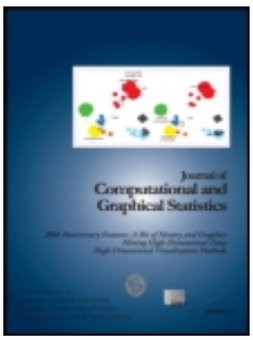


Taylor & Francis
Taylor & Francis Group

Journal of Computational and Graphical Statistics

ISSN: 1061-8600 (Print) 1537-2715 (Online) Journal homepage: http://amstat.tandfonline.com/loi/ucgs20

Robust Autocorrelation Estimation

Christopher C. Chang & Dimitris N. Politis

To cite this article: Christopher C. Chang & Dimitris N. Politis (2016) Robust Autocorrelation Estimation, Journal of Computational and Graphical Statistics, 25:1, 144-166, DOI: 10.1080/10618600.2014.969431

To link to this article: http://dx.doi.org/10.1080/10618600.2014.969431

Accepted author version posted online: 16 Oct 2014.
Published online: 09 Mar 2016.

Robust Autocorrelation Estimation

Christopher C. CHANG and Dimitris N. POLITIS

In this article, we introduce a new class of robust autocorrelation estimators based on interpreting the sample autocorrelation function as a linear regression. We investigate the efficiency and robustness properties of the estimators that result from employing three common robust regression techniques. We discuss the construction of robust autocovariance and positive definite autocorrelation estimates, and their application to AR model fitting. We perform simulation studies with various outlier configurations to compare the different estimators.

Key Words: Regression; Robustness.

1. INTRODUCTION

The estimation of the autocorrelation function plays a central role in time series analysis. For example, when a time series is modeled as an autoregressive (AR) process, the model coefficient estimates are straightforward functions of the estimated autocorrelations Brockwell and Davis (1991), and formulation of autoregressive moving average (ARMA) models for more complex time series can also be done via inspection of the autocorrelation and partial autocorrelation functions.

Recall that the autocovariance function (acvf for short) of a wide-sense stationary time series $\{X_t\}$ is defined as $\gamma(h) := E[(X_{t+h} - \mu)(X_t - \mu)]$ where $\mu := E[X_t]$; similarly, the autocorrelation function (acf for short) is $\rho(h) := \gamma(h)/\gamma(0)$. Given data $X_1, \ldots, X_n$, the classical estimator of the acvf is the sample acvf:

$$\hat{\gamma}(h) := n^{-1} \sum_{j=1}^{n-h} (X_{j+h} - \bar{X})(X_j - \bar{X})$$

for $|h| < n$, where $\bar{X} := n^{-1} \sum_{j=1}^{n} X_j$. The classical estimator of the acf is simply $\hat{\rho}(h) := \hat{\gamma}(h)/\hat{\gamma}(0)$.

However, the classical estimator is not robust; just one contaminated point is enough to corrupt it and mask the real dependence structure. Since it is not uncommon for 10% or more of measured time series values to be outliers (Hampel 1973), this is a serious problem.

Christopher C. Chang and Dimitris N. Politis are with the Department of Mathematics, University of California at San Diego, La Jolla, CA 92093 (E-mails: *chrchang@alumni.caltech.edu*; *dpolitis@ucsd.edu*).

Journal of Computational and Graphical Statistics, Volume 25, Number 1, Pages 144–166
DOI: 10.1080/10618600.2014.969431

To address it, several robust autocorrelation estimators have been proposed (Denby and Martin 1979; Künsch 1983; Ma and Genton 2000; Maronna, Martin, and Yohai 2006). Due to the limitations of older computers, these techniques were not widely adopted; instead, explicit detection and removal of outliers was typically employed. However, today's hardware is powerful enough to support robust estimation in most contexts, and it is far from clear that classical estimation with explicit outlier elimination is more effective in practice than a well-chosen robust estimator.

In this article, we propose a new class of robust autocorrelation estimators, based on constructing an *autoregressive* (AR) scatterplot, and applying robust regression to it.

The remainder of this article is structured as follows: In Section 2, we introduce the new class of robust autocorrelation estimators. Next, in Section 3, we analyze the estimators that result from using three common robust regression techniques, and compare their performance to that of the sample acf. Then, in Sections 4 and 5, we discuss the derivation of autocovariance and positive definite autocorrelation estimates from our initial estimator. We apply our method to robust AR model fitting in Section 6. Finally, we present the results of our simulation study (including one real data example) in Section 7.

2. SCATTERPLOT-BASED ROBUST AUTOCORRELATION ESTIMATION

Assume we have data $X_1, \ldots, X_n$ from the wide-sense stationary time series $\{X_t\}$ discussed in the Introduction. Fix a lag $h < n$ where $h \in \mathbb{Z}^+$, and consider the scatterplot associated with the pairs $\{(X_t - \bar{X}, X_{t+h} - \bar{X})$, for $t \in \{1, \ldots, n-h\}\}$; see (1) for an example with $h = 1$. (Here, and in what follows, we use the R convention of scatterplots, i.e., the pairs (x_i, y_i) correspond to a regression of y on x.)

If the time series $\{X_t\}$ satisfies the causal AR(p) model

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + z_t \tag{1}$$

with z_t being iid $(0, \sigma^2)$, then

$$E[X_{t+h} - \mu | X_t] = (X_t - \mu)\rho(h);$$

this includes the case when $\{X_t\}$ is Gaussian. In the absence of a causal AR(p)—or AR(∞)—model with respect to iid errors, the above equation is modified to read

$$\bar{E}[X_{t+h} - \mu | X_t] = (X_t - \mu)\rho(h)$$

provided the spectral density of $\{X_t\}$ exists and is strictly positive. In the above, $\bar{E}[Y|X]$ denotes the orthogonal projection of Y onto the linear space spanned by X; see, for example, Kreiss, Paparoditis, and Politis (2011).

In either case, it is apparent that the point $(X_t - \bar{X}, X_{t+h} - \bar{X})$ on the scatterplot should tend to be close to the line $y = \rho(h)x$, and one would expect the *slope* of a regression line through the points to be a reasonable estimate of the autocorrelation $\rho(h)$. This works well, as Figure 1 shows. Indeed, it is well known that the (ordinary) least squares (OLS) fit is almost identical to the sample acf for $\frac{h}{n}$ small.

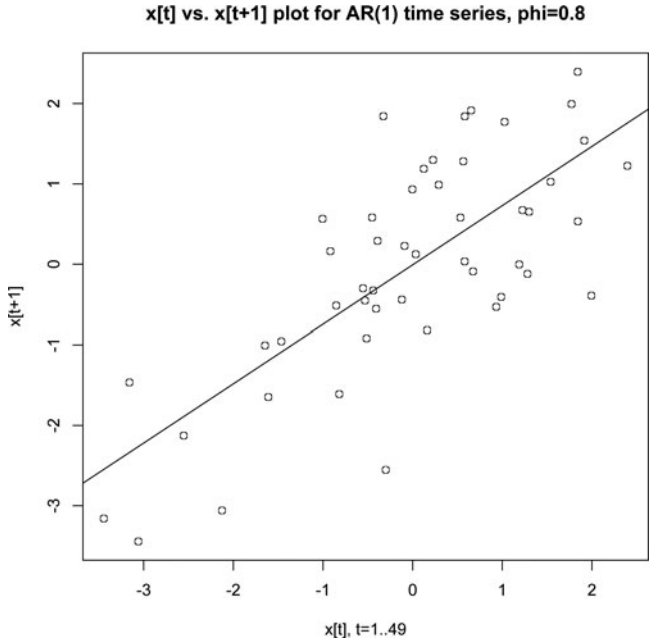


Figure 1. Scatterplot of (X_t, X_{t+1}) for a realization of length 50 from the AR(1) time series $X_t = 0.8X_{t-1} + Z_t$, Z_t iid $N(0, 1)$. Regression line is $y = 0.7372x - 0.00813$.

To elaborate, if the points on the scatterplot are denoted (x_i, y_i), that is, letting $y_i = X_{i+h}$ and $x_i = X_i$, we have

$$\begin{aligned}\hat{\rho}_{\mathrm{OLS}}(h) &= \frac{\sum_{j=1}^{n-h}(x_j - \bar{x})(y_j - \bar{y})}{\sum_{j=1}^{n-h}(x_j - \bar{x})^2} \\ &= \frac{\sum_{j=1}^{n-h}(X_{j+h} - \bar{X}_{(h+1)\ldots n})(X_j - \bar{X}_{1\ldots(n-h)})}{\sum_{j=1}^{n-h}(X_j - \bar{X}_{1\ldots(n-h)})^2} \\ &\approx \frac{\sum_{j=1}^{n-h}(X_{j+h} - \bar{X})(X_j - \bar{X})}{\frac{n-h}{n}\sum_{j=1}^{n}(X_j - \bar{X})^2} \\ &= \frac{n}{n-h}\hat{\rho}(h),\end{aligned}$$

where the notation $\bar{x}_{a\ldots b} := (b - a + 1)^{-1}\sum_{j=a}^{b} x_j$ and $\bar{x} := \bar{x}_{1\ldots n}$ has been used.

The additional $\frac{n}{n-h}$ factor is expected, since the regression slope is supposed to be an unbiased estimator while the sample acf is biased by construction. The only other difference between $\hat{\rho}_{\mathrm{OLS}}(h)$ and $\hat{\rho}(h)$ is the inclusion/exclusion of the first and last time series points in computing sample mean and variance; the impact of that is negligible.

