Robust Autocorrelation Estimation

Christopher C. Chang and Dimitris N. Politis

Department of Mathematics

University of California at San Diego

La Jolla, CA 92093-0112 USA.

E-mails: chrchang@alumni.caltech.edu, dpolitis@ucsd.edu

24 Apr 2014

Abstract

In this paper, 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. Construction of robust autocovariance and positive definite autocorrelation estimates is discussed, as well as application to AR model fitting. Simulation studies with various outlier configurations are performed in order to compare the different estimators.

Keywords: autocorrelation; 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 [1].

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 [4], this is a serious problem.

To address it, several robust autocorrelation estimators have been proposed [3] [9] [12] [13]. 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 the paper at hand, 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 paper 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

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

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

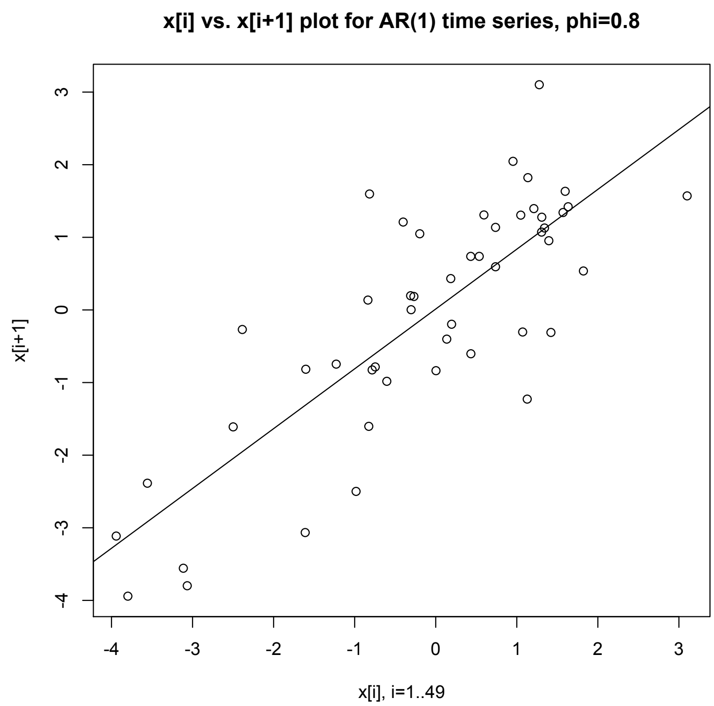

Figure 1: Scatterplot of (X_t, X_{t+1}) for a realization of length 50 from the AR(1) time series $X_t = 0.8X_{t-1}+Z_t$, Z_t iid $N(0,1)$. Regression line is $y = 0.82375x + 0.01289$.

orthogonal projection of Y onto the linear space spanned by X; see e.g. Kreiss, Paparoditis, and Politis [11].

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.

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

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

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 $\hat{\rho}_{OLS}(h)$ 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 $\hat{\rho}_{OLS}(h)$ 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 $\hat{\rho}_{ROBUST}(h)$ to be the estimator of slope β_1 in the straight line regression

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

using *any* robust method to fit the regression.

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 $\hat{\rho}_{ROBUST}(h)$.

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; i.e. 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. $\hat{\rho}_{LTS}$. 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 $|r|_{(1)} \leq \ldots \leq |r|_{(n-h)}$ 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. $\hat{\rho}_{MM}$. An M-estimate [5] 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 utilizes 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 $\frac{1}{2}$.) It has superior robustness, but is inefficient.

MM-estimates, pioneered by Yohai [20], 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 et al. [13].

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 [3]:

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.

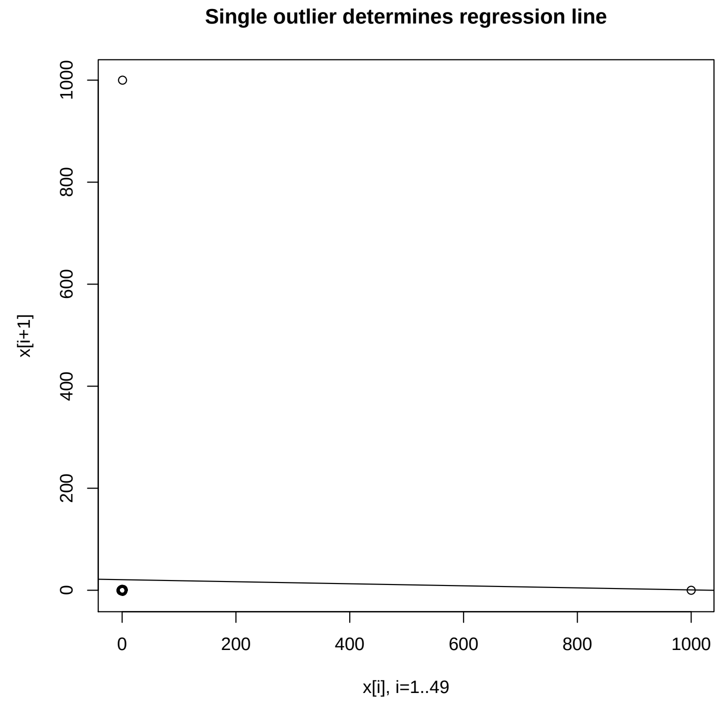

Figure 2: Degenerate regression line from 50 N(0,1) points contaminated by one outlier at 1000.

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 [12] 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 [6] [7] who analyze 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 [2]. We will perform simulations to investigate whether our robust regression estimates are able to keep up.

