AR parameter ϕ_i has a large variance in this case, which affects the precision of the filtered values.

In time series we often face distributions with tails heavier than the Gaussian.[55–58] Estimators should remain reliable in case of such departures from normality. Therefore we considered maximal absolute bias and minimal efficiency for AR models with t-distributed innovations of different degrees of freedom. Already for three degrees of freedom the results were similar to those under normality. Only estimators based on partial autocorrelation lose considerable efficiency. Note that three degrees of freedom correspond to the heaviest tails possible for which the acf is defined under t-distributions. Our simulations agree with the theoretical result that the sample acf of a linear process is still $\sqrt{n}$ -consistent without the need of fourth moments, see Ref 59.

Robustness Under Contamination

Additive outliers are known to be particularly harmful for the estimation of dependence parameters. Such outliers describe, e.g., measurement errors, where a certain value ω is added to the observation at time t, see Ref 60. While for the empirical acf the effect of an outlier increases in ω due to monotonicity, for robust estimators this influence is generally bounded. However, while the influence is still monotone in ω for rank-based or other monotone estimators, it can even decrease for very large values of ω for other estimators, e.g., for so called redescenders like S-estimators, see Figure 4. This means that different outlier sizes are worst-case for different estimators. Since we are interested in estimators with a good overall performance the outlier size ω is sampled from a normal distribution with mean 0 and variance $a^2 \cdot \gamma(0)$, $a \in \{5, 10, 20\}$. This produces some small perturbations, inconspicuous outliers which are favorable for monotone estimators, as well as very large outliers which redescenders can cope well with. Note that the outlier variance is proportional to the process variance. In general, smaller values of a favor monotone estimators, while larger values favor redescenders and GK-approaches.

Furthermore, we contaminate an increasing number $n_0 \in \{5, 10, 15, 20, 25\}$ of values of the original time series of length $n = 100$ to investigate how many outliers an estimator can deal with. It was argued by Ma and Genton[43] that estimators of the autocovariance cannot be expected to cope with more than 25% contaminated observations since one outlier can enter two pairs of observations entering

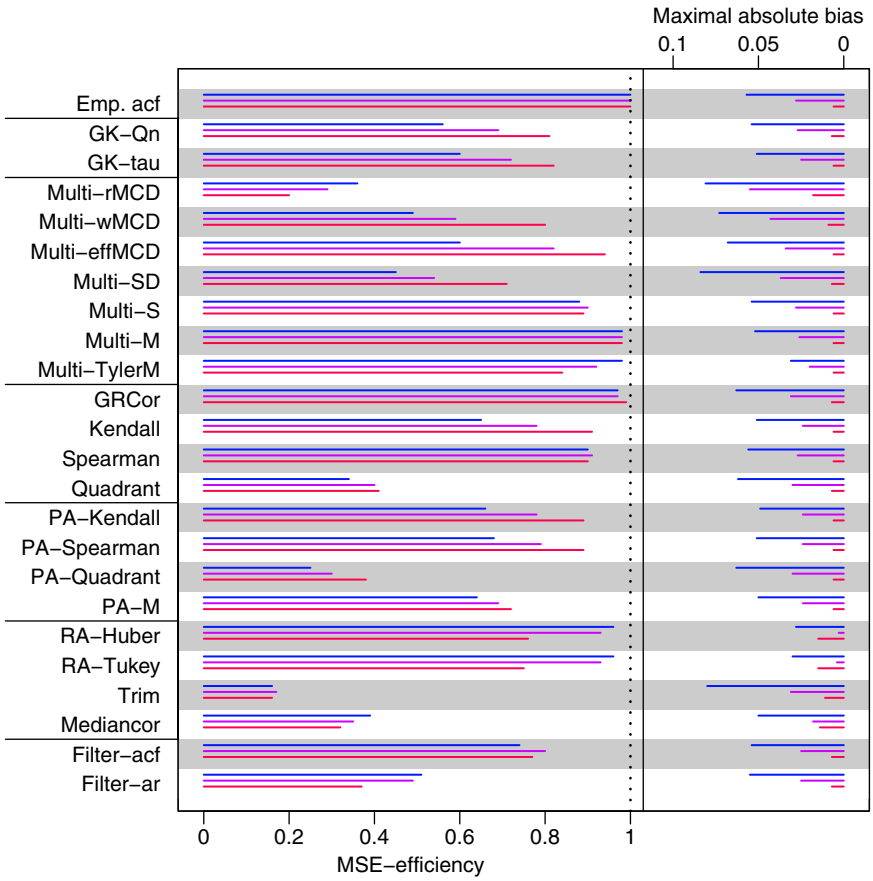


FIGURE 3 | Efficiency (left) and bias (right) for $n = 50, 100, 500$ (from top to bottom in each panel) for an AR[0.4] model with normal innovations.