Since $\hat{\rho}_{\mathrm{OLS}}(h)$ is based on simple OLS regression, the implication is that if we run a robust linear regression on the pairs $\{(X_t, X_{t+h})\}$, we should get a robust estimate of autocorrelation. We, therefore, define $\hat{\rho}_{\mathrm{ROBUST}}(h)$ to be the estimator of slope β_1 in the

straight line regression

$$X_t = \beta_0 + \beta_1 X_{t-h} + \varepsilon \tag{2}$$

using *any* robust method to fit the regression.

Remark 2.1. Our robust estimator $\hat{\rho}_{\text{ROBUST}}(h)$ is not guaranteed to be in the range $[-1, 1]$. As an example, consider the $n = 3$ time series dataset $\{1, 2, 0\}$; for $h = 1$, the slope of the straight line regression equals -2. If an estimator of $\rho(h)$ for a fixed h is sought, then an easy fix is to "clip" the estimator $\hat{\rho}_{\text{ROBUST}}(h)$ to the values -1 and 1. In other words, estimate $\rho(h)$ by $\hat{\rho}_{\text{ROBUST}}(h)$ when the latter is in the interval $[-1, 1]$; else, estimate $\rho(h)$ by 1 when $\hat{\rho}_{\text{ROBUST}}(h) > 1$, and by -1 when $\hat{\rho}_{\text{ROBUST}}(h) < -1$. On the other hand, if one wishes to estimate $\rho(k)$ for many lags, for example, $k = 1, \ldots, h$, a more elaborate fix is required; in that respect, see Section 5, and in particular, Remark 5.1.
However, these fixes are frequently unnecessary. The simulations, discussed later, just involve the basic robust estimator without either fix applied.

In what follows, we investigate in more detail three possibilities for the aforementioned robust regression fitting that result in three different robust acf estimators. Note, however, that other robust regression techniques could alternatively be used in the context of producing $\hat{\rho}_{\text{ROBUST}}(h)$.

1. $\hat{\rho}_{L1}$. Recall that a residual of a linear regression is the vertical distance between the point (x_i, y_i) and the regression line; that is, given the regression line $y = \beta_1 x + \beta_0$, we have $r_i(\beta) = (\beta_1 x_i + \beta_0) - y_i$. The simplest robust regression technique, L1 regression, selects β to minimize $\sum_i |r_i(\beta)|$ instead of the usual $\sum_i r_i(\beta)^2$; the effect is to find a "median regression line."

2. $\hat{\rho}_{\text{LTS}}$. Least trimmed squares regression, or LTS for short, takes a different approach: instead of changing the pointwise loss function, we use the usual squared residuals but throw the largest values out of the sum. More precisely, define $|r|_{(1)} \le \cdots \le |r|_{(n-h)}$ to be the ordered residual absolute values. α-trimmed squares minimizes

$$\hat{\sigma} := \left(\sum_{j=1}^{\lceil (1-\alpha)(n-h) \rceil} |r|_{(j)}^2 \right)^{1/2}.$$

   We look at α-trimmed squares for $\alpha = \frac{1}{2}$ (i.e., we minimize the sum of the smallest 50% of the residuals).

3. $\hat{\rho}_{\text{MM}}$. An M-estimate (Huber 1973) minimizes

$$L(\beta) := \sum_{i=1}^{n} \ell \left(\frac{r_i(\beta)}{\hat{\sigma}} \right)$$

   for some pointwise loss function ℓ, where $\hat{\sigma}$ is an estimate of the scale of the errors. It is efficient, but not resistant to outliers in the x values. A "redescending" M-estimate

uses a loss function that decreases to zero at the tails (as opposed to a monotone increasing loss function).

In contrast, an S-estimate (S for "scale") minimizes a robust estimate of the scale of the residuals:

$$\hat{\beta} := \underset{\beta}{\operatorname{argmin}}\, \hat{\sigma}(\mathbf{r}(\beta)),$$

where $\mathbf{r}(\beta)$ denotes the vector of residuals and $\hat{\sigma}$ satisfies

$$\frac{1}{n}\sum_{j=1}^{n-h} \ell\left(\frac{r_j}{\hat{\sigma}}\right) = \delta.$$

(δ is usually chosen to be $1/2$.) It has superior robustness, but is inefficient.

MM-estimates, pioneered by Yohai (1987), combine these two techniques in a way intended to retain the robustness of S-estimation while gaining the asymptotic efficiency of M-estimation. Specifically, an initial robust-but-inefficient estimate $\hat{\beta}_0$ is computed, then a scale M-estimate of the residuals, and finally the iteratively reweighted least squares algorithm is used to identify a nearby $\hat{\beta}$ that satisfies the redescending M-estimate equation.

For further discussion on the above three robust regression techniques, see Maronna, Martin, and Yohai (2006).

3. THEORETICAL PROPERTIES

3.1 GENERAL

We focus our attention on normal efficiency and two measures of robustness (breakdown point and influence function). Relative *normal efficiency* is the ratio between the asymptotic variance of the classical estimator and that of another estimator under consideration, assuming Gaussian residuals and no contamination. This is a measure of the price we are paying for any robustness gains. The *breakdown point* (BP) is the asymptotic fraction of points that can be contaminated without entirely masking the original relation. Now, in the case of time series modeled by ARMA (autoregressive moving average) processes, we distinguish two types of outliers (Denby and Martin 1979):

1. *Innovation outliers* affect all subsequent observations, and can be observed in a pure ARMA process with a heavy-tailed innovation distribution.
2. *Additive outliers* or *replacement outliers* that exist outside the ARMA process and do not affect other observations. For second-order stationary data, the difference between the two notions is minimal—a replacement outlier functions like a slightly variable additive outlier—so for brevity we just concern ourselves with additive outliers.

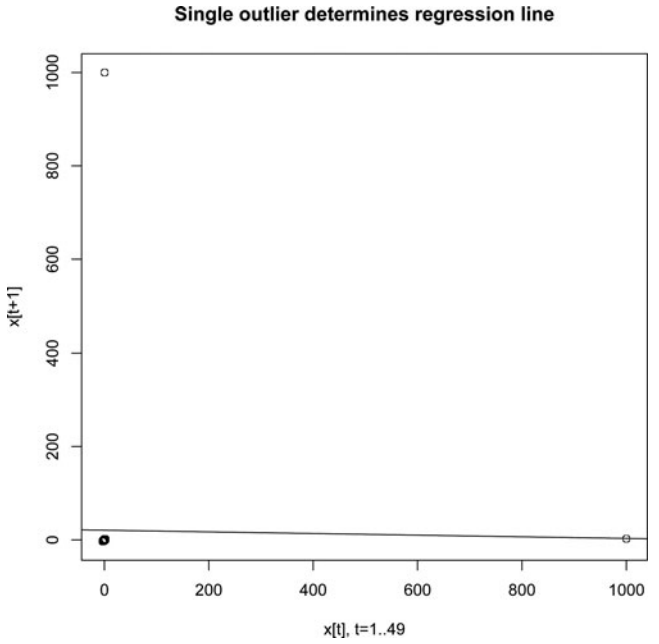


Figure 2. Degenerate regression line from 50 $N(0,1)$ points contaminated by one outlier at 1000.

For additive outliers, the classical autocorrelation estimator has a breakdown point of zero since a single very large outlier is enough to force the estimate to a neighborhood of $\frac{-1}{n-h}$; see Figure 2 for an illustration. Since one additive outlier influences the position of at most two points in the regression, our robust autocorrelation estimators will exhibit BPs at least half that of the robust regression techniques they are built on; see Ma and Genton (2000) for a more exhaustive discussion on "temporal breakdown point."

The impact of an innovation outlier on the regression line varies. For instance, only one point is moved off the regression line in the AR(1) case, but three points are affected in the MA(1) case. So in the former scenario, our robust autocorrelation estimators can be expected to fully inherit the BPs of the robust regressors with respect to innovation outliers, but we cannot expect as much reliability with MA models; for more details, see Genton and Lucas (2003, 2005) who analyzed in detail how dependence can adversely affect the BP.

Perhaps surprisingly, infinite variance symmetric alpha-stable innovation distributions result in a faster sample acf convergence rate than the finite variance innovation case (Davis and Mikosch 2000).

Next, the *influence function* (IF) describes the impact on the autocorrelation estimate $\hat{\rho}_{\text{ROBUST}}$ of adding an infinitesimal probability of an innovation or additive outlier. Given an innovation or residual distribution F, let $\hat{\rho}_{\text{ROBUST}}(F)$ denote the value the robust autocorrelation estimate converges to, as the number of observations increases without bound. Then the influence function can be described as

$$IF(x, \hat{\rho}_{\text{ROBUST}}, F) = \lim_{\epsilon \to 0^+} \frac{\hat{\rho}_{\text{ROBUST}}((1-\epsilon)F + \epsilon \Delta_x) - \hat{\rho}_{\text{ROBUST}}(F)}{\epsilon}$$

for x such that this limit exists, where Δ_x denotes a probability point mass at x. This is a measure of the asymptotic bias caused by observation contamination (Ma and Genton 2000). For example, the value of the classical estimator's influence function increases without bound as $|x| \to \infty$, since the numerator in the limit converges to a nonzero constant while the denominator goes to zero.

3.2 L1

Because the x-coordinates are not fixed, $\hat{\rho}_{L1}$ does not inherit all the asymptotic robustness advantages normally enjoyed by L1 regression. Any outlier in the middle of the time series appears as both an x- and a y-coordinate, and while L1 regression shrugs off the y outlier, the x outlier point can have an extreme influence on it. Therefore, the BP is zero in the additive outliers case and the influence function increases without bound again. Since, if the underlying process is AR(1), an additive outlier can have an effect similar to that of two adjacent innovation outliers, the same results hold in the innovation outliers case.

3.3 LTS

LTS regression exhibits the highest possible breakdown point (1/2). It is robust with respect to both x- and y-outliers, so $\hat{\rho}_{\text{LTS}}$ retains the 1/2 BP in the AR(1) innovation outliers case and has a BP of at least 1/4 with respect to additive outliers. The influence function decreases to zero at the tails since the probability of mistaking the outlier for a "real" point declines exponentially in n.

