

## **EXAMINING THE PROPERTIES OF A SIMPLE ESTIMATOR BASED ON TRANSFORMED CAUCHY VARIABLES**

**ERHARD RESCHENHOFER**

Department of Statistics and Operations Research

University of Vienna

Oskar-Morgenstern-Platz 1

1090 Vienna

Austria

e-mail: [erhard.reschenhofer@univie.ac.at](mailto:erhard.reschenhofer@univie.ac.at)

### **Abstract**

This paper investigates some theoretical properties of a simple estimator for the location parameter of a Cauchy distribution and a truncated Cauchy distribution, respectively, which can also be used for the robust estimation of the first-order autocorrelation of a stationary time series. Because of its simplicity and its robustness against heteroscedasticity and extreme values, this estimator is particularly useful for the analysis of serial correlation in financial time series. When applied to suitably transformed observations from a Cauchy distribution, this estimator is more efficient than competing estimators such as the sample median, trimmed means and other estimators based on sample order statistics.

---

2010 Mathematics Subject Classification: 62E15, 62F10.

Keywords and phrases: truncated Cauchy distribution, sample median, bias, Cramer-Rao lower bound.

Received August 23, 2017

### 1. Introduction

The ratio

$$X = \frac{Y_1}{Y_2} \quad (1)$$

of two jointly normal variables  $Y_1$  and  $Y_2$  with zero means, standard deviations  $\sigma_1$  and  $\sigma_2$ , and correlation

$$-1 < \rho = \frac{E(Y_1 Y_2)}{\sigma_1 \sigma_2} < 1 \quad (2)$$

has a Cauchy distribution with probability density function

$$f(x; \rho) = \frac{1}{\pi} \frac{b}{(x - a)^2 + b^2}, \quad (3)$$

where the location parameter

$$-\infty < a = \rho \frac{\sigma_1}{\sigma_2} < \infty \quad (4)$$

specifies both the median and the mode and the scale parameter

$$0 < b = \sqrt{1 - \rho^2} \frac{\sigma_1}{\sigma_2} \quad (5)$$

specifies the interquartile range (see Jamnik [5], 1971; for the distribution of the ratio of two arbitrary normal variables, see Cedilnik et al. [2]; Marsaglia [6]).

Considering the special case where  $b = 1$ , Rothenberg et al. [11] and Bloch [1] tried to find estimators  $\hat{a}(X_1, \dots, X_n)$  for  $a$  that are more efficient than the sample median based on a random sample  $X_1, \dots, X_n$  from the density (3). Rothenberg et al. [11] showed that the trimmed mean based on the middle 24% of the sample order statistics has an asymptotic relative efficiency (A.R.E.) of 88% as compared to 81% for the

sample median. Bloch [1] found a weighted average of five sample order statistics, which has an A.R.E. of 95%.

In a time series setting, where  $Y_1$  and  $Y_2$  are successive observations from a stationary process and  $\rho$  is the first-order autocorrelation (see Hurwicz [4]; Zieliński [12]), the density (3) reduces to

$$f(x; \rho) = \frac{\sqrt{1 - \rho^2}}{\pi} \frac{1}{x^2 - 2\rho x + 1} \quad (6)$$

and the restriction  $b = 1$  implies that  $\rho = 0$ . It will be shown in the next section that in this case the sample mean of suitably transformed observations has a relative efficiency of 97%. The transformation allows to switch from a Cauchy distribution with infinite support to a truncated Cauchy distribution defined over the finite interval  $[-1, 1]$ . Moreover, the nontrivial case of  $\rho \neq 0$  will also be investigated. However, the theoretical properties critically depend on the assumption that  $\rho$  is small, which is usually satisfied in financial applications. Section 3 concludes the paper.

## 2. Theoretical Results

The maximum likelihood (ML) estimator  $\hat{\rho}_{ML}$  for the parameter  $\rho$  in (6) is asymptotically efficient. It is obtained by equating the score function

$$S(\rho) = \sum_{t=1}^n \frac{\partial}{\partial \rho} \log(f(x_t; \rho)) = n \frac{-\rho}{1 - \rho^2} + \sum_{t=1}^n \frac{2x_t}{x^2 - 2\rho x + 1} \quad (7)$$

to zero. Unfortunately, numerical methods are required to solve this equation. There exists no explicit solution. The asymptotic variance of  $\hat{\rho}_{ML}$  is given by the inverse of the Fisher information  $I(\rho)$ , which is the Cramer-Rao lower bound (CRLB) on the variance of any unbiased estimator of  $\rho$ . We have