Next, the *influence function* (IF) describes the impact on the autocorrelation estimate $\hat{\rho}_{ROBUST}$ of adding an infinitesimal probability of an innovation or additive outlier. Given an innovation or residual distribution F, let $\hat{\rho}_{ROBUST}(F)$ 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

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

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 [12]. 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.

Remark 3.1. As a final note, our robust estimator $\hat{\rho}_{ROBUST}(h)$ 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 $\hat{\rho}_{ROBUST}(h)$ to the values -1 and 1. In other words, estimate $\rho(h)$ by $\hat{\rho}_{ROBUST}(h)$ when the latter is in the interval [-1, 1]; else, estimate $\rho(h)$ by 1 when $\hat{\rho}_{ROBUST}(h) > 1$, and by -1 when $\hat{\rho}_{ROBUST}(h) < -1$. On the other hand, if it desired to estimate $\rho(k)$ for many lags, e.g. $k = 1, \ldots, h$, a more elaborate fix is required; in that respect, see Section 5, and in particular, Remark 5.1.

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 ($\frac{1}{2}$). It is robust with respect to both x- and y-outliers, so $\hat{\rho}_{LTS}$ retains the $\frac{1}{2}$ BP in the AR(1) innovation outliers case and has a BP of at least $\frac{1}{4}$ 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% [18]; this weakness is clearly visible in our simulations.

3.4 MM

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

The normal efficiency is actually a user-adjustable parameter. In practice, it it is usually chosen to be between 0.7 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 mean μ:

$$
\begin{aligned}
X_t - \mu &= \beta_1(X_{t-h} - \mu) + \text{error, since this line should have zero intercept} \\
X_t &= \beta_0 + \beta_1 X_{t-h} + \text{error} \qquad (3) \\
\text{where} \quad \beta_0 &= \mu(1-\beta_1) \qquad \text{(combining eq. (2) and (3)).} \\
\text{Finally, let} \quad \hat{\mu}_h &:= \frac{\hat{\beta}_0}{1-\hat{\beta}_1}
\end{aligned}
$$

   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 $k = 1, \ldots, p$ in order to boil them down to a single estimate.

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

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

Remark 4.1. The estimators $\hat{\rho}_{ROBUST}(h)$ and $\hat{\gamma}_{ROBUST}(h)$ 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, e.g., 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 $\hat{\rho}_{ROBUST}(h)$ and/or $\hat{\gamma}_{ROBUST}(h)$. Several powerful resampling methods for time series data have been developed in the last 25 years, e.g., blocking methods, AR-sieve bootstrap, frequency domain methods, etc.; see the recent review by Kreiss and Paparoditis [10] and the references therein. It is unclear at the moment which of these methods would be preferable in the context of $\hat{\rho}_{ROBUST}(h)$ and $\hat{\gamma}_{ROBUST}(h)$. A benchmark method that typically works under the weakest assumptions is *Subsampling*; see Politis, Romano, and Wolf [17].

There are other approaches to robust autocovariance estimation in the literature; see e.g. Ma and Genton [12] 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 $\hat{\rho}_{ROBUST}(|i-j|)$. Although intuitive, $\hat{\Sigma}$ is neither consistent for Σ as $n \to \infty$, nor is it positive definite; see Wu and Pourahmadi [19] and the references therein.

Following Politis [16], we may define a 'flat-top' taper as the function κ satisfying

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

where $c_\kappa \geq 1$ is a constant, and $g(x)$ is some function such that $|g(x)| \leq 1$. Let the taper's l-scaled version be denoted as $\kappa_l(x) := \kappa(x/l)$. Taking κ of trapezoidal shape, i.e., 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 [14] 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 $n \to \infty$.

In order to "robustify" the tapered estimator of McMurry and Politis [14], we propose $\hat{\Sigma}_{\kappa,l}$ as an estimator of Σ where the i,j element of $\hat{\Sigma}_{\kappa,l}$ is given by $\kappa_l(i-j)\hat{\rho}_{ROBUST}(|i-j|)$. 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 an orthogonal matrix, and $D = \text{diag}(d_1, \ldots, d_n)$ which are the eigenvalues of $\hat{\Sigma}_{\kappa,l}$). Define

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

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 [14].

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 $\hat{\gamma}_{ROBUST}(0)\hat{\Sigma}_{\kappa,l}^{(\epsilon)}$ 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 $\hat{\rho}_{ROBUST}(h)$ 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 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 e.g. Brockwell and 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 eq. (1), i.e., $X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{k-p} + z_t$. 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:

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

and

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

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), i.e., to use $\hat{\gamma}_{ROBUST}(0)$ in place of $\hat{\gamma(0)}$, and $\hat{\rho}_{ROBUST}(k)$ instead of $\hat{\rho}(k)$. The resulting robust estimate of the coefficient vector $\underline{\phi}_p := (\phi_1, \ldots, \phi_p)'$ is given by

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

where for any $m \le n$ we let $\underline{\hat{\rho}}_m := (\hat{\rho}_{ROBUST}(1), \ldots, \hat{\rho}_{ROBUST}(m))'$ 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.

6.2 Extended Yule-Walker

The 'extended' Yule-Walker equations are identical to the usual ones of eq. (6) except letting $k = 1, \ldots, p'$ for some value $p' \ge p$, i.e., they are an overdetermined system: p' equations with p unknowns $\underline{\phi}_p = (\phi_1, \ldots, \phi_p)'$. Politis [15] 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 $p' \ge p$, and let $S_{p',p}$ be the $p' \times p$ matrix with jth column equal to $(\hat{\rho}_{ROBUST}(1-j), \hat{\rho}_{ROBUST}(2-j), \ldots, \hat{\rho}_{ROBUST}(p'-j))$; alternatively, we could define $S_{p',p}$ as the upper left $p' \times p$ submatrix of the autocorrelation matrix $\hat{\Sigma}_{\kappa,l}^{(\epsilon)}$ defined in Section 5.

