Taylor & Francis
Taylor & Francis Group

Communications in Statistics - Simulation and Computation

ISSN: 0361-0918 (Print) 1532-4141 (Online) Journal homepage: https://www.tandfonline.com/loi/lssp20

Heteroscedasticity-robust estimation of autocorrelation

E. Reschenhofer

To cite this article: E. Reschenhofer (2019) Heteroscedasticity-robust estimation of autocorrelation, Communications in Statistics - Simulation and Computation, 48:4, 1251-1263, DOI: 10.1080/03610918.2017.1408826

To link to this article: https://doi.org/10.1080/03610918.2017.1408826

Published online: 16 Jan 2018.

COMMUNICATIONS IN STATISTICS—SIMULATION AND COMPUTATION®
2019, VOL. 48, NO. 4, 1251–1263
https://doi.org/10.1080/03610918.2017.1408826

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Heteroscedasticity-robust estimation of autocorrelation

E. Reschenhofer

Department of Statistics and Operations Research, Faculty of Business, Economics and Statistics, University of Vienna, Oskar-Morgenstern-Platz 1, Vienna, Austria

ABSTRACT
This paper proposes estimators of the first-order autocorrelation that are based on suitably transformed ratios of successive observations. The new estimators are given by simple functions of the observations. Numerical optimization is not required. Simulations show that they are highly robust against extreme values and clusters of high volatility and are therefore particularly useful for the estimation of serial correlation in return series. Besides, the results of the simulation study also call into question the common practice of correcting the small-sample bias of conventional estimators.

ARTICLE HISTORY
Received 17 August 2017
Accepted 14 November 2017

KEYWORDS
Truncated Cauchy distribution; Location parameter; Bias correction; Serial correlation; Return series

MSC
62F10; 62M10

1. Introduction

Given n consecutive observations $y_1, \dots, y_n$ from a zero-mean stationary process, the standard estimator of the first-order autocorrelation ρ is the least squares (LS) estimator

$$\widehat{\rho} = \frac{\sum_{t=2}^{n} y_t y_{t-1}}{\sum_{t=1}^{n-1} y_t^2}, \tag{1}$$

which is neither mean-unbiased nor median-unbiased. Many methods were proposed to reduce the finite-sample bias of this estimator (see, e.g., Bartlett 1946; Quenouille, 1949; Hurwicz 1950; Kendall 1954; Marriott and Pope 1954; White 1961; Andrews, 1993). However, it is a priori not clear whether bias correction is a good thing. MacKinnon and Smith (1998) showed that reducing the bias of an estimator may increase its variance or even its mean squared error (MSE). Things get even more complicated when we look at the whole distribution of an estimator rather than just focus only on its bias and variance. Since the LS estimator $\widehat{\rho}$ of ρ can take values outside the open interval $(-1,1)$ with positive probability, its risk will be infinite if the squared error loss function is replaced by a loss function that is based on the Kullback–Leibler divergence (Kullback and Leibler 1951; Schroeder and Zielinski 2010). However, this problem can easily be fixed by using the closely related estimator

$$\widehat{\rho}_B = \frac{2\sum_{t=2}^{n} y_t y_{t-1}}{\sum_{t=1}^{n-1} y_t^2 + \sum_{t=2}^{n} y_t^2} \tag{2}$$

(Burg 1967, 1975) instead of the LS estimator. But there are more serious problems. Let us assume, for illustration, that ρ is large and the sample size n is small. In this case, the distribution of the LS estimator (and related estimators) is extremely skewed to the left. While the median and the mean of this distribution are less than ρ, its mode is greater than ρ. Bias

CONTACT E. Reschenhofer erhard.reschenhofer@univie.ac.at Department of Statistics and Operations Research, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria.

correction can easily be achieved by multiplying the estimator by a suitable constant greater than 1. Unfortunately, this transformation has an undesirable side effect. The peak of the sampling distribution moves further away from ρ, which reduces the probability of obtaining an estimate reasonably close to the true value and also increases the probability of obtaining an estimate greater than 1.

In macroeconomic and financial applications, where sample sizes are usually large, the correction of a small-sample bias is of lower priority. More pressing is the need to deal with heteroscedasticity. Robustness against heteroscedasticity can be achieved by using a rolling-window approach. Unfortunately, there are two conflicting objectives to trade off when choosing the length of the estimation window, robustness versus accuracy. This paper tries to offer a way out of this predicament by designing estimators that are functions of the ratios

$$x_t = \frac{y_t}{y_{t-1}} \tag{3}$$

of two successive observations. With regard to robustness against heteroscedasticity, the use of these ratios corresponds to a choice of an estimation window of minimal length 2. There is no further need of an estimation window. All ratios can be used at the same time, hence accuracy is not corrupted.

The next two sections deal with bias-corrected estimators and estimators based on the ratios (3), respectively. In Section 4, the results of a simulation study are presented which compares the new estimators proposed in this paper with conventional estimators. The simulation study addresses both the bias in small samples and the robustness issue in large samples. Section 5 concludes.

2. Bias correction

Let $y_1, \ldots, y_n$ be observations of a Gaussian AR(1) process with zero mean and innovation variance σ^2. The LS estimator is also the conditional maximum likelihood (ML) estimator obtained by maximization of the log likelihood

$$\log(f(y_2, \ldots, y_n | y_1; \rho, \sigma^2)) = -\frac{n-1}{2}\log(2\pi\sigma^2) - \sum\nolimits_{t=2}^{n} \frac{(y_t - \rho y_{t-1})^2}{2\sigma^2} \tag{4}$$

conditioned on the first observation. The loss in efficiency due to the omission of the first observation is negligible if n is large. The conditional ML estimator has the same asymptotic distribution as the exact ML estimator and is therefore asymptotically efficient. The greatest disadvantage of the exact ML estimator is that the likelihood

$$f(y_1, \ldots, y_n; \rho, \sigma^2) = f(y_1; \rho, \sigma^2) f(y_2, \ldots, y_n | y_1; \rho, \sigma^2) \tag{5}$$

requires numerical maximization. There exists no simple solution in terms of $y_1, \ldots, y_n$.

Moreover, the log marginal density

$$\log(f(y_1; \rho, \sigma^2)) = -\frac{1}{2}\log\left(\frac{2\pi\sigma^2}{1-\rho^2}\right) - \frac{y_1^2(1-\rho^2)}{2\sigma^2} \tag{6}$$

is undefined in the unstable case when $|\rho| \geq 1$.