It also exhibits the optimal convergence rate, but has a very low normal efficiency of around 7% (Rousseeuw and Leroy 1987); this weakness is clearly visible in our simulations.

3.4 MM

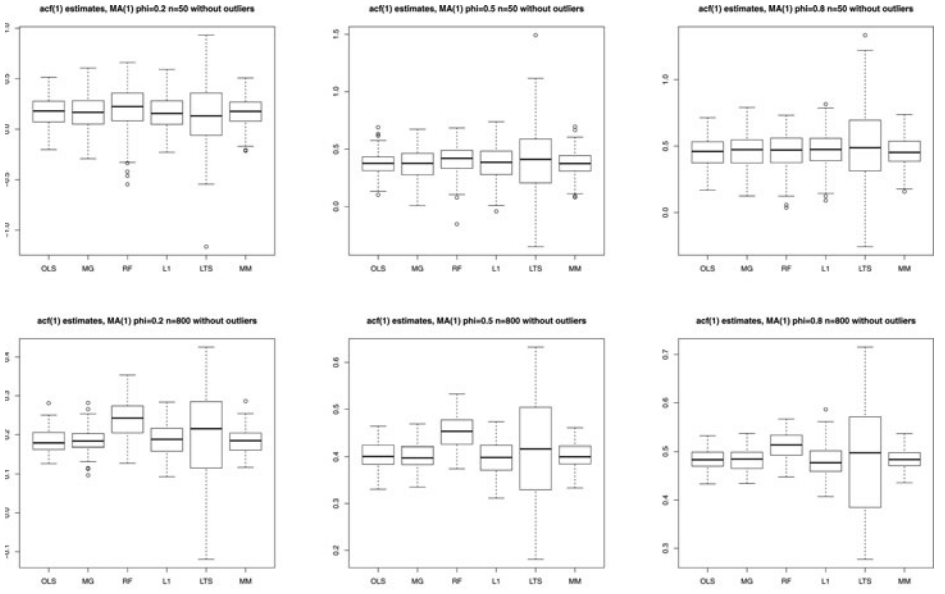
MM-estimates also have an asymptotic breakdown point of 1/2 and are resistant to both x- and y-outliers, so $\hat{\rho}_{\text{MM}}$ has a BP of 1/2 in the innovation outliers case and at least 1/4 in the additive outliers case.

The normal efficiency is actually a user-adjustable parameter. In practice, it it is usually chosen to be between 0.70 and 0.95; aiming for an even higher normal efficiency results in too large a region where the MM-estimate tracks the performance of the classical estimator rather than exhibiting the S-estimate's robustness. We use 0.85 in our simulations.

4. ROBUST AUTOCOVARIANCE ESTIMATION

Our robust acf estimators can be converted into autocovariance estimators via multiplication by a robust estimate of variance. This could proceed as follows:

1. First, obtain a robust estimate of location. Building on Equation (2), from each autocorrelation regression we perform, we can derive an estimate of the process



		MSE					
ϕ	n	OLS	MG	RF	L1	LTS	MM
0.2	50	0.02231	0.03094	0.04245	0.03368	0.10682	0.02423
	800	0.00117	0.00148	0.00475	0.00205	0.01431	0.00128
0.5	50	0.01154	0.01863	0.01539	0.02135	0.08164	0.01274
	800	0.00071	0.00098	0.00410	0.00111	0.01115	0.00072
0.8	50	0.01305	0.01806	0.01846	0.01902	0.08802	0.01396
	800	0.00051	0.00054	0.00141	0.00130	0.01146	0.00049

(a)

Figure 3. Basic MA(1) simulation results (no outliers).

mean μ:

$$X_t - \mu = \beta_1(X_{t-h} - \mu) + \varepsilon, \text{ since this line should have zero intercept}$$

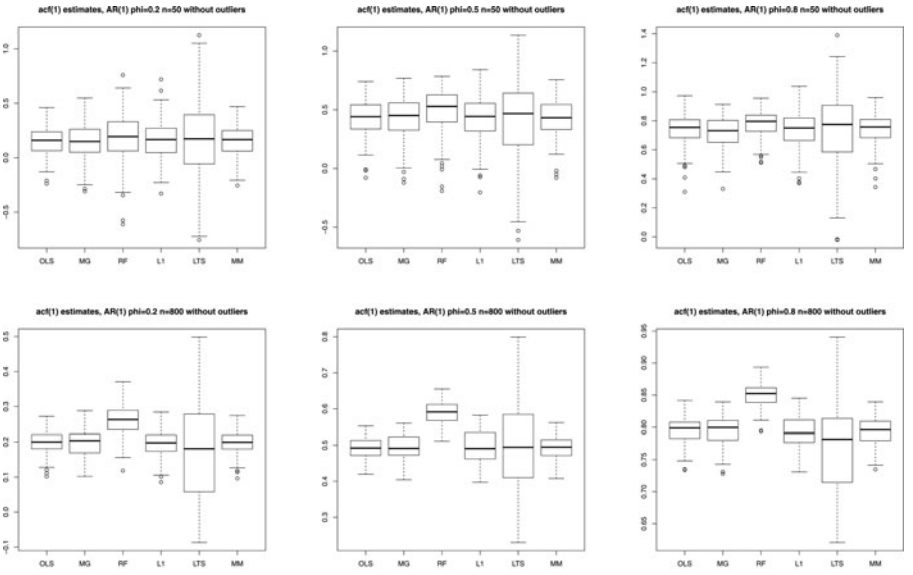
$$X_t = \beta_0 + \beta_1 X_{t-h} + \varepsilon \tag{3}$$

where $\beta_0 = \mu(1 - \beta_1)$ (combining Equations (2) and (3)).

Finally, let $\hat{\mu}_h := \dfrac{\hat{\beta}_0}{1 - \hat{\beta}_1}$,

where $\hat{\beta}_0, \hat{\beta}_1$ are robust estimates of β_0, β_1 in the linear regression (3).

2. Each value of $h > 0$ used in the above step will yield a distinct estimator of location denoted by $\hat{\mu}_h$. We can now use the median of the distinct values $\hat{\mu}_k$ for $k = 1, \ldots,$ (some) p to arrive to a single, better estimator for μ denoted by $\hat{\mu}$. For example, we can use L1 notions and take $\hat{\mu}$ as the median of the values $\{\hat{\mu}_k$ for $k = 1, \ldots, p\}$. Alternatively, LTS regression can be applied to $\hat{\mu}_k$ for $k = 1, \ldots, p$ to boil them down to a single estimate.



		MSE					
ϕ	n	OLS	MG	RF	L1	LTS	MM
0.2	50	0.01883	0.03093	0.04120	0.03020	0.12476	0.01977
	800	0.00138	0.00164	0.00664	0.00211	0.02034	0.00142
0.5	50	0.02628	0.03400	0.03150	0.03551	0.11902	0.02841
	800	0.00110	0.00157	0.00912	0.00212	0.01639	0.00139
0.8	50	0.01379	0.01774	0.00783	0.01844	0.06027	0.01404
	800	0.00062	0.00067	0.00286	0.00063	0.00686	0.00062

(a)

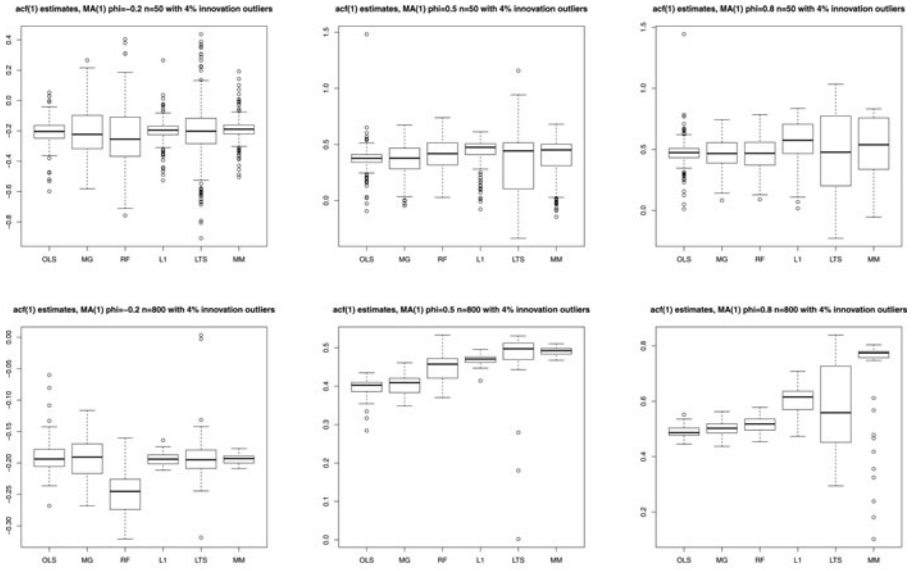
Figure 4. Basic AR(1) simulation results (no outliers).

3. Since $(X_t - \mu)^2 = \gamma(0) + \varepsilon$, we can robustly estimate $\gamma(0)$ by using L1 (or LTS) on the centered sample values $(X_t - \hat{\mu})^2$ for $t = 1, \ldots, n$. For example, let $\hat{\gamma}_{\text{ROBUST}}(0)$ be the median of $(X_t - \hat{\mu})^2$ for $t = 1, \ldots, n$.