As mentioned above, the extended Yule-Walker equations with $k = 1, \ldots, p' > p$ is an overdetermined system. Thus, we can not 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 [15] , 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 $\hat{\rho}_{ROBUST}(h)$ were used but also because of the use of the 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 $\phi \in \{0.2, 0.5, 0.8\}$, $n \in \{50, 200, 800\}$, and Z_t i.i.d. N(0, 1). We estimated the lag-1 and lag-2 autocorrelations in different ways, and compared them to the true values ($\frac{\phi}{1+\phi^2}$ and 0, respectively). We did the same thing for the AR(1) model $X_t = \phi X_{t-1} + Z_t$; the true autocorrelations are ϕ and ϕ^2 in this case.

The estimators that were constructed were $\hat{\rho}_{ROBUST}(h)$ 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, and Ma and Genton's [12] robust autocorrelation estimator (denoted MG).

ϕ	n	Estimator	Avg. $\hat{\rho}(1)$	MSE	Avg. $\hat{\rho}(2)$	MSE
0.2	50	OLS	.16815	.01669	-.04035	.02312
		MG	.17428	.02465	-.03364	.03676
		L1	.15741	.02938	-.04618	.03622
		LTS	.12148	.11283	-.06980	.13513
		MM	.16728	.01731	-.04223	.02533
	200	OLS	.18238	.00458	-.01316	.00546
		MG	.18174	.00629	-.01753	.00714
		L1	.18875	.00827	-.01559	.00861
		LTS	.18659	.04622	-.02034	.04300
		MM	.18328	.00489	-.01330	.00574
	800	OLS	.19202	.00120	.00173	.00108
		MG	.19266	.00127	.00152	.00135
		L1	.19457	.00190	.00080	.00213
		LTS	.20289	.01614	.00342	.01447
		MM	.19253	.00122	.00154	.00123
0.5	50	OLS	.35834	.01685	-.03677	.02702
		MG	.36166	.02319	-.02692	.03660
		L1	.35859	.02194	-.01190	.03290
		LTS	.38351	.07726	.00142	.10233
		MM	.35940	.01748	-.02757	.02745
	200	OLS	.39859	.00216	-.00520	.00516
		MG	.39992	.00308	-.00571	.00707
		L1	.39810	.00520	-.00163	.00862
		LTS	.40652	.03394	.01994	.04868
		MM	.39731	.00252	-.00528	.00560
	800	OLS	.39746	.00094	-.00465	.00183
		MG	.39809	.00111	-.00344	.00239
		L1	.39897	.00175	-.00113	.00258
		LTS	.39555	.01439	.00574	.01894
		MM	.39780	.00100	-.00395	.00199
0.8	50	OLS	.45355	.01053	-.05546	.03023
		MG	.45369	.01663	-.06168	.04081
		L1	.44862	.01992	-.06792	.04046
		LTS	.46865	.08112	-.06159	.12601
		MM	.45345	.01106	-.05628	.03074
	200	OLS	.48315	.00242	-.00775	.00667
		MG	.48289	.00322	-.00604	.00877
		L1	.48248	.00470	-.00235	.00847
		LTS	.49077	.02759	.02308	.03534
		MM	.48340	.00256	-.00730	.00663
	800	OLS	.48415	.00055	-.00434	.00166
		MG	.48349	.00067	-.00541	.00186
		L1	.48356	.00121	-.00320	.00202
		LTS	.47204	.01296	.00645	.01402
		MM	.48402	.00059	-.00436	.00166

Table 1: Uncontaminated MA(1) simulation results; averages of 200 trials.

ϕ	n	Estimator	Avg. $\hat{\rho}(1)$	MSE	Avg. $\hat{\rho}(2)$	MSE
0.2	50	OLS	.16358	.02592	.02875	.01956
		MG	.15360	.03837	.02700	.03465
		L1	.17565	.03564	.01710	.03553
		LTS	.18526	.11907	-.00201	.12429
		MM	.16702	.02758	.02804	.02197
	200	OLS	.20110	.00439	.02818	.00552
		MG	.20064	.00550	.02512	.00688
		L1	.19851	.00733	.02330	.00762
		LTS	.19576	.04101	.02079	.03917
		MM	.20009	.00459	.02691	.00562
	800	OLS	.19193	.00125	.04054	.00123
		MG	.19286	.00162	.04009	.00146
		L1	.19139	.00206	.04056	.00211
		LTS	.19555	.01551	.05124	.01590
		MM	.19191	.00137	.04066	.00124
0.5	50	OLS	.44600	.01603	.18352	.02630
		MG	.44176	.02597	.18445	.03796
		L1	.45312	.02454	.19821	.03591
		LTS	.46085	.09045	.21308	.11105
		MM	.44471	.01738	.18691	.02687
	200	OLS	.48241	.00417	.23662	.00681
		MG	.47893	.00494	.23194	.00776
		L1	.48157	.00635	.23560	.00937
		LTS	.48630	.03007	.22803	.03912
		MM	.48229	.00429	.23674	.00699
	800	OLS	.49777	.00100	.24495	.00157
		MG	.49708	.00125	.24396	.00202
		L1	.49994	.00147	.24465	.00210
		LTS	.50000	.00983	.24269	.01308
		MM	.49796	.00105	.24512	.00165
0.8	50	OLS	.72894	.01682	.52273	.04186
		MG	.70482	.02413	.48780	.05783
		L1	.72172	.02256	.51311	.05671
		LTS	.69385	.06811	.49295	.15527
		MM	.72896	.01790	.51800	.04563
	200	OLS	.78556	.00191	.61795	.00502
		MG	.78135	.00235	.61327	.00565
		L1	.78586	.00291	.61878	.00646
		LTS	.78713	.01646	.61040	.03228
		MM	.78498	.00193	.61847	.00489
	800	OLS	.79622	.00045	.63450	.00142
		MG	.79563	.00052	.63324	.00166
		L1	.79702	.00076	.63717	.00185
		LTS	.80020	.00573	.64809	.00765
		MM	.79634	.00048	.63522	.00149

Table 2: Uncontaminated AR(1) simulation results; averages of 200 trials.