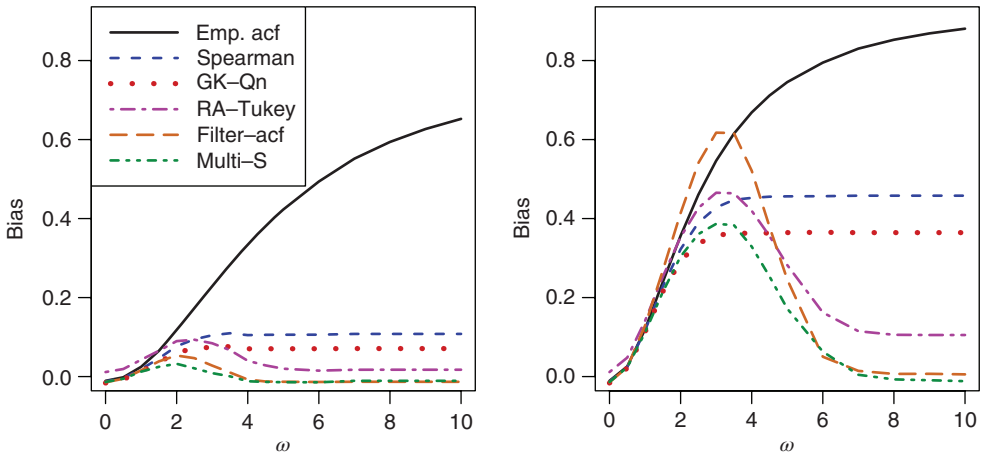


FIGURE 4 | Simulated bias of a contaminated AR[0] model with $n = 100$ and a patch of 5 (left) or 20 outliers (right).

the calculation of the correlation. Accordingly it is not reasonable to choose n_0 larger than 25. Moreover, we consider both isolated and patchy outliers, since these will have different effects. All results are based on 1000 simulation runs each.

We first treat the situation of isolated outliers, which drive the sample acf toward zero. The positions of the outliers are chosen at random for each time series. We first show the results for the AR[0.8] model with $a = 5$, which corresponds to an outlier variance of $25\gamma(0)$. As one can see in Figure 5, the empirical acf becomes useless already for $n_0 = 5$ outliers. In the same situation, some robust alternatives lose more than half of their efficiency. Nevertheless they are all preferable