4. Finally, we define our robust autocovariance estimator for general h as

$$\hat{\gamma}_{\text{ROBUST}}(h) = \hat{\rho}_{\text{ROBUST}}(h) \cdot \hat{\gamma}_{\text{ROBUST}}(0).$$

Remark 4.1. The estimators $\hat{\rho}_{\text{ROBUST}}(h)$ and $\hat{\gamma}_{\text{ROBUST}}(h)$ are robust *point* estimators of the values $\rho(h)$ and $\gamma(h)$. To construct confidence intervals and hypothesis tests based on these estimators, an approximation to their sampling distribution would be required. Such an approximation appears analytically intractable at the moment without imposing assumptions that are so strong to make the whole setting uninteresting, for example, assuming that the time series $\{X_t\}$ is Gaussian. To avoid such unreliastic assumptions, the practitioner may use a bootstrap approximation to the sampling distribution of the estimators $\hat{\rho}_{\text{ROBUST}}(h)$ and/or $\hat{\gamma}_{\text{ROBUST}}(h)$. Several powerful resampling methods for time series data have been developed in the last 25 years, for example, blocking methods, AR-sieve bootstrap, frequency domain methods, etc.; see the recent review by Kreiss and Paparoditis



		MSE					
ϕ	n	OLS	MG	RF	L1	LTS	MM
-0.2	50	0.00872	0.02375	0.04153	0.00757	0.05108	0.00838
	800	0.00120	0.00103	0.00449	0.00010	0.00238	0.00006
0.5	50	0.01526	0.02190	0.02358	0.01486	0.07528	0.03368
	800	0.00078	0.00083	0.00377	0.00501	0.01336	0.00826
0.8	50	0.01460	0.01763	0.01875	0.03258	0.09201	0.07459
	800	0.00048	0.00089	0.00164	0.01661	0.02725	0.07413

(a)

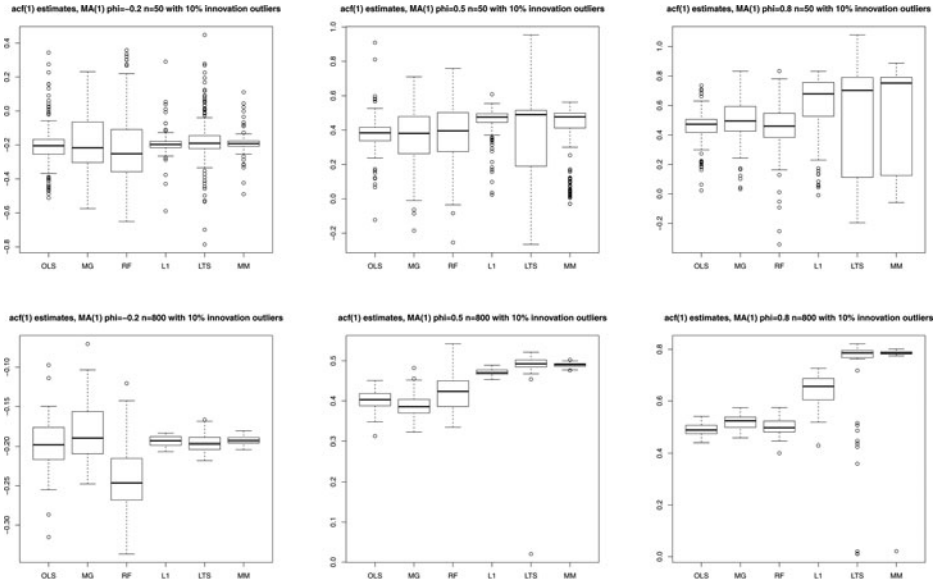
Figure 5. MA(1) with 4% innovation outlier contamination.

(2011) and the references therein. It is unclear at the moment which of these methods would be preferable in the context of $\hat{\rho}_{\text{ROBUST}}(h)$ and $\hat{\gamma}_{\text{ROBUST}}(h)$. A benchmark method that typically works under the weakest assumptions is *subsampling*; see Politis, Romano, and Wolf (1999).

There are other approaches to robust autocovariance estimation in the literature; see, for example, Ma and Genton (2000) and the references therein.

5. ROBUST POSITIVE DEFINITE ESTIMATION OF AUTOCORRELATION MATRIX

Let Σ denote the $n \times n$ matrix with i, j element $\Sigma_{i,j} := \rho(|i - j|)$; in other words, Σ is the autocorrelation matrix of the data $(X_1, \ldots, X_n)$ viewed as a vector. An immediate way to robustly estimate the autocorrelation matrix Σ is by plugging our robust correlation estimates. For example, define a matrix $\hat{\Sigma}$ that has i, j element given by $\hat{\rho}_{\text{ROBUST}}(|i - j|)$. Although intuitive, $\hat{\Sigma}$ is neither consistent for Σ as $n \to \infty$, nor is it positive definite; see Wu and Pourahmadi (2009) and the references therein.



		MSE					
ϕ	n	OLS	MG	RF	L1	LTS	MM
-0.2	50	0.01476	0.02723	0.04344	0.00428	0.02088	0.00303
	800	0.00143	0.00145	0.00468	0.00004	0.00016	0.00003
0.5	50	0.01034	0.02816	0.02636	0.00938	0.05276	0.02776
	800	0.00065	0.00122	0.00277	0.00504	0.01147	0.00798
0.8	50	0.01214	0.01920	0.02574	0.05146	0.12033	0.11011
	800	0.00051	0.00163	0.00139	0.02749	0.08644	0.09194

(a)

Figure 6. MA(1) with 10% innovation outlier contamination.

Following Politis (2011), we may define a "flat-top" taper as the function κ satisfying

$$\kappa(x) = \begin{cases} 1 & \text{if } |x| \le 1 \\ g(x) & \text{if } 1 < |x| \le c_\kappa \\ 0 & \text{if } |x| > c_\kappa, \end{cases}$$

where $c_\kappa \ge 1$ is a constant, and $g(x)$ is some function such that $|g(x)| \le 1$. Let the taper's l-scaled version be denoted as $\kappa_l(x) := \kappa(x/l)$. Taking κ of trapezoidal shape, that is, letting $g(x) = 2 - |x|$ and $c_\kappa = 2$, yields a simple taper that has been shown to work well in practice. McMurry and Politis (2012) introduced a consistent estimator of Σ defined as the $n \times n$ matrix with i, j element given by $\kappa_l(i-j)\hat{\rho}(|i-j|)$; here, l serves the role of a bandwidth parameter satisfying $l \to \infty$ but $l/n \to 0$ as $n \to \infty$.

To "robustify" the tapered estimator of McMurry and Politis (2012), we propose $\hat{\Sigma}_{\kappa,l}$ as an estimator of Σ where the i, j element of $\hat{\Sigma}_{\kappa,l}$ is given by $\kappa_l(i-j)\hat{\rho}_{\text{ROBUST}}(|i-j|)$. Note, however, that $\hat{\Sigma}_{\kappa,l}$ is not guaranteed to be positive definite. To address this problem, let $\hat{\Sigma}_{\kappa,l} = T_n D T_n^t$ be the spectral decomposition of $\hat{\Sigma}_{\kappa,l}$. Since $\hat{\Sigma}_{\kappa,l}$ is symmetric, T_n will be

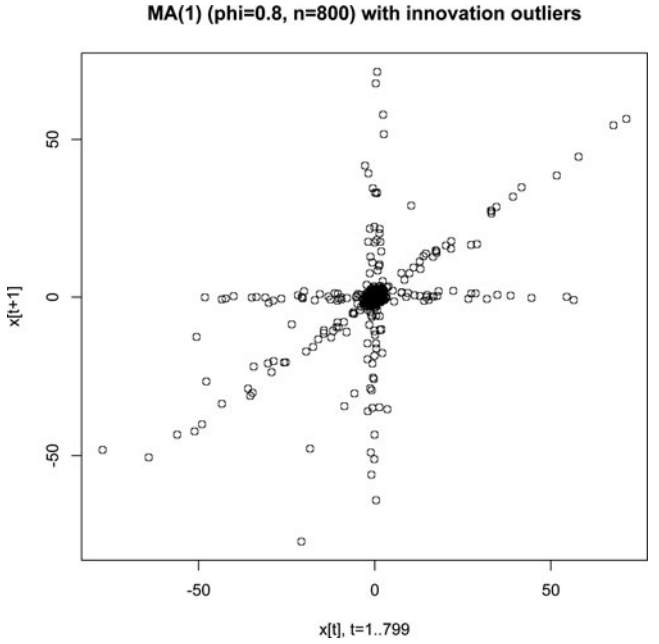


Figure 7. X_t versus X_{t+1} plot for the MA(1) model $X_t = Z_t + 0.8Z_{t-1}$ with innovation outliers. With an innovation outlier at Z_t, (X_{t-1}, X_t) usually lies on the vertical line, (X_t, X_{t+1}) on the diagonal, and (X_{t+1}, X_{t+2}) on the horizontal. The robust estimators tend to fit the diagonal line.

an orthogonal matrix, and $D = \operatorname{diag}(d_1, \ldots, d_n)$ which are the eigenvalues of $\hat{\Sigma}_{\kappa,l}$. Define

$$D^{(\epsilon)} := \operatorname{diag}\big(d_1^{(\epsilon)}, \ldots, d_n^{(\epsilon)}\big), \quad \text{with } d_i^{(\epsilon)} := \max\big(d_i, \epsilon/n^{\zeta}\big), \tag{4}$$