As expected, the OLS (classical) estimator performed best in the no contamination case; see Tables 1 and 2. However, the MM estimator's performance was nearly indistinguishable from OLS's. The L1 and Ma-Genton estimators were somewhat less efficient, with MSEs roughly 1.5x to 2x that of the OLS estimator; LTS's known terrible normal efficiency was clearly in evidence.

Sample size did not affect the performance of the estimators relative to each other, but a larger sample size reduced the downward bias of them all.

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 percent N(0, 1) and 10 or 4 percent N(0, 625).

From Table 3 we can see that for $\phi = -0.2$, the Ma-Genton, L1, and MM estimators do a substantially 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 $\rho(1)$ cluster toward ϕ instead of $\frac{\phi}{1+\phi^2}$. The reason for that is that any innovation outlier not immediately followed by a second one creates a point on the scatterplot of the form $(x+\epsilon_1, \phi x+\epsilon_2)$ where $|x| >> |\epsilon_i|$; 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 Figure 3 for an illustration. The Ma-Genton estimator, not being based on linear regression, is not affected by this pattern.

ϕ	Contam. %	n	Estimator	Avg. $\hat{\rho}(1)$	MSE	Avg. $\hat{\rho}(2)$	MSE
-0.2	4	50	OLS	-.19785	.01077	-.02147	.01086
			MG	-.18534	.02753	-.03046	.03823
			L1	-.17678	.00886	-.01231	.00707
			LTS	-.16145	.07133	-.01445	.05180
			MM	-.18117	.00742	-.00452	.00842
		800	OLS	-.19071	.00112	-.00615	.00154
			MG	-.18134	.00169	-.00437	.00191
			L1	-.19446	.00010	.00032	.00011
			LTS	-.18800	.00164	.00205	.00058
			MM	-.19424	.00006	.00028	.00007
	10	50	OLS	-.19866	.01171	-.01778	.01596
			MG	-.18570	.02871	-.06054	.03894
			L1	-.18540	.00228	-.00367	.00309
			LTS	-.16230	.03224	-.00381	.02583
			MM	-.18483	.00187	-.00340	.00314
		800	OLS	-.19148	.00112	-.00318	.00155
			MG	-.17732	.00167	-.00538	.00205
			L1	-.19368	.00004	-.00017	.00006
			LTS	-.19485	.00022	-.00025	.00013
			MM	-.19312	.00002	-.00011	.00004
0.5	4	50	OLS	.34683	.04265	-.05107	.01870
			MG	.36554	.02180	-.05204	.04099
			L1	.42316	.01562	-.02221	.01304
			LTS	.35159	.08751	-.04056	.07510
			MM	.38470	.02550	-.02562	.01032
		800	OLS	.39890	.00067	-.00156	.00148
			MG	.39308	.00119	-.00587	.00252
			L1	.46748	.00475	-.00097	.00014
			LTS	.45444	.01428	-.00032	.00088
			MM	.48818	.00786	-.00121	.00010
	10	50	OLS	.37823	.00939	-.03809	.01739
			MG	.34596	.02506	-.06730	04132
			L1	.43980	.01020	-.00796	.00501
			LTS	.33623	.06072	-.00761	.01634
			MM	.36369	.03569	.00016	.00302
		800	OLS	.39977	.00083	-.00338	.00181
			MG	.39091	.00120	-.00774	.00246
			L1	.47008	.00501	.00072	.00007
			LTS	.49193	.01064	.00257	.00022
			MM	.48947	.00805	.00006	.00004
0.8	4	50	OLS	.46616	.01131	-.04233	.03611
			MG	.46974	.01749	-.05979	.03702
			L1	.55699	.03682	-.00934	.01905
			LTS	.43306	.10956	-.03247	.05134
			MM	.49038	.07561	-.01341	.01176
		800	OLS	.48720	.00054	-.00442	.00168
			MG	.49182	.00087	-.01179	.00261
			L1	.59438	.01447	-.00013	.00013
			LTS	.55985	.02922	.00078	.00066
			MM	.68805	.06670	-.00008	.00010
	10	50	OLS	.45878	.00836	-.04923	.01955
			MG	.48799	.01891	-.06083	.04586
			L1	.61845	.04446	-.00939	.00426
			LTS	.46685	.12663	-.01234	.01295
			MM	.50545	.11626	-.00938	.00443
		800	OLS	.48400	.00063	-.00768	.00170
			MG	.51178	.00147	-.00867	.00247
			L1	.63333	.02528	-.00110	.00005
			LTS	.71091	.08715	-.00049	.00014
			MM	.76902	.09312	-.00084	.00004

Table 3: MA(1) simulation results with innovation outliers; averages of 200 trials.