It is well known that the LS estimator is biased in finite samples. Its mean is to order $1/n^3$ given by

$$E\left(\widehat{\rho}\right) = \left(1 - \frac{2}{n} + \frac{4}{n^2} - \frac{2}{n^3}\right)\rho + \frac{2}{n^2}\left(\rho^3 + \rho^5\right) + \ldots, \tag{7}$$

(White 1961; for earlier results see Bartlett 1946; Hurwicz 1950; Kendall 1954; Marriott and Pope 1954; for the unstable case see Le Breton and Pham 1989). Assuming that the bias is approximately proportional to $1/n$, Quenouille (1949) proposed to reduce the bias to order $1/n^2$ simply by calculating the sample correlation coefficient not only for the whole sample but also for the first and second half separately. Indeed, the mean of his estimator

$$\widehat{\rho}_Q = 2\widehat{\rho} - \frac{1}{2}\left(\widehat{\rho}_1 + \widehat{\rho}_2\right) \tag{8}$$

is given by

$$E\left(\widehat{\rho}_Q\right) = 2\left(1 - \frac{2}{n}\right)\rho - \frac{1}{2}\left(1 - \frac{2}{\frac{n}{2}} + 1 - \frac{2}{\frac{n}{2}}\right)\rho + O\left(\frac{1}{n^2}\right) = \rho + O\left(\frac{1}{n^2}\right). \tag{9}$$

However, Orcutt and Winokur (1969) found in a Monte Carlo study that a direct application of the approximation of $E(\widehat{\rho})$ to the power of $1/n$ (Kendall 1954; Marriott and Pope 1954) yields an estimator with a smaller mean square error than Quenouille's (1949) estimator.

In the present case, where the true mean is known to be zero, the approximation equation

$$E\left(\widehat{\rho}_Q\right) = \left(1 - \frac{2}{n}\right)\rho + O\left(\frac{1}{n^2}\right) \tag{10}$$

of Marriott and Pope (1954) implies that the bias of the estimator

$$\widehat{\rho}_{MP} = \left(1 + \frac{2}{n-2}\right)\widehat{\rho} \tag{11}$$

is only of order $1/n^2$ and the estimator is therefore less biased than the LS estimator. A further reduction of the bias can be achieved by using White's (1961) approximation of the expected value. Substituting $\widehat{\rho}$ for $E(\widehat{\rho})$ in this approximation yields

$$\widehat{\rho} \approx w\left(\rho\right) = \gamma\rho + \psi\rho^3 + \psi\rho^5, \tag{12}$$

where

$$\lambda = 1 - \frac{2}{n} + \frac{4}{n^2} - \frac{2}{n^3},\ \psi = \frac{2}{n^2},$$

and solving for ρ yields

$$w^{-1}\left(\widehat{\rho}\right) \approx \rho. \tag{13}$$

Clearly, this approximation is useful only if the value of $\widehat{\rho}$ is close to $E(\widehat{\rho})$. The inverse function of w can be determined numerically to any desired precision. Using the standard formulas for series reversion obtained by plugging w^{-1} into w and equating the first seven coefficients (see, e.g., Abramowitz and Stegun 1970), we obtain

$$\widehat{\rho}_W = \frac{1}{\gamma}\widehat{\rho} - \frac{\psi}{\gamma^4}\widehat{\rho}^3 + \frac{3\psi^2 - \gamma\psi}{\gamma^7}\widehat{\rho}^5 + \frac{8\gamma\psi^2 - 12\psi^3}{\gamma^{10}}\widehat{\rho}^7. \tag{14}$$

Andrews (1993) considered more general AR(1) models that allow for an intercept, a time trend and even a unit root. He proposed a bias correction method which replaces the LS estimate by that value which implies a median equal to the LS estimate. To establish the unbiasedness of Andrews' estimator $\hat{\rho}_A$ in the stable case, where $|\rho| < 1$, we use Burg's (1967, 1975) estimator $\widehat{\rho_B}$ instead of the LS estimator for the definition of $\hat{\rho}_A$ because it only takes values between −1 and 1. Since the median function

$$m(\rho) = med(\widehat{\rho_B}; \rho) \tag{15}$$

of $\widehat{\rho_B}$ is strictly increasing on the parameter space (−1,1), we have

$$\hat{\rho}_A = m^{-1}(\widehat{\rho_B}) \tag{16}$$

and

$$\rho = m^{-1}m(\rho) = m^{-1}(med(\widehat{\rho_B}; \rho)) = med(m^{-1}(\widehat{\rho_B}); \rho) = med(\widehat{\rho_A}; \rho) \tag{17}$$

for all $\rho \in (-1, 1)$. The estimator $\widehat{\rho_A}$ is therefore a median-unbiased estimator. Andrews (1993) pointed out that the properties of his estimator depend on the specification of the distribution of the innovations (normality, homoskedasticity) and of the form of autocorrelation (first-order AR). However, perhaps the most serious drawback is that both the median function m and its inverse must be evaluated numerically. Tables of values of m are provided only for a very limited number of scenarios. Also Tanizaki (2000), who extended Andrew's estimator to higher-order AR models, used Monte Carlo techniques rather than analytical techniques.

3. Using cauchy distributed ratios

Looking for a more convenient alternative to Andrew's (1993) median-unbiased estimator, Zieliński (1999) revisited Hurwicz's (1950) proposal to use the median of the ratios of successive observations as an estimator of ρ. He was able to prove Hurwicz's (1950) conjecture that the estimator

$$\widehat{\rho_H} = med\left(\frac{y_2}{y_1}, \ldots, \frac{y_n}{y_{n-1}}\right) \tag{18}$$

is median-unbiased. The reason for Hurwicz (1950) to consider the median rather than the mean is that $x_t = y_t/y_{t\text{-}1}$ is the ratio of two centered normal variables with corr$(y_t, y_{t\text{-}1}) = \rho$ and is therefore a Cauchy variable with density

$$f(x; \rho) = \frac{1}{\pi\theta}\frac{\theta^2}{(x-\rho)^2 + \theta^2} = \frac{\sqrt{1-\rho^2}}{\pi}\frac{1}{x^2 - 2\rho x + 1}, \tag{19}$$

where the location parameter $\rho\sigma_t/\sigma_{t\text{-}1} = \rho$ specifies both the median and the mode and the scale parameter

$$\theta = \frac{\sqrt{1-\rho^2}\sigma_t}{\sigma_{t-1}} = \sqrt{1-\rho^2} \tag{20}$$

specifies the interquartile range (see Jamnik 1971; for the distribution of the ratio of any two jointly normal variables see Cedilnik et al. 2004; Marsaglia 2006). Since the Cauchy distribution does not have any finite moments, the sample mean cannot be used for the estimation of ρ. The sample median is an obvious alternative (for other estimators based on sample order statistics see Rothenberg et al. 1964; Bloch 1966), but its asymptotic relative efficiency (A.R.E.)