$$\begin{aligned}
I(\rho) &= E\left(-\sum_{t=1}^n \frac{\partial^2}{\partial \rho} \log(f(x_t; \rho))\right) \\
&= n\left(\frac{1+\rho^2}{(1-\rho^2)^2} - \frac{4\sqrt{1-\rho^2}}{\pi} \int_{-\infty}^{\infty} \frac{x^2}{(x^2-2\rho x+1)^3} dx\right) \\
&= n\left(\frac{1+\rho^2}{(1-\rho^2)^2} - \frac{4\sqrt{1-\rho^2}}{\pi} \frac{2+4\rho^2}{16(1-\rho^2)^2} \int_{-\infty}^{\infty} \frac{1}{x^2-2\rho x+1} dx\right) \\
&= n\left(\frac{1+\rho^2}{(1-\rho^2)^2} - \frac{\sqrt{1-\rho^2}}{2\pi} \frac{1+2\rho^2}{(1-\rho^2)^2} \frac{2}{2\sqrt{1-\rho^2}} \arctan\left(\frac{2(x-\rho)}{2\sqrt{1-\rho^2}}\right)\right) \Bigg|_{-\infty}^{\infty} \\
&= \frac{n}{2(1-\rho^2)^2}. \tag{8}
\end{aligned}$$

The asymptotic variance of the  $p$ -sample quantile is given by

$$\frac{1}{n} \frac{p(1-p)}{(f(\rho; \rho))^2} = \frac{1}{n} p(1-p)\pi^2(1-\rho^2), \tag{9}$$

hence the A.R.E. of the sample median  $\hat{\rho}_{med}$ , which is the 0.5-sample quantile, is

$$\frac{2(1-\rho^2)^2}{\pi^2(1-\rho^2)} = \frac{8}{\pi^2}(1-\rho^2). \tag{10}$$

The fact that the A.R.E. decreases as  $|\rho|$  increases and goes to zero as  $|\rho| \rightarrow 1$  is not as bad as it looks because only small values of  $\rho$  are relevant for financial applications. For  $|\rho| = 0.0, 0.1,$  and  $0.2$ , the values 0.811, 0.802, and 0.778, respectively, are obtained. Moreover, it may be expected that the A.R.E. of the sample median will be much higher when the variables  $X_1, X_2, X_3, \dots$  are not i.i.d. but are rather obtained from a time series  $Y_0, Y_1, Y_2, Y_3, \dots$  via  $X_t = Y_t / Y_{t-1}$ . This is due to the fact that the sample median will probably be much less affected than the ML

estimator by dependencies between successive ratios. In case of a series of asset returns, approximate independence can be achieved by omitting every third return, i.e.,

$$X_1 = \frac{Y_1}{Y_0}, \quad X_2 = \frac{Y_4}{Y_3}, \quad X_3 = \frac{Y_7}{Y_6}, \dots, \quad (11)$$

because the autocorrelation at lags higher than one is negligible and (conditional) heteroscedasticity is already largely eliminated by the definition of  $X_t$  as the ratio of two successive returns.

It follows from

$$\frac{2x}{x^2 - 2\rho x + 1} = \frac{2\frac{1}{x}}{\left(\frac{1}{x}\right)^2 - 2\rho\frac{1}{x} + 1} \quad (12)$$

that neither the score function (7) nor the ML estimator will change when an observation  $X_t$  is replaced by its inverse. Reschenhofer [9] therefore proposed to switch from the sample  $X_1, \dots, X_n$  to the sample  $Z_1, \dots, Z_n$ , where

$$Z_t = \begin{cases} X_t, & \text{if } |X_t| \leq 1, \\ \frac{1}{X_t}, & \text{if } |X_t| > 1. \end{cases} \quad (13)$$

The latter sample has the advantage that all moments exist because the variables  $Z_t$  are defined over the finite interval  $[-1, 1]$ . Their density is given by

$$f^*(z; \rho) = 2f(z; \rho) = 2\frac{\sqrt{1-\rho^2}}{\pi} \frac{1}{z^2 - 2\rho z + 1}, \quad -1 \leq z \leq 1, \quad (14)$$

which is a special case of the truncated Cauchy distribution introduced by Nadarajah and Kotz [7]. Using suitable integration rules for rational functions (see, e.g., Gradshteyn and Ryzhik [3], pp. 79-80), we obtain

$$\begin{aligned}
E(Z_t) &= \int_{-1}^1 z f^*(z) dz = \frac{2\sqrt{1-\rho^2}}{\pi} \int_{-1}^1 \frac{z}{z^2 - 2\rho z + 1} dz \\
&= \frac{2\sqrt{1-\rho^2}}{\pi} \left( \frac{1}{2} \log(z^2 - 2\rho z + 1) \Big|_{-1}^1 + \rho \int_{-1}^1 \frac{1}{z^2 - 2\rho z + 1} dz \right) \\
&= \frac{\sqrt{1-\rho^2}}{\pi} \log\left(\frac{1-\rho}{1+\rho}\right) + \rho \frac{2\sqrt{1-\rho^2}}{\pi} \frac{1}{\sqrt{1-\rho^2}} \arctan\left(\frac{x-\rho}{\sqrt{1-\rho^2}}\right) \Big|_{-1}^1 \\
&= \rho + \frac{\sqrt{1-\rho^2}}{\pi} \log\left(\frac{1-\rho}{1+\rho}\right) = h(\rho), \tag{15}
\end{aligned}$$