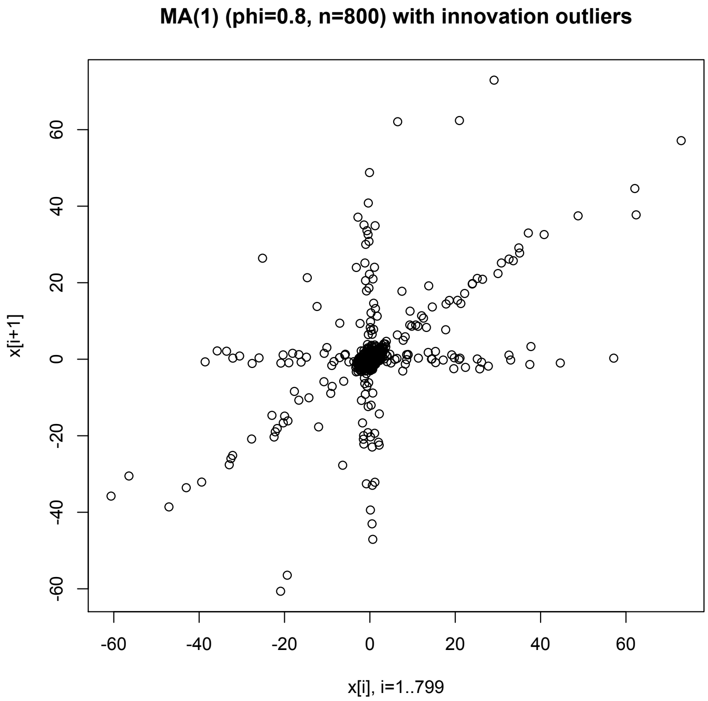

Figure 3: X_t vs. X_{t+1} plot for the MA(1) model $X_t = Z_t + 0.8Z_{t-1}$ with innovation outliers. With an innovation outlier at Z_t, (X_{t-1}, X_t) usually lies on the vertical line, (X_t, X_{t+1}) on the diagonal, and (X_{t+1}, X_{t+2}) on the horizontal. The robust estimators tend to fit the diagonal line.

ϕ	Contam. %	n	Estimator	Avg. $\hat{\rho}(1)$	MSE	Avg. $\hat{\rho}(2)$	MSE
-0.2	4	50	OLS	-.22576	.02227	.02086	.01577
			MG	-.20588	.03829	.00309	.03341
			L1	-.20750	.01018	.02877	.01126
			LTS	-.18120	.06381	.01468	.05601
			MM	-.20344	.00958	.03090	.00692
		800	OLS	-.19775	.00145	.03814	.00091
			MG	-.20515	.00160	.04088	.00161
			L1	-.19949	.00009	.03948	.00010
			LTS	-.19582	.00053	.03674	.00062
			MM	-.19983	.00005	.03962	.00007
	10	50	OLS	-.20993	.01148	.02600	.00906
			MG	-.22251	.03201	.03557	.03274
			L1	-.20522	.00131	.04086	.00157
			LTS	-.19185	.01821	.04927	.01699
			MM	-.20467	.00070	.04440	.00126
		800	OLS	-.19992	.00125	.04020	.00173
			MG	-.21774	.00209	.03777	.00178
			L1	-.20048	.00004	.03959	.00005
			LTS	-.19996	.00016	.03800	.00017
			MM	-.20039	.00002	.03967	.00003

ϕ	Contam. %	n	Estimator	Avg. $\hat{\rho}(1)$	MSE	Avg. $\hat{\rho}(2)$	MSE
0.5	4	50	OLS	.46520	.00956	.19043	.02019
			MG	.48690	.02355	.19364	.04024
			L1	.49198	.00532	.22763	.00791
			LTS	.48511	.02905	.19989	.04247
			MM	.49183	.00377	.23505	.00649
		800	OLS	.49840	.00097	.24964	.00152
			MG	.53888	.00282	.26705	.00255
			L1	.49969	.00006	.25023	.00009
			LTS	.50038	.00039	.25076	.00039
			MM	.49966	.00004	.24984	.00007
	10	50	OLS	.43619	.04085	.17919	.02531
			MG	.55736	.02541	.23512	.03739
			L1	.48964	.00309	.23227	.00662
			LTS	.48506	.00814	.24911	.01338
			MM	.49613	.00106	.24566	.00249
		800	OLS	.49832	.00086	.24440	.00151
			MG	.59379	.00993	.28994	.00367
			L1	.49924	.00003	.24811	.00007
			LTS	.49902	.00012	.24713	.00018
			MM	.49941	.00002	.24885	.00004
0.8	4	50	OLS	.74099	.01184	.53776	.03219
			MG	.81219	.01316	.61055	.03626
			L1	.77752	.00572	.59431	.01536
			LTS	.76933	.02006	.59469	.03543
			MM	.77987	.00425	.59493	.01619
		800	OLS	.79691	.63449	.00037	.00104
			MG	.88504	.00760	.74106	.01129
			L1	.80011	.00003	.63971	.00008
			LTS	.80013	.00016	.64059	.00040
			MM	.80001	.00002	.64043	.00006
	10	50	OLS	.72992	.01731	.53105	.03459
			MG	.89090	.01489	.70164	.02350
			L1	.79232	.00165	.61721	.00596
			LTS	.79450	.00806	.61635	.01611
			MM	.79677	.00068	.62704	.00465
		800	OLS	.79714	.00046	.63659	.00137
			MG	.93719	.01892	.80715	.02847
			L1	.79990	.00001	.63971	.00004
			LTS	.79943	.00008	.64034	.00012
			MM	.80009	.00001	.64031	.00003

Table 4: AR(1) simulation results with innovation outliers; averages of 200 trials.