is already relatively low (81%) in the simplest case ($\theta = 1, \rho = 0$) and decreases further as $|\rho|$ increases. We must therefore replace the median by a more efficient estimator if we want to keep using the ratios $x_t = y_t/y_{t\text{-}1}$. Their main advantage is that they are not affected by (conditional) heteroscedasticity as long as the (conditional) variances of successive observations $y_{t\text{-}1}$ and y_t are approximately of the same size.

Pretending that the ratios are i.i.d., we may use a (quasi-) ML approach. The derivative of

$$\log\left(f\left(x;\rho\right)\right) = -\log\left(\pi\right) + \frac{1}{2}\log\left(1-\rho^2\right) - \log\left(x^2 - 2\rho x + 1\right) \tag{21}$$

with respect to ρ can be written as

$$-\frac{\rho}{1-\rho^2} + \frac{2x}{x^2 - 2\rho x + 1} = -\frac{\rho}{1-\rho^2} + \frac{2\frac{1}{x}}{\left(\frac{1}{x}\right)^2 - 2\rho\frac{1}{x} + 1}, \tag{22}$$

hence the ML estimator $\widehat{\rho_1}$ based on the random sample $x_2, \ldots, x_n$ remains unchanged when each observation x_t with absolute value greater than 1 is replaced by its inverse $1/x_t$. As far as ML estimation is concerned, there is no difference between the original sample and the new sample. However, there is a huge difference when it comes to calculating the sample mean or other sample moments.

To overcome the main weakness of the Cauchy distribution, the nonexistence of its moments, Nadarajah and Kotz (2006) introduced a truncated version with probability density

$$f^*\left(z\right) = \frac{1}{\theta}\frac{1}{D}\left\{1 + \left(\frac{z-\rho}{\theta}\right)^2\right\}^{-1} = \frac{1}{D}\frac{\theta}{\theta^2 + \left(z-\rho\right)^2}, \quad -\infty < A \le z \le B < \infty, \tag{23}$$

where

$$D = arctan\left(\beta\right) - arctan\left(\alpha\right), \; \alpha = \frac{A-\rho}{\theta}, \;\; \beta = \frac{B-\rho}{\theta}, \tag{24}$$

and ρ and θ are the location parameter and the scale parameter, respectively. Using the ordinary hypergeometric function represented by the hypergeometric series

$$F\left(a, b, c, v\right) = 1 + \frac{a\cdot b}{c\cdot 1}v + \frac{a\left(a+1\right)b\left(b+1\right)}{c\left(c+1\right)\cdot 1\cdot 2}v^2 + \ldots, \tag{25}$$

they derived explicit expressions

$$E(Z^k) = \frac{\rho^k}{D}\sum\nolimits_{j=0}^{k}\frac{1}{j+1}\binom{k}{j}\left(\frac{\theta}{\rho}\right)^j\left\{\beta^{j+1}F\left(1, \frac{j+1}{2}, 1+\frac{j+1}{2}, -\beta^2\right)\right.$$
$$\left. -\,\alpha^{j+1}F\left(1, \frac{j+1}{2}, 1+\frac{j+1}{2}, -\alpha^2\right)\right\} \tag{26}$$

for the noncentral moments.

In the case of the ratios of successive observations of a Gaussian AR(1) process, we have $A = -1$ and $B = 1$. Since $y_t/y_{t\text{-}1}$ and $y_{t\text{-}1}/y_t$ have the same distribution, we might wish to select always that one with the smaller absolute value. All values would then be between -1 and 1 and the new density would just be two times the old density on the interval [−1,1]. The location parameter would still be the mode of the distribution. Clearly, we cannot expect that some nonparametric estimator of the mode (e.g., the half sample mode; see Robertson-Cryer

1974; Bickel and Frühwirth 2006), which is based on the "truncated" ratios

$$z_t = sign(y_{t-1}y_t)\frac{\min(|y_{t-1}|, |y_t|)}{\max(|y_{t-1}|, |y_t|)}, \tag{27}$$

is more accurate than the sample median of the original ratios $x_t = y_t/y_{t\text{-}1}$. Again, a (quasi-) ML approach makes more sense. As pointed out at the end of the previous section, it does not make any difference whether the original ratios x_t or the "truncated" ratios z_t are used in the maximization. But the truncated Cauchy distribution has other advantages. Since it is defined over a finite interval, it has all its moments. We may therefore also put the sample mean to use.

In general, the sample mean is a biased estimator of the mode of a truncated Cauchy distribution. We must therefore take care of the bias. The first step is to derive a simple expression for the expected value. For $A = -1$ and $B = 1$, we have

$$-\alpha = \frac{1+\rho}{\theta} = \frac{1+\rho}{\sqrt{1-\rho^2}} = \frac{\sqrt{1+\rho}}{\sqrt{1-\rho}} = \frac{\sqrt{1-\rho^2}}{1-\rho} = \frac{1}{\beta} \tag{28}$$

and

$$D = arctan(\beta) - arctan(\alpha) = arctan(\beta) + arctan(-\alpha) = arctan(\beta) + arctan\left(\frac{1}{\beta}\right) = \frac{\pi}{2} \tag{29}$$