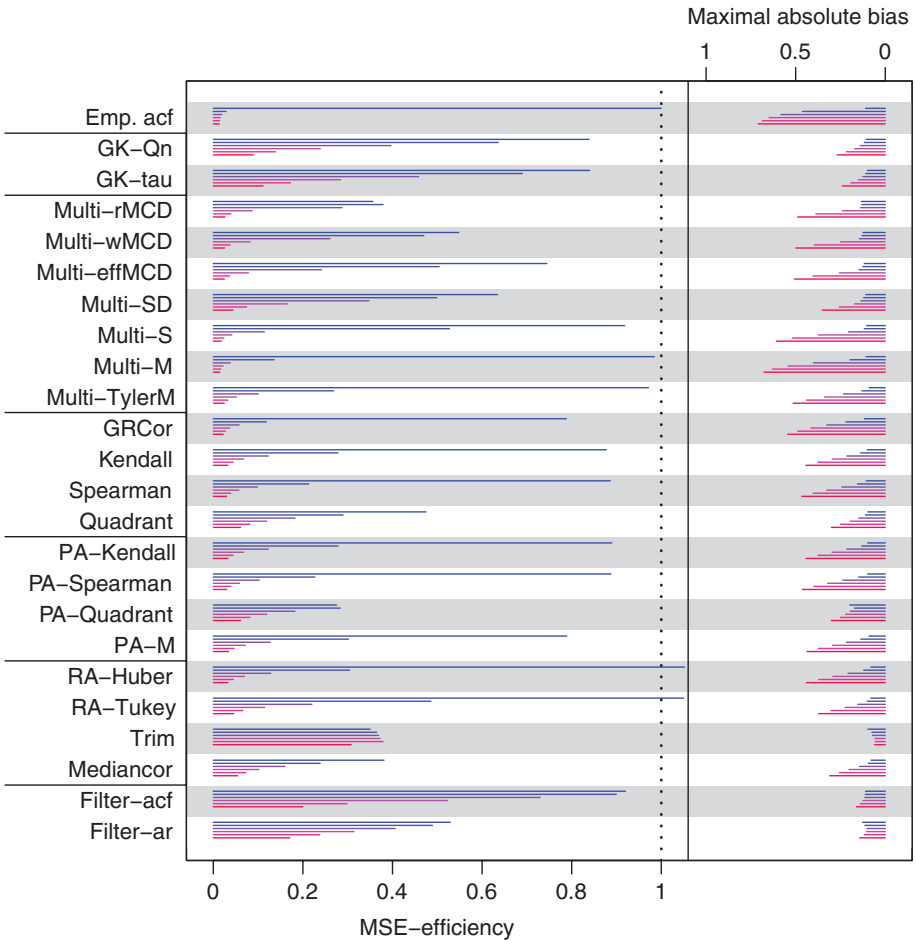


FIGURE 5 | Efficiency (left) and bias (right) for a contaminated AR[0.8] model with $n = 100$ and $n_0 = 0, 5, 10, 15, 20, 25$ (from top to bottom in each panel) isolated outliers and $a = 5$.

to the empirical acf. Their efficiency is at least 4.2 times larger.

Estimators which cope especially well with additive outliers are those based on robust filtering and to some extent also the GK approaches.

If one increases the outlier variance to $400\gamma(0)$, monotone estimators lose little more efficiency and get a somewhat larger bias, as expected, see Figure 6. Furthermore the filter methods and GK approaches gain efficiency compared to smaller outlier variances. For the latter this seems at first to be surprising since GK estimators are not known to be redescending, but it is consistent with the multivariate setting. The influence function derived in Ref 61 tends to zero along the axes at least in the elliptical model. This means that a small fraction of outliers in only one dimension will have nearly no influence on the estimation as long as the outliers are large. In the case of isolated outliers we expect only one variable of each pair of observations to be contaminated and therefore a redescending behavior of the GK methods.

The estimators generally behave better for models with small absolute autocorrelations. This is not surprising, since the bias effect is more limited there. In models with rather small absolute autocorrelations other robust estimators like RA and GK approaches outperform the Filter-acf, which seems to behave especially well if the autocorrelations have large absolute values.

Patchy (consecutive) outliers increase the sample acf at small time lags toward one. For our simulations we add the same value ω generated from a $N(0, a^2 \cdot \gamma(0))$ distributed random variable to successive observations at times $\{51, \ldots, 50 + n_0\}$. This resembles a temporarily level shift of the same height. We first look at the AR[0]-model with an outlier variance $\gamma(0)100$. Again the estimation by the empirical acf is useless already for $n_0 = 5$ outliers, see Figure 7. Robust estimators can cope much better with this situation and rarely preserve less than half of their efficiency, reaching values between 3.5 and 22 times the