From Table 4, we can see that our robust regression estimators $\hat{\rho}_{ROBUST}$ 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 4. The Ma-Genton estimator performs about as poorly as OLS here; interestingly, their errors tend to be in opposite directions.

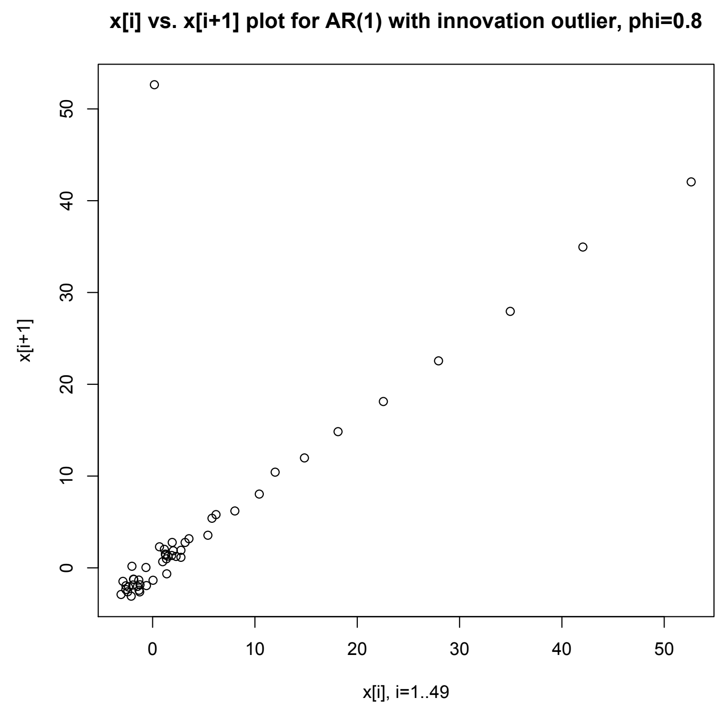

Figure 4: (x_t, x_{t+1}) plot for a realization of length $n = 50$ from the AR(1) time series $X_t = 0.8X_{t-1} + Z_t$ with one innovation outlier.

ϕ_1, ϕ_2	Contam. %	n	Estimator	Avg. $\hat{\rho}(1)$	MSE	Avg. $\hat{\rho}(2)$	MSE
0.5, 0.1	4	50	OLS	.49805	.01063	.30211	.05545
			MG	.58988	.02912	.36280	.03815
			L1	.53168	.00797	.34237	.01120
			LTS	.54746	.03708	.37782	.04575
			MM	.54553	.00746	.35229	.00849
		800	OLS	.55232	.00110	.37241	.00147
			MG	.64522	.00940	.43532	.00545
			L1	.55813	.00034	.37576	.00018
			LTS	.61018	.00749	.38777	.00166
			MM	.57101	.00151	.37671	.00010
	10	50	OLS	.50255	.01400	.30139	.04159
			MG	.70744	.04561	.44201	.03910
			L1	.53896	.00265	.35993	.00599
			LTS	.59708	.01678	.37594	.01517
			MM	.56102	.00395	.36632	.00395
		800	OLS	.55514	.00106	.37448	.00143
			MG	.74715	.03765	.50552	.01800
			L1	.55795	.00021	.37708	.00011
			LTS	.63746	.00934	.38582	.00082
			MM	.55693	.00030	.37730	.00005

ϕ_1, ϕ_2	Contam. %	n	Estimator	Avg. $\hat{\rho}(1)$	MSE	Avg. $\hat{\rho}(2)$	MSE
0.6, 0.3	4	50	OLS	.73869	.02945	.66437	.04485
			MG	.85771	.01835	.79481	.03038
			L1	.83786	.01150	.75807	.02043
			LTS	.85036	.02913	.77909	.03403
			MM	.85324	.01226	.77049	.01785
		800	OLS	.85081	.00060	.80749	.00101
			MG	.96729	.01227	.94209	.01663
			L1	.90582	.00242	.83797	.00066
			LTS	.91584	.00372	.84984	.00160
			MM	.91882	.00383	.84796	.00121
	10	50	OLS	.70326	.04351	.62912	.05943
			MG	.91934	.01062	.86392	.01527
			L1	.84315	.00793	.76536	.01284
			LTS	.88486	.01410	.81898	.01243
			MM	.88306	.00714	.78694	.00987
		800	OLS	.84891	.00076	.80323	.00119
			MG	.98441	.01621	.96065	.02146
			L1	.90649	.00248	.83758	.00062
			LTS	.92019	.00409	.85299	.00166
			MM	.92185	.00421	.84226	.00084

Table 5: AR(2) simulation results with innovation outliers, averages of 200 trials. 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.

Moving on to the AR(2) case of Table 5, we see that with innovation outliers, the L1 and MM robust estimators exhibit much better performance than OLS given a small (50) sample size, but the difference fades with a larger sample size. The Ma-Genton estimator performs 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, innovations are i.i.d. N(0, 1)).

The Ma-Genton and MM-based estimators do the best here; the MM estimator seems to have a slight edge over MG although this is not uniformly evident in the entries of Table 6. 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, and LTS generally acted as a much less efficient MM.