Noting that $F(a,b,c,v) = F(b,a,c,v)$ and using further properties of F (see Gradshteyn and Ryzhik 2007, pp. 1006–1007), we obtain

$$\begin{aligned} E\,(Z) &= \frac{\rho}{D}\Bigg[\beta F\left(\frac{1}{2}, 1, \frac{3}{2}, -\beta^2\right) - \alpha F\left(\frac{1}{2}, 1, \frac{3}{2}, -\alpha^2\right) \\ &\quad + \frac{\theta}{2\rho}\left\{\beta^2 F(1,1,2,-\beta^2) - \alpha^2 F(1,1,2,-\alpha^2)\right\}\Bigg] \\ &= \frac{\rho}{D}\left[\beta\frac{\arctan(\beta)}{\beta} - \alpha\frac{\arctan(\alpha)}{\alpha} + \frac{\theta}{2\rho}\left\{\beta^2\frac{\log(1+\beta^2)}{\beta^2} - \alpha^2\frac{\log\left(1+\alpha^2\right)}{\alpha^2}\right\}\right] \\ &= \rho + \frac{\theta}{\pi}\log\left(\frac{1+\beta^2}{1+\frac{1}{\beta^2}}\right) = \rho + \frac{\theta}{\pi}\log(\beta^2) = \rho + \frac{\sqrt{1-\rho^2}}{\pi}\log\left(\frac{1-\rho}{1+\rho}\right) = h\,(\rho)\,. \end{aligned} \tag{30}$$

Thus, the estimator

$$\widehat{\rho_2} = h^{-1}(\bar{Z}) \tag{31}$$

will take a value close to ρ if the value of $\bar{Z}$ is close to $E(Z)$. The inverse function of h can be approximated by an odd polynomial of sufficiently high order or simply by

$$\widehat{\rho_3} = \hat{h}^{-1}(\bar{Z}) = -1 + 2\Phi_{0.295}(\bar{Z}) \tag{32}$$

(see Figure 1), where Φ_σ denotes the cumulative distribution function of a normal distribution with mean 0 and standard deviation σ.

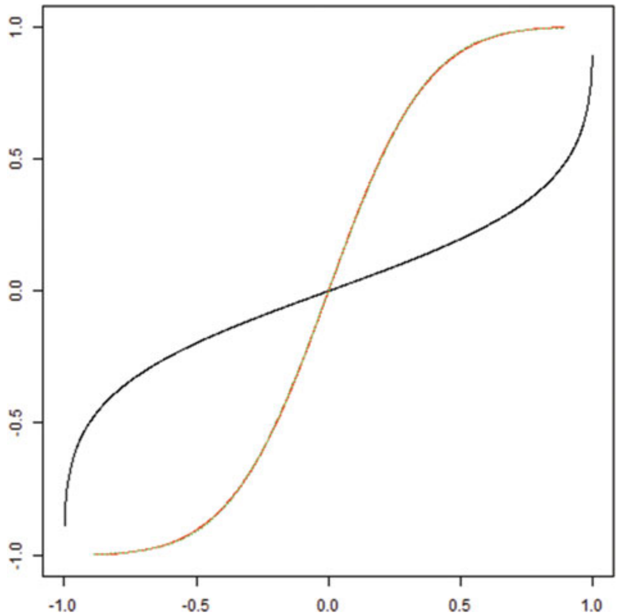

Figure 1. Graphs of the function h (black), its inverse function h^{-1} (red), and the simple approximation $\hat{h}^{-1}$ (green) of h^{-1}.

4. Simulations

In this section, the new estimators $\widehat{\rho}_W$ (14), $\widehat{\rho}_1$ (ML estimator based on ratios), $\widehat{\rho}_2$ (31), $\widehat{\rho}_3$ (32) of the first- order autocorrelation ρ are compared with the existing estimators $\widehat{\rho}_{ML}$ (ML estimator based on observations), $\widehat{\rho}$ (1), $\widehat{\rho}_B$ (2), $\widehat{\rho}_{CC}$ (33), $\widehat{\rho}_{MP}$ (11), $\widehat{\rho}_H$ (18) both in case of a standard AR(1) model and in case of deviations from the standard model. In the latter case, our focus is on nonnormality and heteroscedasticity. The sample counterpart

$$\widehat{\rho}_{CC} = \frac{\sum_{t=1}^{n-1} y_t y_{t+1}}{\sqrt{\sum_{t=1}^{n-1} y_t^2 \sum_{t=2}^{n} y_t^2}} \tag{33}$$

of the correlation coefficient between y_t and y_{t-1} is also included in the comparison. Of particular interest is the performance of the simple robust estimators $\widehat{\rho}_2$ and $\widehat{\rho}_3$.

We consider the ARMA(p,q)-GARCH(r,s) model

$$y_t = \sum_{k=1}^{p} \phi_k y_{t-k} + \sum_{k=1}^{q} \theta_k u_{t-k} + u_t, \tag{34}$$

where

$$u_t = \sigma_t z_t,$$
$$\sigma_t^2 = \alpha_0 + \sum_{k=1}^{r} \alpha_k u_{t-k}^2 + \sum_{k=1}^{s} \beta_k \sigma_{t-k}^2.$$

In the simulations, the ARMA order (p,q) is either (0,0) or (1,0), the first-order autocorrelation ρ is 0 if $(p,q) = (0,0)$ or $\rho = \phi_1 = 0.8$ if $(p,q) = (1,0)$, the GARCH order (r,s) is either (0,0) or (1,1), the GARCH parameters are $\alpha_0 = 1$, $\alpha_1 = 0.1$, $\beta_1 = 0.8999$, and the GARCH innovations z_t are either i.i.d. $N(0,1)$ or $t(3)$. For each case, 1,000,000 time series of length