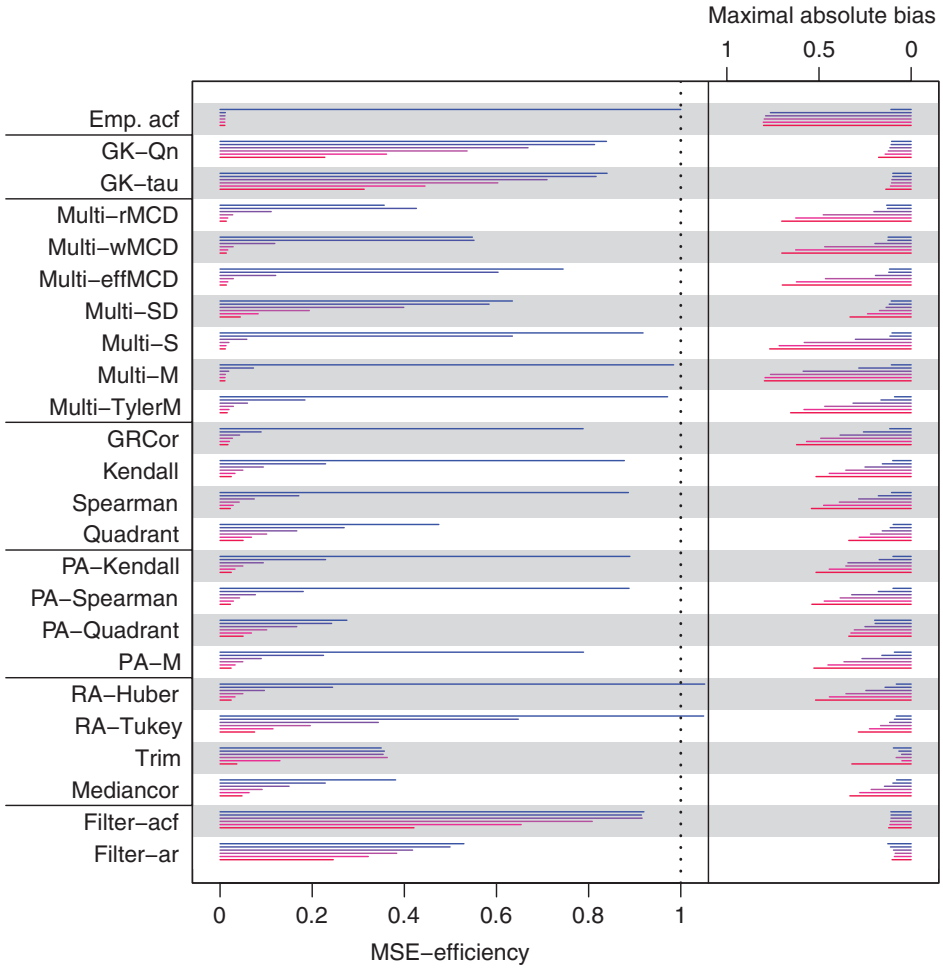


FIGURE 6 | Efficiency (left) and bias (right) for a contaminated AR[0.8] model with $n = 100$ and $n_0 = 0, 5, 10, 15, 20, 25$ (from top to bottom in each panel) isolated outliers and $a = 20$.

efficiency of the empirical acf in our experiments. Estimators based on SD and MCD and to some extent also the multivariate S-estimator perform well. Even for large amounts of outliers they are little biased and they lose only little efficiency. Different kinds of reweighting which boost efficiency in the clean model do not significantly increase the bias or vulnerability to outliers and should be preferred. Recall that GK approaches are not redescending under patchy outliers as can be seen in Figure 4.

If the acf attains large positive values, a few consecutive outliers can even improve the estimation by canceling the small sample bias. All estimators behave better compared to the AR[0] model in these cases, with rank and RA-estimators improving most. Patchy outliers seem to have the largest impact if the acf contains values close to −1. We observe the largest maximal absolute biases and the smallest minimal efficiencies for the AR[−0.4] and AR[−0.8] models. The only exceptions are the filter methods which perform even better than in the AR[0] case. This agrees with higher efficiencies for larger absolute correlations in the absence of contamination.

Nonlinear Models

As suggested by a referee, we also consider nonlinear models, specifically GARCH models, introduced by Bollerslev.[62] They allow the variance of the process to depend on the past observations and are popular for modeling financial time series (see for example Refs 63,64). A GARCH process $(X_t)_{t\in\mathbb{Z}}$ of order $(1,1)$ is defined as

$$X_t = \sigma_t a_t \quad \text{with} \quad \sigma_t^2 = \sigma_0^2 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

with parameters $\sigma_0, \alpha_1, \beta_1 \in \mathbb{R}_+$ and innovations $(a_t)_{t\in\mathbb{Z}}$ usually forming Gaussian or t-distributed white noise. In our simulations we consider a process of order (1,1) with $\sigma_0 = 0.05$, $\alpha_1 = 0.1$, $\beta_1 = 0.85$ and standard normal innovations $(a_t)_{t\in\mathbb{Z}}$ (abbreviated as