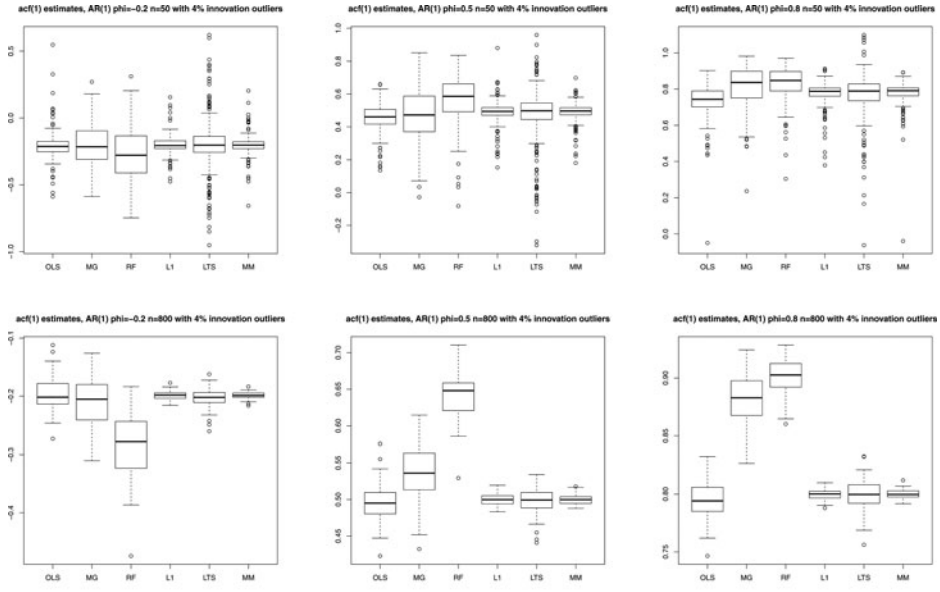
where $\epsilon \geq 0$ and $\zeta > 1/2$ are two constants. The choices $\zeta = 1$ and $\epsilon = 1$ works well in practice as in McMurry and Politis (2012).

Finally, we define

$$\hat{\Sigma}_{\kappa,l}^{(\epsilon)} := T_n D^{(\epsilon)} T_n^t. \tag{5}$$

By construction, $\hat{\Sigma}_{\kappa,l}^{(\epsilon)}$ is nonnegative definite when $\epsilon = 0$, and strictly positive definite when $\epsilon > 0$. Furthermore, $\hat{\Sigma}_{\kappa,l}^{(\epsilon)}$ inherits the robustness and consistency of $\hat{\Sigma}_{\kappa,l}$. Finally, note that the matrix that equals $\hat{\gamma}_{\text{ROBUST}}(0)\hat{\Sigma}_{\kappa,l}^{(\epsilon)}$ is positive definite, robust, and consistent estimator of the autocovariance matrix of the data $(X_1, \ldots, X_n)$ viewed as a vector.

Remark 5.1. As mentioned at the end of Section 3.1, the estimator $\hat{\rho}_{\text{ROBUST}}(h)$ is not necessarily in the interval $[-1, 1]$. The construction of the nonnegative definite estimator $\hat{\Sigma}_{\kappa,l}^{(\epsilon)}$ fixes this problem. In particular, the whole vector of autocorrelations $(\rho(0), \rho(1), \ldots, \rho(n-1))$—where of course $\rho(0) = 1$—is estimated by the first row of matrix $\hat{\Sigma}_{\kappa,l}^{(\epsilon)}$. As long as $\epsilon \geq 0$, this vector estimator can be considered as the first part of a nonnegative definite sequence, and hence it can be considered as the first part of the autocorrelation sequence of some stationary time series; see, for example, Brockwell and



		MSE					
ϕ	n	OLS	MG	RF	L1	LTS	MM
-0.2	50	0.01221	0.02652	0.04076	0.00522	0.05486	0.00659
	800	0.00094	0.00197	0.01052	0.00007	0.00035	0.00004
0.5	50	0.00934	0.02537	0.02717	0.00598	0.04063	0.00432
	800	0.00071	0.00252	0.02085	0.00008	0.00040	0.00005
0.8	50	0.01402	0.01325	0.00987	0.00535	0.02231	0.00647
	800	0.00033	0.00710	0.01053	0.00002	0.00021	0.00001

(a)

Figure 8. AR(1) with 4% innovation outlier contamination.

Davis (1991). Hence, the vector estimator enjoys all the properties of an autocorrelation sequence, including the property of belonging to the interval $[-1, 1]$.

6. ROBUST AR MODEL FITTING

6.1 ROBUST YULE-WALKER ESTIMATES

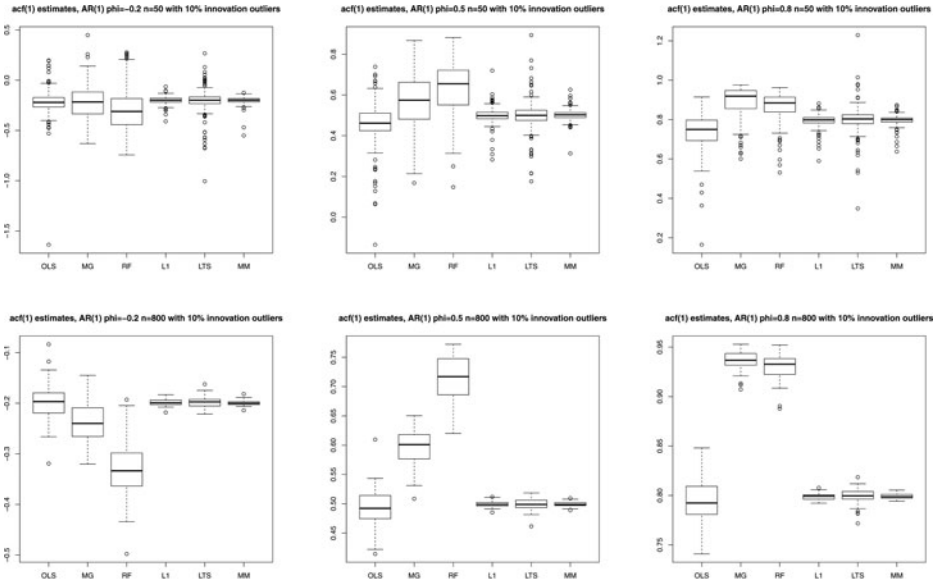
Consider again the causal AR(p) model of Equation (1), that is, $X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{k-p} + z_t$. Now we do not need to assume that the driving noise z_t is iid; it is sufficient that z_t is a mean zero, white-noise with variance σ^2.

In this context, autocovariance estimates can be directly used to derive AR coefficient estimates via the Yule-Walker equations:

$$\hat{\gamma}(0) = \phi_1 \hat{\gamma}(-1) + \cdots + \phi_p \hat{\gamma}(k-p) + \sigma^2$$

and

$$\hat{\rho}(k) = \phi_1 \hat{\rho}(k-1) + \cdots + \phi_p \hat{\rho}(k-p) \text{ for } k = 1, \ldots, p. \tag{6}$$



		MSE					
ϕ	n	OLS	MG	RF	L1	LTS	MM
-0.2	50	0.02157	0.02968	0.05325	0.00134	0.01702	0.00166
	800	0.00137	0.00276	0.002134	0.00004	0.00016	0.00003
0.5	50	0.01374	0.02147	0.03335	0.00161	0.00578	0.00083
	800	0.00120	0.01017	0.04705	0.00003	0.00011	0.00002
0.8	50	0.01338	0.01437	0.00983	0.00107	0.00510	0.00070
	800	0.00050	0.01853	0.01700	0.00001	0.00007	0.00001

(a)

Figure 9. AR(1) with 10% innovation outlier contamination.

However, if the classical autocovariance estimates are used in the above, a single outlier of size B perturbs the ϕ coefficient estimates by $O(B/n)$; a pair of such outliers can perturb $\hat{\phi}_1$ by $O(B^2/n)$.

A simple way to attempt to address this vulnerability is to plug robust autocovariance estimates into the linear system (6), that is, to use $\hat{\gamma}_{\text{ROBUST}}(0)$ in place of $\hat{\gamma(0)}$, and $\hat{\rho}_{\text{ROBUST}}(k)$ instead of $\hat{\rho}(k)$. The resulting robust estimate of the coefficient vector $\underline{\phi}_p := (\phi_1, \ldots, \phi_p)'$ is given by

$$\underline{\hat{\phi}}_{p,\text{ROBUST}} := S_p^{-1} \, \underline{\hat{\rho}}_p, \tag{7}$$

where for any $m \leq n$ we let $\underline{\hat{\rho}}_m := (\hat{\rho}_{\text{ROBUST}}(1), \ldots, \hat{\rho}_{\text{ROBUST}}(m))'$ and $\underline{\rho}_m := (\rho(1), \ldots, \rho(m))'$. Furthermore, in the above, S_p is the upper left $p \times p$ submatrix of the autocorrelation matrix $\hat{\Sigma}_{\kappa,l}^{(\epsilon)}$ defined in the previous section. Since S_p is a Toeplitz matrix, its inverse S_p^{-1} can be found via fast algorithms such as the Durbin-Levinson algorithm.

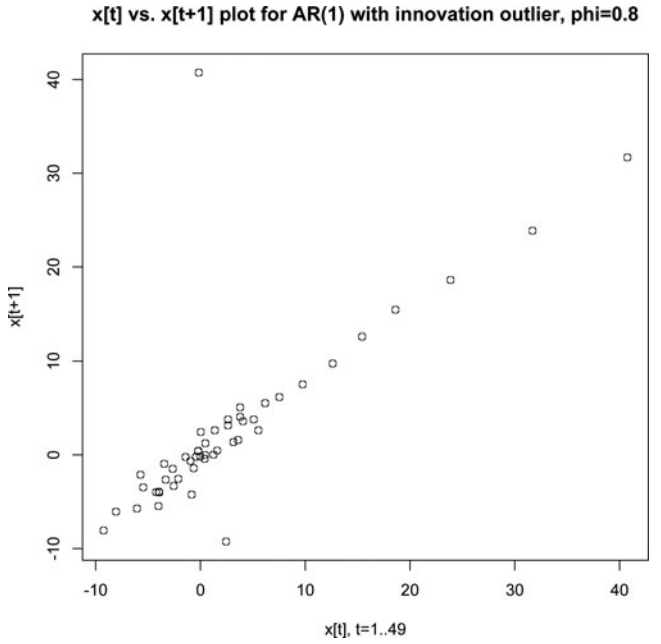


Figure 10. (x_t, x_{t+1}) plot for a realization of length $n = 50$ from the AR(1) time series $X_t = 0.8X_{t-1} + Z_t$ with one innovation outlier.