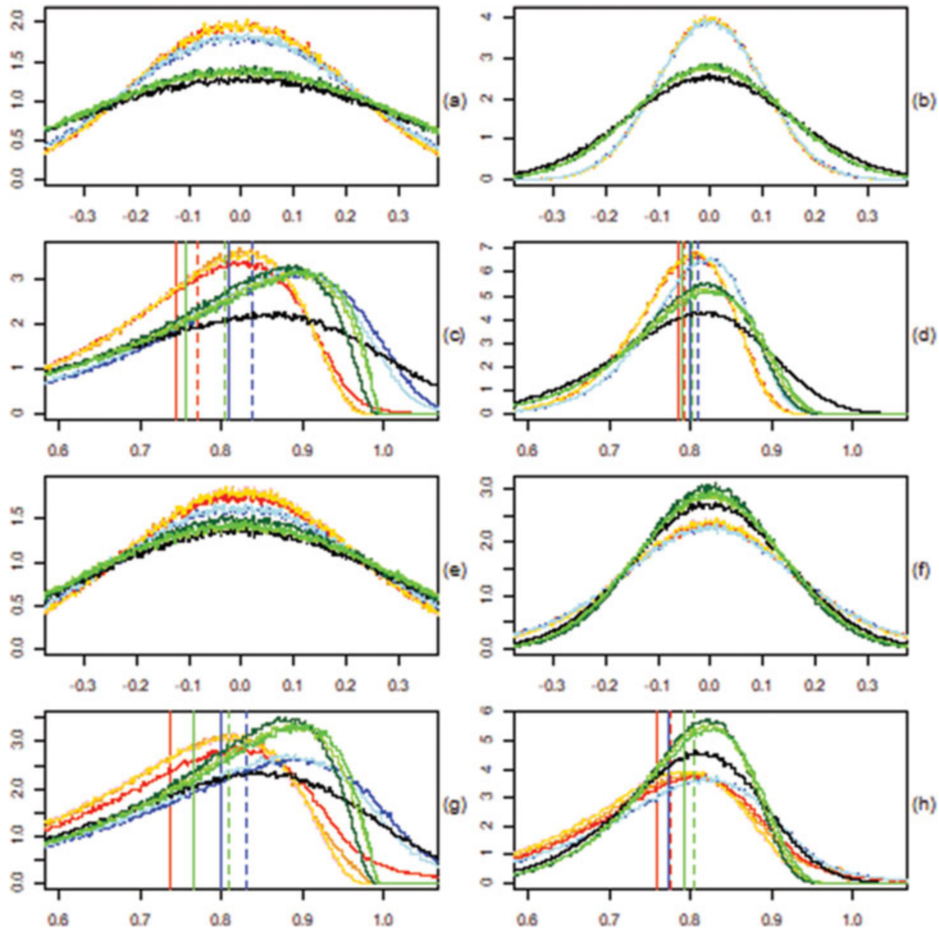

Figure 2. Sampling distributions of the estimators $\widehat{\rho}_{ML}$ (orange), $\widehat{\rho}$ (red), $\widehat{\rho}_B$ (pink), $\widehat{\rho}_{CC}$ (gold), $\widehat{\rho}_{MP}$ (blue), $\widehat{\rho}_W$ (lightblue), $\widehat{\rho}_H$ (black), $\widehat{\rho}_1$ (darkgreen), $\widehat{\rho}_2$ (green), $\widehat{\rho}_3$ (gray), with means (vertical lines) and medians (dotted lines) under different assumptions, Gaussian white noise: (a) $\rho = 0$, $n = 25$, (b) $\rho = 0$, $n = 100$, Gaussian AR(1): (c) $\rho = 0.8$, $n = 25$, (d) $\rho = 0.8$, $n = 100$, GARCH(1,1) with $t(3)$ innovations: (e) $\rho = 0$, $n = 25$,, (f) $\rho = 0$, $n = 100$, AR(1)-GARCH(1,1) with $t(3)$ innovations: (g) $\rho = 0.8$, $n = 25$, (h) $\rho = 0.8$, $n = 100$.

$n = 25$ and $n = 100$, respectively, are generated. All computations are carried out with the free statistical software R (R Core Team 2017).

The densities of the sampling distributions of the competing estimators of ρ are estimated by histograms with bin-width 0.002. Smoothing, e.g., by kernel density estimators, is not appropriate because it obscures what happens at the boundaries of the stable region. Figures 2 and 3 show the histograms and boxplots, respectively. The sampling distributions are very similar within certain groups of estimators. The first group consists of the ML estimator $\widehat{\rho}_{ML}$, the LS estimator $\widehat{\rho}$, Burg's (1967, 1975) estimator $\widehat{\rho}_B$, and the sample autocorrelation $\widehat{\rho}_{CC}$. The only noteworthy difference is that the LS estimator takes values outside the critical region, which is, of course, most noticeable when ρ is large, n is small, and the innovations come from a fat-tailed distribution (see Figure 2.g). Since there is practically no difference between $\widehat{\rho}_B$ and $\widehat{\rho}_{CC}$, the latter will be omitted from further analysis. The second group consists of Marriott and Pope's (1954) estimator $\widehat{\rho}_{MP}$ and the estimator $\widehat{\rho}_W$, which is based on White's (1961) approximation of the bias of $\widehat{\rho}$. If $\rho = 0$, the bias is not an issue (see Figures 2.a, 2.b, 2.e, 2.f). But if $\rho = 0.8$, these two "bias-corrected" estimators are indeed much less biased