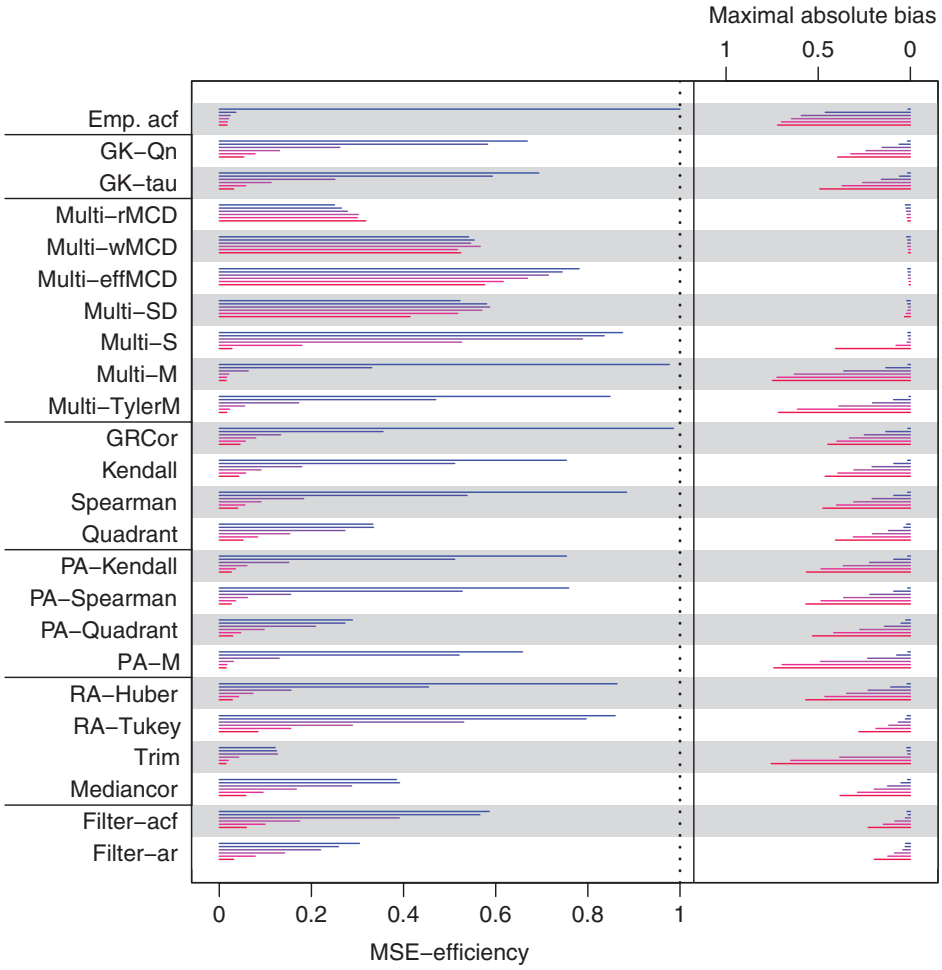


FIGURE 7 | Efficiency (left) and bias (right) for a contaminated AR[0] model with $n = 100$ and outlier patches of length $n_0 = 0, 5, 10, 15, 20, 25$ and $a = 10$.

GARCH[0.05,0.1,0.85]) which is a realistic parameter setting (see Ref 65). Results under patchy outliers can be seen in Figure 8. In the clean model robust procedures are a little more efficient than in the linear case, which might be due to the heavy tails of the marginal distribution of GARCH processes.[66] Under patchy and also under isolated outliers the results (not shown here) are comparable to the results for linear processes.

For modeling GARCH processes the autocorrelations of the squared process $(X_t^2)_{t\in\mathbb{Z}}$ are also of interest (see Ref 67). Since estimation of the acf of $(X_t^2)_{t\in\mathbb{Z}}$ is harder, we choose a time series of length $n = 1000$. It turns out that most estimators are substantially biased, which might be a result of the skewed distribution of the process $(X_t^2)_{t\in\mathbb{Z}}$. Using parameters $\alpha_1 \in \{0, 0.1, 0.2\}$ and $\beta_1 \in \{0, 0.5, 0.7\}$ we find that in addition to the empirical acf only the RA approach with the Huber function and the median correlation yield an acceptable bias.

Positive-Semidefiniteness

We have mentioned the problem of positive-semidefiniteness repeatedly. Our simulations reveal that this is mainly a problem of little efficient estimators like quadrant correlation and the 50% trimming (median) approach. We never noticed problems for multivariate approaches except for the raw MCD, which occasionally produces indefinite estimations if the model is close to being nonstationary. It turns out that consistency corrections for the approaches based on univariate transformations often destroy definiteness. Whereas the difference between the original and the enforced positive-semidefinite estimation is negligible for the RA-estimators, we observed changes up to 0.08 for trimmed and median based correlation. There can be even greater discrepancies for the Filter-AR estimator, which might be caused by some instability of our implementation of this procedure. We rarely noticed indefinite estimations by the variance based approaches. Enforcing