6.2 EXTENDED YULE-WALKER

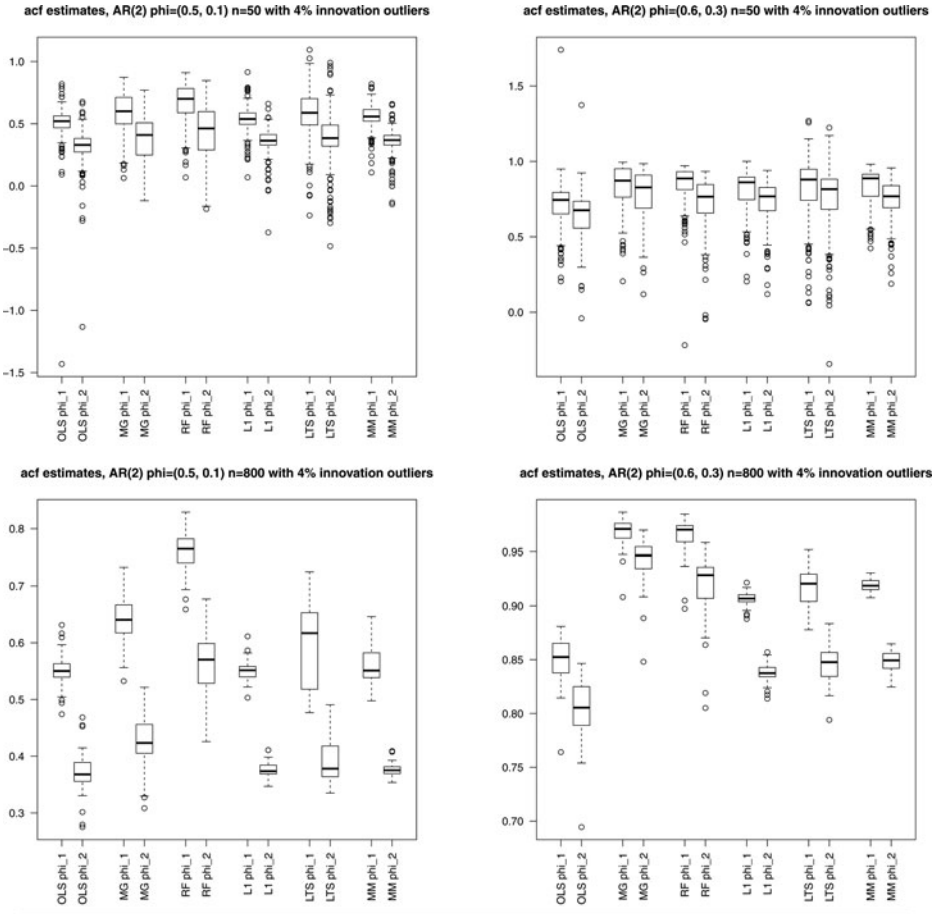
The "extended" Yule-Walker equations are identical to the usual ones of Equation (6) except letting $k = 1, \ldots, p'$ for some value $p' \geq p$, that is, they are an overdetermined system: p' equations with p unknowns $\underline{\phi}_p = (\phi_1, \ldots, \phi_p)'$. Politis (2009) noted that the extended Yule-Walker equations can be used to provide a more robust estimator of $\underline{\phi}_p$. For example, in the AR(1) case with $p = 1$, letting $p' = 2$ suggests that $\hat{\gamma}(1)/\hat{\gamma}(0)$ and $\hat{\gamma}(2)/\hat{\gamma}(1)$ are equally valid as estimators for ϕ_1 that could be combined to yield an improved one.

Generalizing this idea, fix $p' \geq p$, and let $S_{p',p}$ be the $p' \times p$ matrix with jth column equal to $(\hat{\rho}_{\text{ROBUST}}(1-j), \hat{\rho}_{\text{ROBUST}}(2-j), \ldots, \hat{\rho}_{\text{ROBUST}}(p'-j))$; alternatively, we could define $S_{p',p}$ as the upper left $p' \times p$ submatrix of the autocorrelation matrix $\hat{\Sigma}^{(\epsilon)}_{\kappa,l}$ defined in Section 5.

As mentioned above, the extended Yule-Walker equations with $k = 1, \ldots, p' > p$ is an overdetermined system. Thus, we cannot expect that $\underline{\rho}_{p'}$ exactly equals $S_{p',p}\underline{\phi}_p$. Define the discrepancy $\underline{\eta} = \underline{\rho}_{p'} - S_{p',p}\underline{\phi}_p$ from which it follows that

$$\underline{\rho}_{p'} = S_{p',p}\underline{\phi}_p + \underline{\eta}. \tag{8}$$

As suggested by Politis (2009) , Equation (8) can be viewed as a linear regression (with "errors-in-variables") having response $\underline{\rho}_{p'}$, design matrix $S_{p',p}$, parameter vector $\underline{\phi}_p$, and error vector $\underline{\eta}$. Finding OLS estimates of $\underline{\phi}_p$ in the regression model (8) gives doubly robust estimates: robust because $\hat{\rho}_{\text{ROBUST}}(h)$ were used but also because of the use of the



		MSE					
ϕ_1, ϕ_2	n	OLS	MG	RF	L1	LTS	MM
0.5, 0.1	50	.03157, .02985	.02954, .03689	.03770, .05072	.01108, .01249	.03841, .05011	.01028, .01164
	800	.00084, .00139	.00868, .00444	.04332, .03728	.00034, .00017	.00705, .00177	.00115, .00011
0.6, 0.3	50	.04267, .05308	.02182, .02743	.01571, .03576	.02180, .02677	.04242, .05428	.01504, .02137
	800	.00051, .00080	.01243, .01673	.01198, .01173	.00244, .00063	.00399, .00143	.00387, .00122

(a)

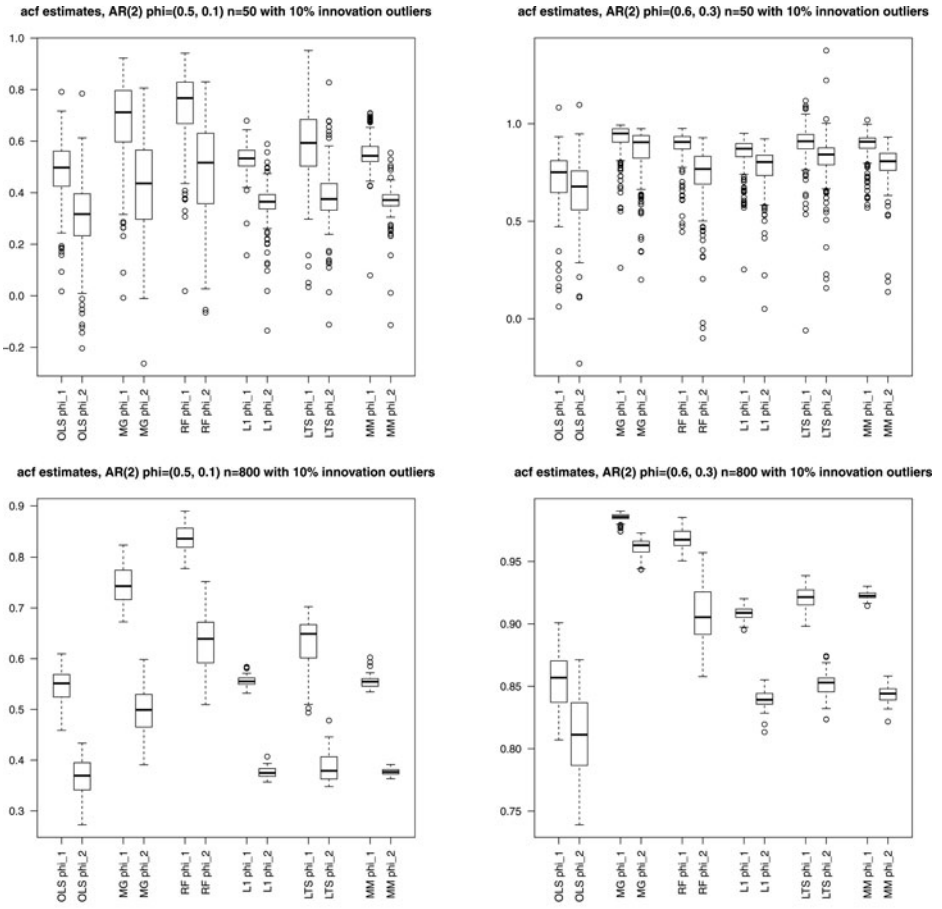
Figure 11. AR(2) with 4% innovation outlier contamination. True $(\rho(1), \rho(2))$ is $(\frac{5}{9}, \frac{17}{45})$ in the $(\phi_1, \phi_2) = (0.5, 0.1)$ case, and $(\frac{6}{7}, \frac{57}{70})$ in the $(\phi_1, \phi_2) = (0.6, 0.3)$ case.

extended Yule-Walker equations. One could alternatively use a robust regression technique to fit regression (8); the details are obvious and are omitted.