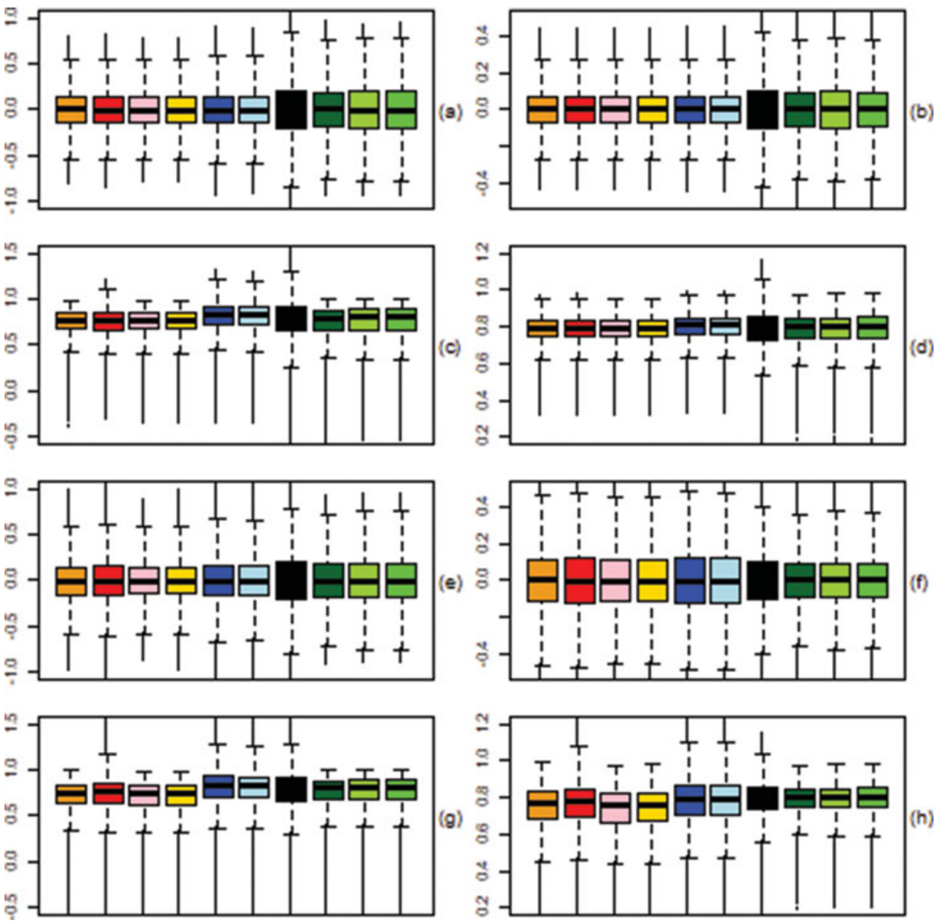

Figure 3. Boxplots of the estimators $\widehat{\rho}_{ML}$ (orange), $\widehat{\rho}$ (red), $\widehat{\rho}_B$ (pink), $\widehat{\rho}_{CC}$ (gold), $\widehat{\rho}_{MP}$ (blue), $\widehat{\rho}_W$ (lightblue), $\widehat{\rho}_H$ (black), $\widehat{\rho}_1$ (darkgreen), $\widehat{\rho}_2$ (green), $\widehat{\rho}_3$ (gray) under different assumptions, Gaussian white noise: (a) $\rho = 0, n = 25$, (b) $\rho = 0, n = 100$, Gaussian AR(1): (c) $\rho = 0.8, n = 25$, (d) $\rho = 0.8, n = 100$, GARCH(1,1) with $t(3)$ innovations: (e) $\rho = 0, n = 25$,, (f) $\rho = 0, n = 100$, AR(1)-GARCH(1,1) with $t(3)$ innovations: (g) $\rho = 0.8, n = 25$, (h) $\rho = 0.8, n = 100$.

than the LS estimator. However, this improvement comes with a price. The mode of the sampling distribution moves further away from ρ towards the edge of the stable region, which increases the probability of obtaining an estimate greater than 1 and reduces the probability of obtaining an estimate reasonably close to the true value (see Figures 2.c and 2.g). Because of the high agreement between $\widehat{\rho}_{MP}$ and $\widehat{\rho}_W$, only the former will be considered below. The remaining estimators $\widehat{\rho}_H$, $\widehat{\rho}_1$, $\widehat{\rho}_2$, $\widehat{\rho}_3$ are all based on ratios of successive observations. While the first one is never among the best, the other three are the best estimators under conditional heteroscedasticity provided that n is not too small (see Figures 2.f and 2.h), hence they are particularly useful for financial applications. In the following, $\widehat{\rho}_2$ will be omitted because its sampling distribution is very similar to that of $\widehat{\rho}_3$ but the latter is much simpler.

In addition to the plots of the sampling distributions shown in Figure 2, Table 1 summarizes important characteristics of these distributions such as their means, medians, modes (estimated with the help of the function *hsm* of the R package *modeest*), root-mean-square errors (RMSE), mean absolute deviations (MAD), and the probabilities that the value of the estimator falls into specific intervals. These numbers corroborate our interpretation of Figure 2,

Table 1. Comparison of various estimators of the first-order autocorrelation ρ under different assumptions.