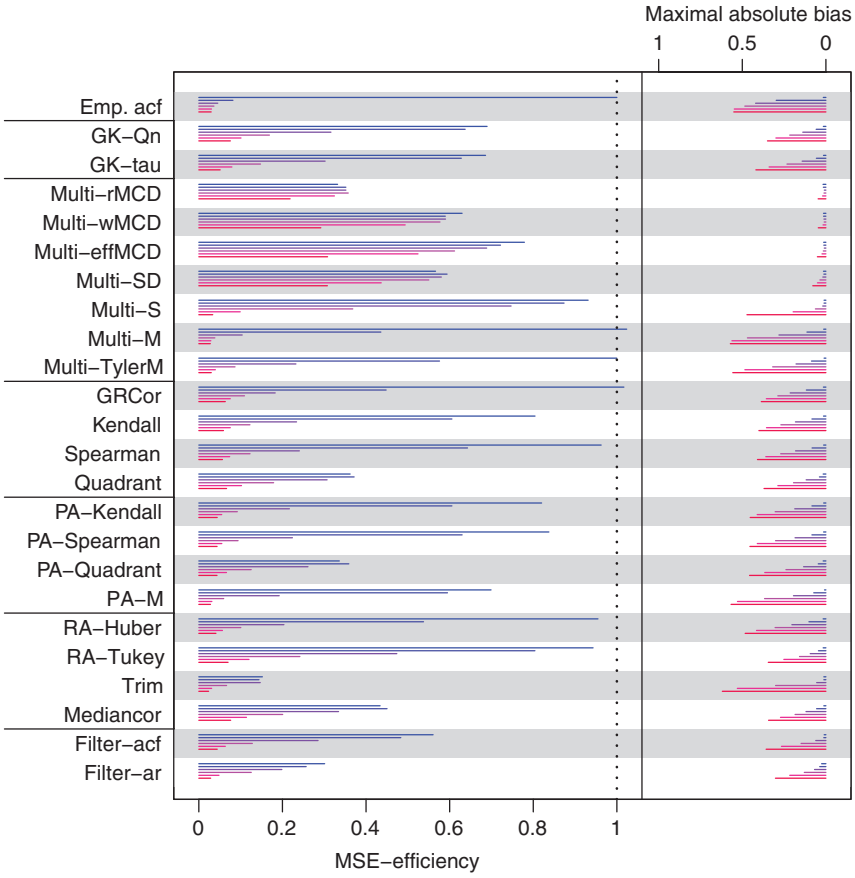


FIGURE 8 | Efficiency under an outlier patch of length $n_0 = 0, 5, 10, 15, 20, 25$ (from top to bottom in each panel) under a GARCH[0.05,0.85,0.1] model with n = 100 and $a = 5$.

positive-semidefiniteness increases the efficiency of trimmed estimators slightly.

CONCLUSION

Many of the proposals for robust autocorrelation estimation are borrowed from the usual correlation estimation applied to all pairs of observations (X_t, X_{t+h}) at a certain time lag h, with the intention of carrying over good robustness properties and high efficiency under normality to the time series context. A problem arising there is that every outlier can enter two pairs of observations, so that the number of contaminated pairs can be up to twice the number of outliers. This problem does not arise for estimators which filter the time series before the acf estimation. However, these do not respect the serial dependence structure like the RA-estimators or they are computationally heavy like the robust filter algorithm. There is a great interest in robust time series analysis nowadays. In the late review process of this paper we have become aware of another work[68] on autocorrelation estimation based on the idea of estimating $\rho(h)$ by regressing X_{t+h} robustly on X_t. Extra manipulations are needed to guarantee that such estimates are positive-semidefinite and do not exceed 1. Moreover, there are many candidate robust regression techniques available, so that a careful inspection of this proposal would have been beyond the scope of this paper. Our simulation study confirms that even a small fraction of contamination can make the empirical acf useless. The robust filter algorithms yield good results even in case of many isolated or patchy outliers, but have a lower efficiency if there is little serial correlation. Estimation based on a reweighted MCD is favorable, if there are patchy outliers. The approach based on the Stahel-Donoho estimator means a good compromise, but it is computationally demanding. If one looks for a relatively quick estimator, the approach based on robust variances seems to be a good choice, since they also generally yield good results. A possible lack of positive-semidefiniteness can easily be fixed by a projection algorithm. Our simulation results are based on additive outliers of random size ω and therefore represent a kind of overall performance for different outlier sizes. In simulations not reported here we also consider other outlier scenarios with fixed outlier sizes. The worst case biases and efficiencies are generally worse there than those presented here. Nevertheless, the results are qualitatively rather similar.

It needs to be kept in mind that in the simulations reported here we focus on the case of innovations from a contaminated Gaussian or at least continuous-symmetric distribution. Results look different, e.g., for count time series as reported in Ref 69 where rank-based estimators performed rather well.

ACKNOWLEDGMENTS

We are very grateful to the referees and the review editor for their careful reading and fruitful remarks which lead to substantial improvements of this paper. The authors were supported by the Collaborative Research Center 823 of the German Research Foundation (project C3: 'Analysis of structural change in time series analysis').