ϕ	n	Contam. Pattern	Estimator	Avg. $\hat{\rho}(1)$	MSE	Avg. $\hat{\rho}(2)$	MSE
-0.2	50	25, 25	OLS	.45142	.42544	-.04117	.00796
			MG	-.15163	.03272	.01986	.03361
			L1	.00949	.04707	-.00093	.00337
			LTS	-.16946	.09823	.02005	.06239
			MM	-.11206	.03400	.01052	.00573
		25, 0	OLS	-.03535	.02868	-.02372	.00719
			MG	-.23079	.03396	.03604	.03247
			L1	-.00350	.03885	-.00308	.00348
			LTS	-.22194	.11734	.04899	.09225
			MM	-.12271	.03509	.01925	.01019
		25, -25	OLS	-.48932	.08450	.00144	.00328
			MG	-.27360	.03155	.01328	.03401
			L1	-.22172	.02783	.00269	.00268
			LTS	-.18993	.09892	.00917	.06258
			MM	-.11939	.04269	.00924	.00569
	800	25, 25	OLS	.21900	.17617	.01230	.00140
			MG	-.20102	.00130	.03990	.00146
			L1	-.12892	.00658	.01522	.00135
			LTS	-.21102	.01470	.04796	.01365
			MM	-.18993	.00220	.01481	.00112
		25, 0	OLS	-.11684	.00753	.02222	.00138
			MG	-.20590	.00128	.03766	.00138
			L1	-.16603	.00271	.01995	.00165
			LTS	-.20086	.01327	.02622	.01303
			MM	-.19773	.00208	.02053	.00118
		25, -25	OLS	-.38096	.03324	.01549	.00167
			MG	-.20358	.00136	.03339	.00181
			L1	-.19913	.00164	.01230	.00178
			LTS	-.19461	.01404	.02058	.01134
			MM	-.18644	.00255	.01274	.00143
0.8	50	25, 25	OLS	.49677	.09497	-.00211	.42308
			MG	.73375	.01678	.48080	.05831
			L1	.71784	.02180	.02922	.38034
			LTS	.76085	.05544	.41964	.17343
			MM	.80826	.03660	.40465	.13308
		25, 0	OLS	.08162	.52154	.04809	.35910
			MG	.69233	.02827	.46751	.06854
			L1	.34407	.25730	.13002	.29226
			LTS	.71216	.07760	.44343	.15611
			MM	.70241	.02940	.42005	.10891
		25, -25	OLS	-.40196	1.44785	.03820	.36352
			MG	.69580	.02682	.50972	.04958
			L1	.07515	.55373	.05980	.34270
			LTS	.73795	.06227	.45480	.13595
			MM	.73087	.01986	.44154	.11602
	800	25, 25	OLS	.68855	.01329	.40271	.05906
			MG	.79444	.00052	.62901	.00148
			L1	.79499	.00071	.59307	.00407
			LTS	.79494	.00556	.62680	.00936
			MM	.79541	.00047	.63165	.00126
		25, 0	OLS	.61925	.03378	.49346	.02354
			MG	.79651	.00048	.63460	.00134
			L1	.78514	.00097	.61991	.00197
			LTS	.80580	.00493	.63532	.00849
			MM	.79848	.00046	.63659	.00125
		25, -25	OLS	.31915	.23308	.39943	.05927
			MG	.79172	.00065	.62866	.00160
			L1	.76533	.00208	.59569	.00394
			LTS	.78979	.00514	.63881	.00996
			MM	.79527	.00050	.63117	.00136

Table 6: AR(1) simulation results with additive outliers; averages of 200 trials. 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.

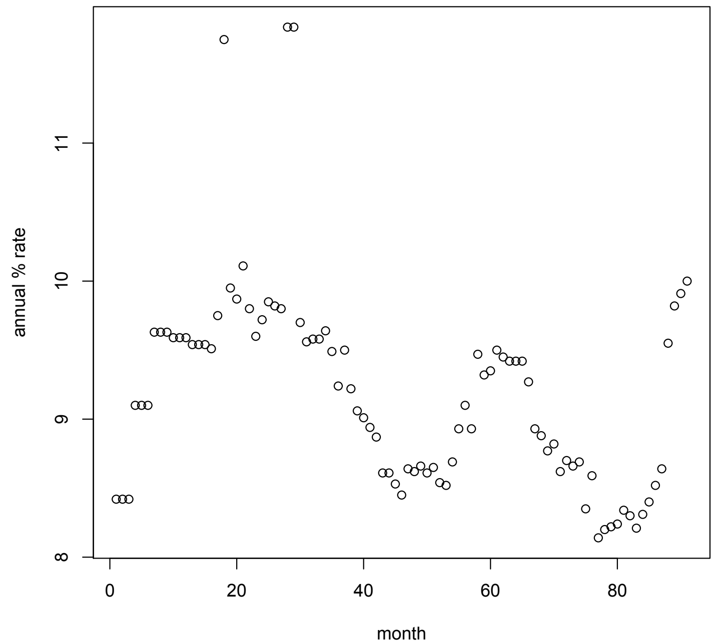

Figure 5: Austrian bank data.

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 data set has been previously discussed and analyzed by Künsch [9], Ma and Genton [12], and Genton and Ruiz-Gazen [8]. The outliers around months 20 and 30 are clearly notable in Figure 5. Table 7 summarizes our findings.

Estimator	Outliers replaced?	$\hat{\rho}(1)$	$\hat{\rho}(3)$	$\hat{\rho}(4)$	$\hat{\rho}(5)$	$\hat{\rho}(6)$	$\hat{\rho}(12)$
OLS	no	.79184	.58923	.51249	.44414	.40440	.08583
	yes	.93911	.77912	.67222	.58084	.49877	.07229
MG	no	.96571	.82703	.73727	.65968	.55046	-.18033
L1	no	.97222	.83459	.78351	.72603	.65957	-.02786
LTS	no	.99451	.95588	.87975	.85556	.36749	-.94203
MM	no	.97194	.81113	.49292	.40119	.34198	.04550

Table 7: Numerical results with Austrian bank data.