Gaussian white noise: (a) $\rho = 0, n = 25$, (b) $\rho = 0, n = 100$								
Gaussian AR(1): (c) $\rho = 0.8, n = 25$, (d) $\rho = 0.8, n = 100$								
		$\hat{\rho}_{ML}$	$\hat{\rho}$	$\hat{\rho}_B$	$\hat{\rho}_{MP}$	$\hat{\rho}_H$	$\hat{\rho}_1$	$\hat{\rho}_3$
(a)	Mean	−0.00009	−0.00007	−0.00009	−0.00008	−0.00038	−0.00029	−0.00024
	Median	0.00000	−0.00020	−0.00019	−0.00021	−0.00002	0.00000	−0.00019
	Mode	0.01100	0.01345	0.01358	0.01462	0.01697	0.00800	0.00436
	SD	0.19730	0.19701	0.19642	0.21414	0.30949	0.27303	0.27182
	RMSE	0.19730	0.19701	0.19642	0.21414	0.30949	0.27303	0.27182
	MAD	0.15887	0.15863	0.15832	0.17242	0.24758	0.22126	0.22113
	$\in[1,\infty)$	0.00000	0.00000	0.00000	0.00000	0.00132	0.00000	0.00000
	$\in[-0.1,0.1]$	0.37988	0.37877	0.37882	0.35032	0.25046	0.27525	0.27117
	$\in[-0.05,0.05]$	0.19579	0.19402	0.19412	0.17877	0.12670	0.14008	0.13729
(b)	Mean	0.00018	0.00018	0.00018	0.00018	0.00008	0.00027	0.00032
	Median	0.00000	0.00023	0.00023	0.00024	−0.00013	0.00000	0.00011
	Mode	0.00200	0.00465	0.00382	0.00475	−0.00037	0.00000	−0.00015
	SD	0.09949	0.09948	0.09946	0.10151	0.15619	0.14019	0.13988
	RMSE	0.09949	0.09948	0.09946	0.10151	0.15619	0.14019	0.13988
	MAD	0.07958	0.07957	0.07957	0.08120	0.12488	0.11228	0.11219
	$\in[1,\infty)$	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
	$\in[-0.1,0.1]$	0.68533	0.68298	0.68311	0.67319	0.47573	0.52236	0.51897
	$\in[-0.05,0.05]$	0.38586	0.38248	0.38252	0.37538	0.24913	0.27806	0.27542
(c)	Mean	0.74637	0.74486	0.74343	0.80963	0.76791	0.75079	0.75757
	Median	0.77500	0.77198	0.77235	0.83911	0.80054	0.79500	0.80418
	Mode	0.81300	0.81781	0.83230	0.88892	0.87377	0.88500	0.89671
	SD	0.13702	0.14221	0.13846	0.15457	0.21329	0.17498	0.18164
	RMSE	0.14714	0.15252	0.14957	0.15487	0.21569	0.18177	0.18653
	MAD	0.10536	0.11067	0.10722	0.12171	0.16237	0.12914	0.13483
	$\in[1,\infty)$	0.00000	0.00244	0.00000	0.05630	0.10447	0.00000	0.00000
	$\in[-0.7,0.9]$	0.62403	0.59131	0.61504	0.47566	0.39979	0.52139	0.47988
	$\in[-0.75,0.85]$	0.34213	0.31966	0.33360	0.24101	0.20520	0.26566	0.24090
(d)	Mean	0.78490	0.78466	0.78459	0.80067	0.79096	0.78696	0.79080
	Median	0.79300	0.79223	0.79228	0.80840	0.79993	0.79900	0.80305
	Mode	0.81200	0.81057	0.81449	0.82712	0.82392	0.81300	0.82330
	SD	0.06269	0.06354	0.06305	0.06484	0.09833	0.08004	0.08431
	RMSE	0.06448	0.06536	0.06491	0.06484	0.09875	0.08109	0.08481
	MAD	0.04934	0.05009	0.04967	0.05115	0.07715	0.06206	0.06562
	$\in[1,\infty)$	0.00000	0.00000	0.00000	0.00000	0.00509	0.00000	0.00000
	$\in[-0.7,0.9]$	0.89673	0.89109	0.89357	0.89351	0.70928	0.82525	0.79653
	$\in[-0.75,0.85]$	0.60841	0.59621	0.60056	0.56462	0.40239	0.49900	0.46282
GARCH(1,1) with $t(3)$ innovations: (e) $\rho = 0, n = 25$, , (f) $\rho = 0, n = 100$								
AR(1)-GARCH(1,1) with $t(3)$ innovations: (g) $\rho = 0.8, n = 25$, (h) $\rho = 0.8, n = 100$								
		$\hat{\rho}_{ML}$	$\hat{\rho}$	$\hat{\rho}_B$	$\hat{\rho}_{MP}$	$\hat{\rho}_H$	$\hat{\rho}_1$	$\hat{\rho}_3$
(e)	Mean	−0.00055	−0.00057	−0.00054	−0.00062	−0.00077	−0.00059	−0.00059
	Median	−0.00100	−0.00067	−0.00064	−0.00073	−0.00109	−0.00100	−0.00086
	Mode	−0.03500	−0.03949	−0.02778	−0.04292	−0.02308	−0.02200	0.01016
	SD	0.22482	0.24072	0.21653	0.26165	0.29731	0.25835	0.26502
	RMSE	0.22482	0.24072	0.21653	0.26165	0.29731	0.25835	0.26502
	MAD	0.17814	0.18480	0.17357	0.20087	0.23588	0.20877	0.21533
	$\in[1,\infty)$	0.00000	0.00173	0.00000	0.00231	0.00137	0.00000	0.00000
	$\in[-0.1,0.1]$	0.34991	0.34031	0.35222	0.31459	0.26675	0.29251	0.27952
	$\in[-0.05,0.05]$	0.17976	0.17383	0.18016	0.15981	0.13540	0.14920	0.14171
(f)	Mean	−0.00012	−0.00016	−0.00010	−0.00017	0.00004	−0.00008	−0.00007
	Median	0.00000	−0.00007	−0.00007	−0.00007	−0.00002	0.00000	0.00012
	Mode	0.00700	−0.00788	0.00092	−0.00804	−0.00358	0.00100	0.00337
	SD	0.18993	0.19707	0.17799	0.20110	0.14744	0.13037	0.13550
	RMSE	0.18993	0.19707	0.17799	0.20110	0.14744	0.13037	0.13550
	MAD	0.14551	0.14782	0.13923	0.15084	0.11749	0.10433	0.10864
	$\in[1,\infty)$	0.00000	0.00114	0.00000	0.00121	0.00000	0.00000	0.00000
	$\in[-0.1,0.1]$	0.43979	0.43405	0.44704	0.42648	0.50383	0.55615	0.53384
	$\in[-0.05,0.05]$	0.23170	0.22722	0.23435	0.22282	0.26592	0.29928	0.28430

(Continued on next page)

Table 1. (Continued).

GARCH(1,1) with $t(3)$ innovations: (e) $\rho = 0, n = 25$, , (f) $\rho = 0, n = 100$

AR(1)-GARCH(1,1) with $t(3)$ innovations: (g) $\rho = 0.8, n = 25$, (h) $\rho = 0.8, n = 100$

		$\hat{\rho}_{ML}$	$\hat{\rho}$	$\hat{\rho}_B$	$\hat{\rho}_{MP}$	$\hat{\rho}_H$	$\hat{\rho}_1$	$\hat{\rho}_3$
(g)	Mean	0.71565	0.73683	0.70615	0.80090	0.77122	0.76006	0.76624
	Median	0.75000	0.76522	0.74186	0.83176	0.80038	0.80200	0.80948
	Mode	0.81300	0.82133	0.81758	0.89275	0.85137	0.87800	0.89427
	SD	0.16423	0.18471	0.16545	0.20077	0.20211	0.16494	0.17057
	RMSE	0.18463	0.19521	0.19022	0.20077	0.20415	0.16971	0.17388
	MAD	0.12916	0.13700	0.13324	0.14778	0.15348	0.12144	0.12668
	$\in[1,\infty)$	0.00000	0.02846	0.00000	0.10142	0.09427	0.00000	0.00000
	$\in[-0.7,0.9]$	0.55242	0.51083	0.54667	0.42321	0.42240	0.54167	0.50086
	$\in[-0.75,0.85]$	0.30298	0.27274	0.29735	0.21484	0.21872	0.27606	0.25149
(h)	Mean	0.74684	0.75779	0.72862	0.77325	0.79219	0.78997	0.79297
	Median	0.76900	0.77682	0.75448	0.79268	0.80006	0.80100	0.80444
	Mode	0.81200	0.80480	0.79152	0.82121	0.80765	0.82200	0.83304
	SD	0.13471	0.14470	0.13221	0.14766	0.09271	0.07667	0.08034
	RMSE	0.14482	0.15073	0.15025	0.15006	0.09304	0.07733	0.08065
	MAD	0.10162	0.10422	0.10395	0.10456	0.07266	0.05946	0.06262
	$\in[1,\infty)$	0.00000	0.01841	0.00000	0.02454	0.00389	0.00000	0.00000
	$\in[-0.7,0.9]$	0.63370	0.62562	0.64040	0.61655	0.73758	0.84192	0.81615
	$\in[-0.75,0.85]$	0.36360	0.35424	0.37035	0.34154	0.42543	0.51459	0.48112

in particular the negative side effects of bias correction and the suitability of the estimators based on the truncated Cauchy distribution for financial data. For example, comparing $\widehat{\rho}$ and $\widehat{\rho}_{MP}$ in scenario (c), where $\rho = 0.8$ and $n = 25$, we see that bias correction increases the mean from 0.74 to 0.81 but also the mode from 0.82 to 0.89 and thereby reduces the probability that the estimate lies within the interval [0,7,0.9] from 0.59 to 0.48. The probability that the estimate is greater than one increases from 0.00 to 0.05. Finally, also the RMSE and the MAD increase (see Table 1.c). For financial applications, a scenario such as (h) is more relevant, where the observations come from a GARCH(1,1) process with $t(3)$ innovations. In this case, the bias is not an issue because $\rho = 0$. Conditional heteroscedasticity and extreme observations are much bigger problems. Using the simple robust estimator $\widehat{\rho}_3$ instead of the LS estimator $\widehat{\rho}$ increases the probability that the estimate lies within the interval $[-0.1,0.1]$ from 0.43 to 0.53. The RMSE and the MAD decrease from 0.20 and 0.15 to 0.14 and 0.11, respectively (see Table 1.f).