REFERENCES

1. Brockwell PJ. Autocovariance. *WIREs Comput Stat* 2009, 1:187–198.
2. Box GEP, Jenkins GM, Reinsel GC. *Time Series Analysis: Forecasting and Control*. 3rd ed. Englewood Cliff, NJ: Prentice Hall; 1994.
3. Brockwell PJ, Davis RA. *Time Series: Theory and Methods*. 2nd ed. New York: Springer; 2006.
4. Wei WWS. *Time Series Analysis: Univariate and Multivariate Methods*. Redwood City: Addison-Wesley; 1990.
5. Vecchia AV, Ballerini R. Testing for periodic autocorrelations in seasonal time series data. *Biometrika* 1991, 78:53–63.
6. Caiado J, Crato N, Peña D. A periodogram-based metric for time series classification. *Comput Stat Data Anal* 2006, 50:2668–2684.
7. Chan W-S. A note on time series model specification in the presence of outliers. *J Appl Stat* 1992, 19:117–124.
8. Deutsch SJ, Richards JE, Swain JJ. Effects of a single outlier on ARMA identification. *Commun Stat Theor Methods* 1990, 19:2207–2227.
9. Maronna RA, Martin RD, Yohai VJ. *Robust Statistics*. Chichester: John Wiley & Sons; 2006.
10. Schlittgen R, Streitberg BHJ. *Zeitreihenanalyse*. München: Oldenbourg; 2001, 414–427.
11. Brockwell PJ. Autoregressive processes. *WIREs Comput Stat* 2011, 3:316–331.
12. Ramsey FL. Characterization of the partial autocorrelation function. *Ann Stat* 1974, 2:1296–1301.
13. Masarotto G. Robust identification of autoregressive moving average models. *Appl Stat* 1987, 36:214–220.
14. Makhoul J. Lattice methods in spectral estimation. *Appl Time Ser Anal* 1981, 2:301–326.
15. Chan W-S, Wei WWS. A comparison of some estimators of time series autocorrelations. *Comput Stat Data Anal* 1992, 14:149–163.
16. Chakhchoukh Y. A new robust estimation method for ARMA models. *IEEE Trans Signal Process* 2010, 58:3512–3522.
17. Bustos OH, Yohai VJ. Robust estimates for ARMA models. *J Am Stat Assoc* 1986, 81:155–168.
18. Morgenthaler S. Robustness. *WIREs Comput Stat* 2011, 3:85–94.
19. Garel B, Hallin M. Rank-based autoregressive order identification. *J Am Stat Assoc* 1999, 94:1357–1371.
20. Ferretti NE, Kelmansky DM, Yohai VJ. Estimators based on ranks for ARMA models. *Commun Stat Theor Methods* 1991, 20:3879–3907.
21. Boudt K, Cornelissen J, Croux C. The gaussian rank correlation estimator: robustness properties. *Stat Comput* 2012, 22:471–483.
22. Croux C, Dehon C. Influence functions of the spearman and kendall correlation measures. *Stat Methods Appl* 2010, 19:497–515.
23. Mottonen J, Koivunen V, Oja H. Robust autocovariance estimation based on sign and rank correlation coefficients. In: *Proceedings of the IEEE Signal Processing Workshop on Higher-Order Statistics*, IEEE, 187–190, 1999.
24. Gather U, Bauer M, Fried R. The identification of multiple outliers in online monitoring data. *Estadistica* 2002, 54:289–338.
25. Taskinen S, Croux C, Kankainen A, Ollila E, Oja H. Influence functions and efficiencies of the canonical correlation and vector estimates based on scatter and shape matrices. *J Multivar Anal* 2006, 97:359–384.
26. Vakili K, Schmitt E. Finding multivariate outliers with fastPCS. *Comput Stat Data Anal* 2014, 69:54–66.
27. Maronna RA. Robust M-estimators of multivariate location and scatter. *Ann Stat* 1976, 4:51–67.
28. Tyler DE. A distribution-free M-estimator of multivariate scatter. *Ann Stat* 1987, 15:234–251.
29. Hubert M, Debruyne M. Breakdown value. *WIREs Comput Stat* 2009, 1:296–302.
30. Davies PL. Asymptotic behaviour of S-estimates of multivariate location parameters and dispersion matrices. *Ann Stat* 1987, 15:1269–1292.
31. Van Aelst S, Rousseeuw P. Minimum volume ellipsoid. *WIREs Comput Stat* 2009, 1:71–82.