7. NUMERICAL WORK

7.1 SIMULATION WITHOUT OUTLIERS

First, we generated time series data $X_1, \ldots, X_n$ according to the MA(1) model $X_t = Z_t + \phi Z_{t-1}$ (with no outliers) with $\phi \in \{0.2, 0.5, 0.8\}$, $n \in \{50, 800\}$, and Z_t iid N(0, 1). We estimated the lag-1 autocorrelations in different ways, and compared them to the true



ϕ_1, ϕ_2	n	MSE					
		OLS	MG	RF	L1	LTS	MM
0.5, 0.1	50	.01857, .02521	04138, .03767	.05084, .04835	.00392, .00619	.01945, .01144	.00416, .00423
	800	.00107, .00162	03673, .01651	.07812, .06693	.00013, .00011	.00818, .00092	.00018, .00004
0.6, 0.3	50	.04177, .05834	01242, .01836	.00863, .03323	.00782, .01384	.01276, .01604	.00629, .01149
	800	.00057, .00099	01624, .02157	.01236, .00935	.00262, .00068	.00412, .00145	.00434, .00091

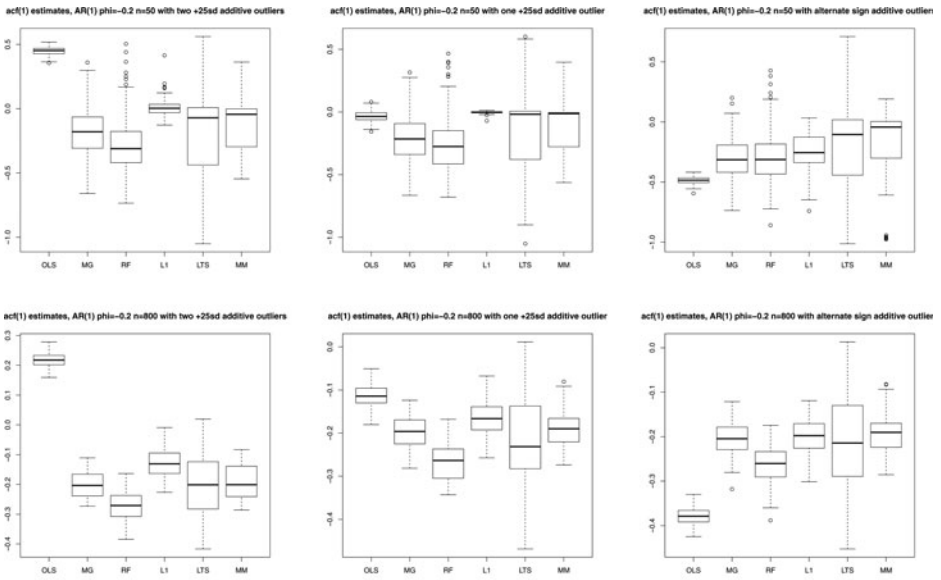
(a)

Figure 12. AR(2) with 10% innovation outlier contamination.

value ($\frac{\phi}{1+\phi^2}$). We did the same thing for the AR(1) model $X_t = \phi X_{t-1} + Z_t$; the true autocorrelation is ϕ in this case.

The estimators that were constructed were $\hat{\rho}_{\text{ROBUST}}(h)$ using the three aforementioned options for robust regression: L1, LTS, and MM. As baselines for comparison, we included ordinary least squares (OLS) regression (which as discussed above is nearly identical to the sample acf, Ma and Genton 2000), robust autocorrelation estimator (denoted MG), and a sample acf/robust filtering approach discussed in Maronna, Martin, and Yohai (2006) (denoted RF) and implemented by Dürre, Fried, and Liboschik (2014).

As expected, the OLS (classical) estimator performed best in the no contamination case; see Figures 3 and 4. However, the MM estimator's performance was nearly indistinguishable



		MSE					
Contam. pattern	n	OLS	MG	RF	L1	LTS	MM
25, 25	50	0.42073	0.03200	0.05025	0.04623	0.10268	0.03537
	800	0.17371	0.00180	0.00822	0.00718	0.01194	0.00322
25, 0	50	0.02932	0.03333	0.05232	0.03891	0.09765	0.03828
	800	0.00799	0.00148	0.00643	0.00313	0.01316	0.00231
25, -25	50	0.08199	0.03762	0.04818	0.02533	0.11194	0.05671
	800	0.03307	0.00162	0.00639	0.00187	0.01199	0.00251

(a)

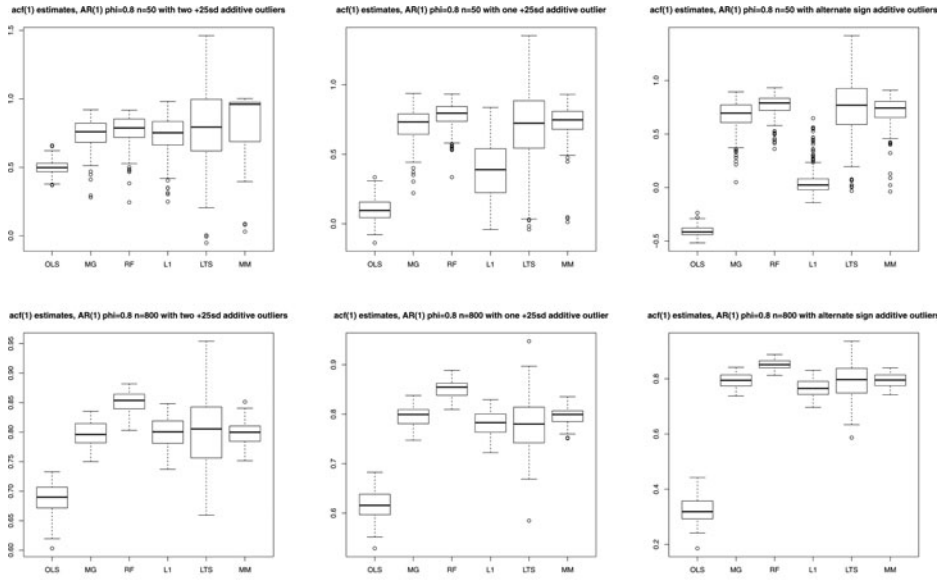
Figure 13. AR(1), $\phi = -0.2$ with additive outliers. In a length-n time series, an "a, b" contamination pattern means that a was added to the $\frac{n}{2}$th element and b was added to the $(\frac{n}{2}+1)$th element.

from OLS's. The L1, Ma-Genton, and robust filtering estimators were somewhat less efficient, with MSEs roughly 1.5× to 2× that of the OLS estimator. LTS's known terrible normal efficiency was clearly in evidence.

7.2 SIMULATION WITH INNOVATION OUTLIERS

Next, we investigated estimator performance when faced with innovation outliers, modifying Z_t to be distributed according to a Gaussian mixture, 90 or 96% $N(0, 1)$ and 10 or 4% $N(0, 625)$.

From Figures 5 and 6 we can see that for $\phi = -0.2$, the L1 and MM estimators do a better job of handling the innovation outliers than the sample acf. However, for larger values of ϕ and large sample sizes, our robust estimates of $\rho(1)$ are significantly biased toward ϕ (when the true value is $\frac{\phi}{1+\phi^2}$). The reason is that any innovation outlier not immediately followed by a second one creates a point on the scatterplot of the form $(x + \varepsilon_1, \phi x + \varepsilon_2)$ where $|x| >> |\varepsilon_i|$; all of these high-magnitude points trace a single line of slope ϕ which are picked up by the robust estimators as the primary signal, and the other high-magnitude outlier points (which bring the OLS estimate in line) are ignored. See



Contam. pattern	n	MSE					
		OLS	MG	RF	L1	LTS	MM
25, 25	50	0.09189	0.01536	0.01300	0.02420	0.06136	0.03676
	800	0.01352	0.00041	0.00294	0.00059	0.00482	0.00039
25, 0	50	0.49503	0.02238	0.00874	0.22382	0.06768	0.02171
	800	0.03502	0.00055	0.00301	0.00102	0.00494	0.00044
25, -25	50	1.46111	0.03628	0.01192	0.56448	0.07636	0.03018
	800	0.22926	0.00062	0.00293	0.00231	0.00520	0.00053

(a)

Figure 14. AR(1), $\phi = 0.8$ with additive outliers.

Figure 7 for an illustration. The Ma-Genton and robust filtering estimators, not being based on linear regression, are not affected by this pattern (though they still fail to outperform the sample acf).

From Figures 8 and 9, we can see that our robust regression estimators $\hat{\rho}_{\text{ROBUST}}$ all shine in the AR(1) case with innovation outliers. This is unsurprising, since an AR(1) innovation outlier only pulls one point off the appropriate regression line, while generating several other high-leverage points right on the true line; see Figure 10. The Ma-Genton and robust filtering estimators generally perform more poorly than OLS here.

Moving on to the AR(2) case of Figures 11 and 12, we see that with innovation outliers, the L1 and MM robust estimators exhibit much better performance than OLS given a small (50) sample size, though the advantage vanishes with a larger sample size. The Ma-Genton and robust filtering estimators perform relatively poorly across the board.

7.3 SIMULATION WITH ADDITIVE OUTLIERS

Next, we investigated the performance of our estimators in the additive outlier case by perturbing one or two elements in the middle of the time series by a large number (where, as before, Z_t is iid N(0, 1)).