5. Concluding remarks

We have explored novel estimators of the first-order autocorrelation ρ that are based on ratios of successive observations. By construction, they are robust against (conditional) heteroscedasticity. Moreover, a suitable transformation makes sure that the denominator is always greater than the numerator, which eliminates the danger of extreme values either due to large numerators or small denominators. In the Gaussian case, this transformation allows to change from a Cauchy distribution, which does not have any finite moments, to a truncated Cauchy distribution, which is defined over a finite interval and therefore has all its moments. Thus, the new estimators are best suited for the case where both outliers and clusters of high volatility are present. Indeed, the results of a simulation study show that they outperform conventional estimators in this case provided that the sample size is not too small. A major advantage of our estimators is their simplicity, which obviates the need for numerical optimization.

The results of the simulation study also corroborate MacKinnon and Smith's (1998) finding that bias correction is not necessarily a good thing because it may increase the MSE. Including also bias-corrected versions of the LS estimator in our simulation study and looking at the whole sampling distribution rather than only at its MSE, we observed that bias correction may also move the mode of the sampling distribution further away from ρ towards the edge of the stable region, which increases the probability of obtaining an estimate greater than 1 and reduces the probability of obtaining an estimate reasonably close to the true value.

Tasks for future research include the application to financial data, the further investigation of the properties of the new estimators, the extension to higher-order autocorrelation as well as to cross-correlation.

Acknowledgement

The author very much appreciates the referee's comments, which helped to improve the manuscript.

ORCID

E. Reschenhofer http://orcid.org/0000-0002-4866-2324

References

Abramowitz, M., and I. A. Stegun. 1970. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. (9th printing). New York: Dover.

Andrews, D. W. K. 1993. Exactly median-unbiased estimation of first order autoregressive/unit root models. *Econometrica* 62:139–165.

Bartlett, M. S. 1946. On the theoretical specification and sampling properties of autocorrelated time-series. *Supplement to the Journal of the Royal Statistical Society* 8:27–41.

Bickel, D. R., and R. Frühwirth. 2006. On a fast, robust estimator of the mode: comparisons to other robust estimators with applications. *Computational Statistics and Data Analysis* 50:3500–30.

Bloch, B. 1966. Note on the estimation of the location parameter of the Cauchy distribution. *Journal of the American Statistical Association* 61:852–55.

Burg, J. P. 1967. *Maximum Entropy Spectral Analysis*. Presented at the 37th Annual International Meeting of the Society of Exploration Geophysicists, Oklahoma City, October, 31st, 1967.

Burg, J. P. 1975. *Maximum Entropy Spectral Analysis*. PhD thesis, Stanford University.

Cedilnik, A., K. Košmelj, and A. Blejec. 2004. The distribution of the ratio of jointly normal variables. *Metodološki zvezki* 1:99–08.

Gradshteyn, I. S., and I. M. Ryzhik. 2007. *Table of Integrals, Series, and Products* (7th edition, edited by A. Jeffrey, D. Zwillinger). Amsterdam: Elsevier/Academic Press.

Hurwicz, L. 1950. Least-squares bias in time series. In *Statistical Inference in Dynamic Models*. T. C. Koopmans (ed.) New York: Wiley. 365–83.

Jamnik, R. 1971. *Verjetnostni račun*. Ljubljana: Mladinska knjiga.

Kendall, M. G. 1954. Note on bias in the estimation of autocorrelation. *Biometrika* 41:403–04.

Kullback, S., and R. A. Leibler. 1951. On information and sufficiency. *Annals of Mathematical Statistics* 22:79–86.

Le Breton, A., and D. T. Pham. 1989. On the bias of the least squares estimator for the first order autoregressive process. *Annals of the Institute of Statistical Mathematics* 41:555–63.

MacKinnon, J. G., and Smith, A. A., Jr. 1998. Approximate bias correction in econometrics. *Journal of Econometrics* 85:205–30.

Marriott, F. H. C., and J. A. Pope. 1954. Bias in the estimation of autocorrelations. *Biometrika* 41:390–02.

Marsaglia, G. 2006. Ratios of normal variables. *Journal of Statistical Software* 16:1–10.

Nadarajah, S., and S. Kotz. 2006. A truncated Cauchy distribution. *International Journal of Mathematical Education in Science and Technology* 37:605–08.

Orcutt, G. H., and H. S. Winokur. 1969. First order autoregression: inference, estimation and prediction. *Econometrica* 37:1–14.
Quenouille, M. H. 1949. Approximate Tests of Correlation in Time-Series. *Journal of the Royal Statistical Society*. Series B 11:68–84.
R Core Team. 2017. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
Robertson, T., and J. D. Cryer. 1974. An iterative procedure for estimating the mode. *Journal of the American Statistical Association* 69:1012–16.
Rothenberg, T. J., F. M. Fisher, and C. B. Tilanus. 1964. A note on estimation from a Cauchy sample. *Journal of the American Statistical Association* 59:460–63.
Schroeder, M., and R. Zieliński. 2010. Estimating AR(1) with entropy loss function. *IM PAN Preprint* 726:1–10.
Tanizaki, H. 2000. Bias correction of OLSE in the regression model with lagged dependent variables. *Computational Statistics & Data Analysis* 34:495–11.
White, J. S. 1961. Asymptotic expansions for the mean and variance of the serial correlation coefficient. *Biometrika* 48:85–94.
Zieliński, R. 1999. A median-unbiased estimator of the AR(1) coefficient. *Journal of Time Series Analysis* 20:477–81.