The L1 and MM estimates of typical month-to-month and season-to-season autocorrelation are more reasonable than the OLS estimate which is overly affected by the outliers. However, the LTS estimator was erratic, overestimating the low lag autocorrelations and yielding a bizarre value of -.94203 for the 12-month autocorrelation.

7.5 Simulation: Robust AR Model Fitting

Finally, we ran a small simulation to illustrate the performance of Politis' [15] robust AR model fitting method when combined with our robust autocorrelation estimators $\hat{\rho}_{ROBUST}$, i.e., the methodology from Section 6.2. The model was an AR(2) contaminated with innovation outliers.

ϕ_1, ϕ_2	n	Estimator	Avg. $\hat{\phi}(1)$	MSE	Avg. $\hat{\phi}(2)$	MSE
0.5, 0.1	50	OLS	.58223	.04698	-.03085	.04182
		L1	.53079	.01876	.05283	.01510
		LTS	.53561	.04073	.04985	.02820
		MM	.55800	.01390	.03558	.01270
	800	OLS	.62381	.01950	.03664	.00769
		L1	.51531	.00071	.09386	.00004
		LTS	.61376	.02175	.01312	.01310
		MM	.52039	.00074	.08972	.00035

Table 8: AR(2) simulation results with innovation outliers (10 percent frequency, scale 25x normal), averages of 50 (with $n = 800$) or 200 (with $n = 50$) trials.

As we can see in Table 8, the robust AR model fitting method yields reasonable results even when it is applied to the raw sample acf. However, performance was noticeably better with $n = 50$ when combining it with the L1 or MM robust autocorrelation estimators; furthermore, in the $n = 800$ case, the performance advantage was overwhelming. Thus, these two methods (Politis' [15] robust AR fitting and $\hat{\rho}_{ROBUST}$) are not redundant; they complement each other very well.

7.6 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, on a five-year-old MacBook Pro). Thus, unless the data set 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 percent 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 9 below). 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.

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

Table 9: 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. For the $n = 80$ case, we processed 20000 series and divided the final times by 100. R and C source code is posted at [url].

8 Conclusions

A class of robust autocorrelation estimators $\hat{\rho}_{ROBUST}$ 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.

8.1 Acknowledgements

Many thanks are due to Tim McMurry for helpful R tips, and to Marc Genton for sharing with us the Austrian bank data.

References

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

[2] Davis, R.A. and Mikosch, T., "The sample autocorrelations of financial time series models," *Nonlinear and Nonstationary Signal Processing*, Cambridge University Press, Cambridge, pp. 247–274, 2000.

[3] Denby, L. and Martin, R.D., 'Robust estimation of the first-order autoregressive parameter,' *Journal of the American Statistical Association*, vol. 74, no. 365, pp. 140–146, 1979.

[4] Hampel, F.R. 'Robust estimation, a condensed partial survey', *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 27, pp. 87–104, 1973.

[5] Huber, P.J. 'Robust regression: Asymptotics, conjectures and Monte Carlo', *Annals of Statistics*, vol. 1, pp. 799–821, 1973.

[6] 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,* vol. 65, 81-94.

[7] Genton, M. G., and Lucas, A. (2005), 'Discussion of the paper "Breakdown and groups" by L. Davies and U. Gather', *Annals of Statistics,* vol. 33, 988-993.

[8] Genton, M. G., and Ruiz-Gazen, A. (2010), 'Visualizing influential observations in dependent data,' it Journal of Computational and Graphical Statistics, vol. 19, 808-825.

[9] Künsch, H. 'Infinitesimal robustness for autoregressive processes', *Ann. Statist.*, vol. 12, pp. 843–863, 1984.

[10] Kreiss, J.-P., and Paparoditis, E. 'Bootstrap methods for dependent data: A review.' *Journal of the Korean Statistical Society*, vol. 40, no. 4, pp. 357-378, 2011.

[11] Kreiss, J.-P., Paparoditis, E. and Politis, D.N. 'On the Range of Validity of the Autoregressive Sieve Bootstrap', *Ann. Statist.*, vol. 39, No. 4, pp. 2103-2130, 2011.

[12] Ma, Y. and Genton, M.G., "Highly robust estimation of the autocovariance function," *Journal of Time Series Analysis*, vol. 21, no. 6, pp. 663–684, 2000.

[13] Maronna, R.A., Martin, R.D., and Yohai, V.J., *Robust Statistics: Theory and Methods*, Wiley, 2006.

[14] McMurry, T. and Politis, D.N., "Banded and tapered estimates of autocovariance matrices and the linear process bootstrap," *Journal of Time Series Analysis*, vol. 31, pp. 471-482, 2010. [Corrigendum: vol. 33, 2012.]

[15] Politis, D.N., "An algorithm for robust fitting of autoregressive models," *Economics Letters*, vol. 102, no. 2, pp. 128–131, 2009.

[16] Politis, D.N., 'Higher-order accurate, positive semi-definite estimation of large-sample covariance and spectral density matrices', *Econometric Theory*, vol. 27, no. 4, pp. 703-744, 2011.

[17] Politis, D.N., Romano, J.P. and Wolf, M. *Subsampling*, Springer, New York, 1999.

[18] Rousseeuw, P.J. and Leroy, A.M., *Robust Regression and Outlier Detection*, John Wiley & Sons, New York, 1987.

[19] Wu, W.B. and Pourahmadi, M., "Banding sample autocovariance matrices of stationary processes," *Statistica Sinica*, vol. 19, no. 4, pp. 1755–1768, 2009.

[20] Yohai, V.J., "High breakdown-point and high efficiency estimates for regression," *The Annals of Statistics*, vol. 15, pp. 642–656, 1987.