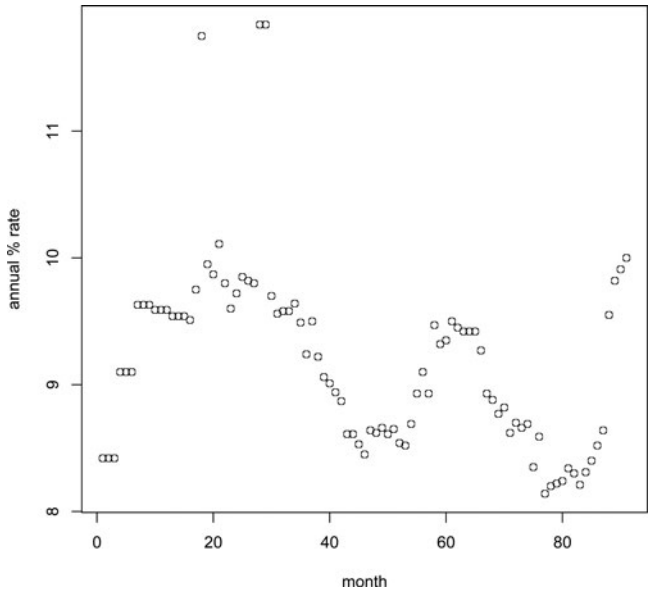


Figure 15. Austrian bank data.

The Ma-Genton and MM-based estimators do the best here. The OLS estimator performed especially badly in the $\phi = 0.8$ case, L1 was fairly good but failed the $\phi = 0.8$ with $n = 50$ case (while, conversely, RF was generally mediocre but shined with $\phi = 0.8$, $n = 50$), and LTS generally acted as a much less efficient MM.

7.4 REAL DATA EXPERIMENT: AUSTRIAN BANK INTEREST RATES

We also applied our robust estimators to some real-world data, namely monthly interest rates of an Austrian bank over a 91-month period; this dataset has been previously discussed and analyzed by Künsch (1983), Ma and Genton (2000), and Genton and Ruiz-Gazen (2010). The three outliers above 11% around months 20 and 30 are clearly notable in Figure 15. Table 1 summarizes our findings.

The L1 and MM (as well as MG and RF) estimates of typical month-to-month and season-to-season autocorrelation are more reasonable than the OLS estimate which is

Table 1. Numerical results with Austrian bank data

Estimator	Outliers replaced?	$\hat{\rho}(1)$	$\hat{\rho}(2)$	$\hat{\rho}(3)$	$\hat{\rho}(4)$	$\hat{\rho}(5)$	$\hat{\rho}(6)$	$\hat{\rho}(12)$
OLS	no	0.79184	0.63702	0.58923	0.51249	0.44414	0.40440	0.08583
	yes	0.93911	0.87238	0.77912	0.67222	0.58084	0.49877	0.07229
MG	no	0.96571	0.91741	0.82703	0.73727	0.65968	0.55046	−0.18033
RF	no	0.91748	0.76950	0.61159	0.48246	0.38337	0.30356	−0.00780
L1	no	0.97222	0.91329	0.83459	0.78351	0.72603	0.65957	−0.02786
LTS	no	0.99451	0.98020	0.95588	0.87975	0.85556	0.36749	−0.94203
MM	no	0.97194	0.94398	0.81113	0.49292	0.40119	0.34198	0.04550

Table 2. Total time to compute $\hat{\rho}(1)$ and $\hat{\rho}(2)$ for 200 AR(2) ($\phi_1 = 0.6, \phi_2 = 0.3$) length-$n$ time series, on a MacBook Pro with a 2.8 GHz Intel Core 2 Duo processor

n	Method	Language	Total time
80	OLS	R	0.69s
		C	0.0011s
	L1	R	0.84s
		C	0.0069s
800	OLS	R	1.2s
		C	0.01s
	L1	R	1.5s
		C	0.14s
8000	OLS	R	5.1s
		C	0.11s
	L1	R	14s
		C	6.6s
80000	OLS	R	70.8s
		C	1.0s
	L1	R	730s
		C	590s

NOTE: For the $n = 80$ case, we processed 20,000 series and divided the final times by 100. R and C source code is posted at *http://www.math.ucsd.edu/politis/DPpublication.html*.

overly affected by the outliers. However, the LTS estimator was erratic, overestimating the low lag autocorrelations and yielding a bizarre value of −0.94203 for the 12-month autocorrelation.

7.5 COMPUTATIONAL COST

When performing these simulations, the *relative* computational cost of the robust estimators (including Ma-Genton) tended to exceed that of the classical estimator by one to two orders of magnitude. (The exact multiplier depends on the type of robust regression employed, and how it was implemented.) However, the *absolute* computation time for each full set of simulations was never more than a few days (in R). Thus, unless the dataset needs to be analyzed in real time and/or is much larger than what we have simulated, the extra time requirement is not a problem.

L1 regression is especially tractable. The `rq()` function in Koenker et al.'s R `quantreg` package came within 25% of `lm()` (and was actually faster than `glm()`) for length-80 and 800 time series; this is close enough that choice of implementation language has a much greater impact (see Table 2). For very long time series, the difference between L1 regression's O(n^2) time complexity OLS's O(n) makes itself felt, but L1 remains realistic until n is in the millions.

8. CONCLUSIONS

A class of robust autocorrelation estimators $\hat{\rho}_{\text{ROBUST}}$ was proposed, based on interpreting the sample acf as a linear (auto)regression to be fitted via robust regression techniques. Three

members of this class, based on L1, LTS, and MM robust regression, respectively, were compared with the sample acf and the Ma-Genton robust estimator in a simulation study, after a brief discussion of their theoretical properties.

MM regression exhibited an excellent combination of efficiency and robustness. Interestingly, L1 also performed well despite its inferior theoretical properties. However, the version of LTS we used was too inefficient to outperform the sample acf even in the presence of contamination.

There was one contamination pattern we simulated that none of our robust estimators handled well, namely innovation outliers in the context of a MA(1) model, where many of the high-magnitude elements of the time series arguably exhibit a well-defined but different autocorrelation than the rest of the series. If this type of outlier may be present, we recommend that the Ma-Genton estimator be used instead. (It would be useful to develop automatic diagnostics to guard against this case.) Conversely, our MM and L1-based estimators clearly outperform the Ma-Genton estimator in the context of AR(p) data with innovation outliers.

ACKNOWLEDGMENTS

Many thanks are due to Tim McMurry for helpful R tips, Marc Genton for sharing the Austrian bank data with us, and Alexander Dürre for Ma-Genton and robust filtering estimator code. The research of Dimitris N. Politis was partially supported by NSF grants DMS 13-08319 and DMS 12-23137.

[Received September 2013. Revised August 2014.]

REFERENCES

Brockwell, P. J., and Davis, R. A. (1991), *Time Series: Theory and Methods*, New York: Springer. [144,156]

Davis, R. A., and Mikosch, T. (2000), "The Sample Autocorrelations of Financial Time Series Models," *Nonlinear and Nonstationary Signal Processing*, Cambridge: Cambridge University Press, pp. 247–274. [149]

Denby, L., and Martin, R. D. (1979), "Robust Estimation of the First-Order Autoregressive Parameter," *Journal of the American Statistical Association*, 74, 140–146. [145,148]

Dürre, A., Fried, R., and Liboschik, T. (2014), "Robust Estimation of (partial) Autocorrelation," *Sonderforschungsbereich (SFB)*, 823. [160]

Hampel, F. R. (1973), "Robust Estimation, a Condensed Partial Survey," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 27, 87–104. [144]

Huber, P. J. (1973), "Robust Regression: Asymptotics, Conjectures and Monte Carlo," *The Annals of Statistics*, 1, 799–821. [147]

Genton, M. G., and Lucas, A. (2003), "Comprehensive Definitions of Breakdown-Points for Independent and Dependent Observations," *Journal of the Royal Statistical Society,* Series B, 65, 81–94. [149]

——— (2005), Discussion of "Breakdown and Groups" by L. Davies and U. Gather, *The Annals of Statistics*, 33, 988–993. [149]

Genton, M. G., and Ruiz-Gazen, A. (2010), "Visualizing Influential Observations in Dependent Data," *Journal of Computational and Graphical Statistics*, 19, 808–825. [163]

Künsch, H. (1984), "Infinitesimal Robustness for Autoregressive Processes," *The Annals of Statistics*, 12, 843–863. [145,163]

Kreiss, J.-P., and Paparoditis, E. (2011), "Bootstrap Methods for Dependent Data: A Review," *Journal of the Korean Statistical Society*, 40, 357–378. [153]

Kreiss, J.-P., Paparoditis, E., and Politis, D. N. (2011), "On the Range of Validity of the Autoregressive Sieve Bootstrap," *The Annals of Statistics*, 39, 2103–2130. [145]

Ma, Y., and Genton, M. G. (2000), "Highly Robust Estimation of the Autocovariance Function," *Journal of Time Series Analysis*, 21, 663–684. [145,149,150,153,160,163]

Maronna, R. A., Martin, R. D., and Yohai, V. J. (2006), *Robust Statistics: Theory and Methods*, New York: Wiley. [145,148,160]

McMurry, T., and Politis, D. N. (2010), "Banded and Tapered Estimates of Autocovariance Matrices and the Linear Process Bootstrap," *Journal of Time Series Analysis*, 31, 471–482. [Corrigendum: vol. 33, 2012.] [154,155]

Politis, D. N. (2009), "An Algorithm for Robust Fitting of Autoregressive Models," *Economics Letters*, 102, 128–131. [158]

——— (2011), "Higher-Order Accurate, Positive Semi-Definite Estimation of Large-Sample Covariance and Spectral Density Matrices," *Econometric Theory*, 27, 703–744. [154]

Politis, D. N., Romano, J. P., and Wolf, M. (1999), *Subsampling*, New York: Springer. [153]

Rousseeuw, P. J., and Leroy, A. M. (1987), *Robust Regression and Outlier Detection*, New York: Wiley. [150]

Wu, W. B., and Pourahmadi, M. (2009), "Banding Sample Autocovariance Matrices of Stationary Processes," *Statistica Sinica*, 19, 1755–1768. [153]

Yohai, V. J. (1987), "High Breakdown-Point and High Efficiency Estimates for Regression," *The Annals of Statistics*, 15, 642–656. [148]