32. Lopuhaa HP. On the relation between S-estimators and M-estimators of multivariate location and covariance. *Ann Stat* 1989, 17:1662–1683.
33. Salibian-Barrera M, Yohai VJ. A fast algorithm for S-regression estimates. *J Comput Graph Stat* 2006, 15.
34. Hubert M, Debruyne M. Minimum covariance determinant. *WIREs Comput Stat* 2010, 2:36–43.
35. Rousseeuw PJ. Multivariate estimation with high breakdown point. *Math Stat Appl B* 1985, 8:283–297.
36. Rousseeuw PJ, Van Driessen K. A fast algorithm for the minimum covariance determinant estimator. *Technometrics* 1999, 41:212–223.
37. Croux C, Haesbroeck G. Influence function and efficiency of the minimum covariance determinant scatter matrix estimator. *J Multivar Anal* 1999, 71:161–190.
38. Gervini D. A robust and efficient adaptive reweighted estimator of multivariate location and scatter. *J Multivar Anal* 2003, 84:116–144.
39. Hadi AS, Imon AHM, Werner M. Detection of outliers. *WIREs Comput Stat* 2009, 1:57–70.
40. Donoho DL. Breakdown properties of multivariate location estimators. Technical report, Harvard University, Boston; 1982. Available at: http://statweb.stanford.edu/%7Edonoho/Reports/Oldies/BPMLE.pdf.
41. Stahel WA. Robust estimation: infinitesimal optimality and covariance matrix estimators. PhD Thesis, ETH, Zurich, Switzerland, 1981.
42. Gnanadesikan R, Kettenring JR. Robust estimates, residuals, and outlier detection with multiresponse data. *Biometrika* 1972, 28:81–124.
43. Ma Y, Genton MG. Highly robust estimation of the autocovariance function. *J Time Ser Anal* 2000, 21:663–684.
44. Rousseeuw PJ, Croux C. Alternatives to the median absolute deviation. *J Am Stat Assoc* 1993, 88:1273–1283.
45. Maronna RA, Zamar RH. Robust estimates of location and dispersion for high-dimensional datasets. *Technometrics* 2002, 44:307–317.
46. Masreliez C. Approximate non-Gaussian filtering with linear state and observation relations. *IEEE Trans Automat Control* 1975, 20:107–110.
47. Martin RD, Thomson DJ. Robust-resistant spectrum estimation. *Proc IEEE* 1982, 70:1097–1115.
48. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2012.
49. Todorov V, Filzmoser P. An object-oriented framework for robust multivariate analysis. *J Stat Softw* 2009, 32:1–47.
50. Nordhausen K, Sirkia S, Oja H, Tyler DE. *ICSNP: Tools for Multivariate Nonparametrics*. R package version 1.0-9, 2012.
51. Sirkia S, Miettinen J, Nordhausen K, Oja H, Taskinen S. *SpatialNP: Multivariate nonparametric methods based on spatial signs and ranks*. R package version 1.1-0, 2012.
52. Rousseeuw P, Croux C, Todorov V, Ruckstuhl A, Salibian-Barrera M, Verbeke T, Koller M, Maechler M. *Robustbase: Basic Robust Statistics*. R package version 0.90-2, 2014.
53. Brent RP. *Algorithms for Minimization Without Derivatives*. Englewood Cliffs, NJ: Prentice-Hall; 1973.
54. Al-Homidan S. Semidefinite and second-order cone optimization approach for the toeplitz matrix approximation problem. *J Numer Math* 2006, 14:1–15.
55. Davis R, Resnick S. Limit theory for the sample covariance and correlation functions of moving averages. *Ann Stat* 1986, 14:533–558.
56. Loretan M, Phillips PCB. Testing the covariance stationarity of heavy-tailed time series: An overview of the theory with applications to several financial datasets. *J Empir Financ* 1994, 1:211–248.
57. Politis DN. Financial time series. *WIREs Comput Stat* 2009, 1:157–166.
58. Rojo J. Heavy-tailed densities. *WIREs Comput Stat* 2013, 5:30–40.
59. Davis RA, Mikosch T. The sample autocorrelations of heavy-tailed processes with applications to ARCH. *Ann Stat* 1998, 26:2049–2080.
60. Fox AJ. Outliers in time series. *J R Stat Soc Ser B Methodol* 1972, 34:350–363.
61. Ma Y, Genton MG. Highly robust estimation of dispersion matrices. *J Multivar Anal* 2001, 78:11–36.
62. Bollerslev T. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 1986, 31:307–327.
63. Bollerslev T, Chou RY, Kroner KF. ARCH modeling in finance: a review of the theory and empirical evidence. *J Econom* 1992, 52:5–59.
64. Duan J-C. The GARCH option pricing model. *Math Finance* 1995, 5:13–32.
65. Lamoureux CG, Lastrapes WD. Heteroskedasticity in stock return data: volume versus GARCH effects. *J Financ* 1990, 45:221–229.
66. Basrak B, Davis RA, Mikosch T. Regular variation of GARCH processes. *Stochastic Processes and their Applications* 2002, 99:95–115.
67. Baillie RT, Chung H. Estimation of GARCH models from the autocorrelations of the squares of a process. *J Time Ser Anal* 2001, 22:631–650.
68. Chang CC, Politis DN. Robust autocorrelation estimation. *Journal of Computational and Graphical Statistics*, (just-accepted), 2014.
69. Fried R, Liboschik T, Elsaied H, Kitromilidou S, Fokianos K. On outliers and interventions in count time series following GLMs. *Aust J Stat* 2014, 43:181–193.