$$\begin{aligned}
E(Z_t^2) &= \frac{2\sqrt{1-\rho^2}}{\pi} \int_{-1}^1 \frac{z^2}{(z^2 - 2\rho z + 1)^3} dz \\
&= \frac{2\sqrt{1-\rho^2}}{\pi} \left( (z + \rho \log(z^2 - 2\rho z + 1)) \Big|_{-1}^1 + (2\rho^2 - 1) \int_{-1}^1 \frac{1}{z^2 - 2\rho z + 1} dz \right) \\
&= \frac{2\sqrt{1-\rho^2}}{\pi} \left( 2 + \rho \log\left(\frac{1-\rho}{1+\rho}\right) \right) + 2\rho^2 - 1, \tag{16}
\end{aligned}$$

and

$$\text{var}(Z_t) = E(Z_t^2) - (E(Z_t))^2 = \rho^2 + \frac{4\sqrt{1-\rho^2}}{\pi} - 1 + \frac{1-\rho^2}{\pi^2} \left( \log\left(\frac{1-\rho}{1+\rho}\right) \right)^2. \tag{17}$$

Reschenhofer [9] proposed the estimator

$$\hat{\rho}_\Phi = -1 + 2\Phi_{0.295}(\bar{Z}), \tag{18}$$

where  $\Phi_\sigma$  denotes the cumulative distribution function of a normal distribution with mean 0 and standard deviation  $\sigma$ , and found in an extensive simulation study that in the presence of extreme values and clusters of high volatility, it outperforms both the sample median of  $X_1, \dots, X_n$  as well as conventional estimators of the autocorrelation  $\rho$

that are based on the original time series  $Y_0, \dots, Y_n$ . Reschenhofer [10] argued that in financial applications, where  $\rho$  is usually small, the much simpler estimator

$$\tilde{\rho} = \frac{\pi}{\pi - 2} \bar{Z} \quad (19)$$

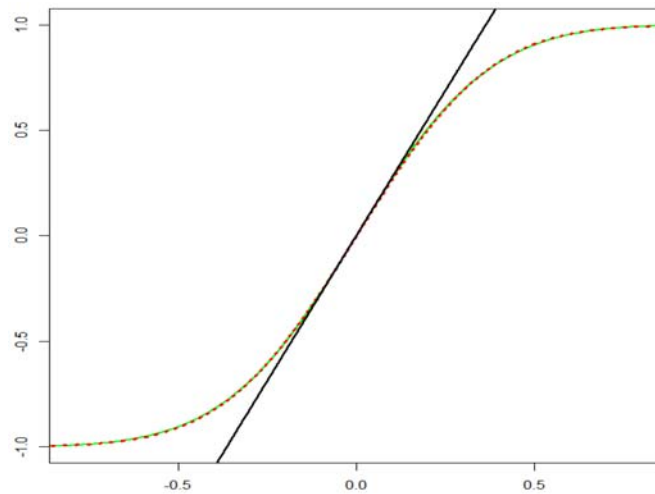
is also adequate. In an empirical study of index returns, he demonstrated the usefulness of the latter estimator for the almost instantaneous detection of a structural break in the autocorrelation without any delay caused by an estimation window. Figure 1 shows that the function

$$\hat{h}_{\Phi}^{-1}(\xi) = -1 + 2\Phi_{0.295}(\xi) \quad (20)$$

is an excellent approximation of the inverse function of  $h$  and that the same is true for the function

$$\tilde{h}^{-1}(\xi) = \frac{\pi}{\pi - 2} \xi \quad (21)$$

in a neighbourhood of zero. All computations are carried out with the free statistical software R (R Core Team [8]).



**Figure 1.** Graphs of  $h^{-1}$  (green),  $\hat{h}_{\Phi}^{-1}$  (red), and the linear approximation  $\tilde{h}^{-1}$  (black).

Because of its robustness and simplicity, the estimator  $\tilde{\rho}$  may become a standard tool for the analysis of the serial correlation in return series. To check the relative efficiency of  $\tilde{\rho}$ , which is approximately unbiased in a neighbourhood of zero, we must compare its variance

$$\text{var}(\tilde{\rho}) = \left( \frac{\pi}{\pi - 2} \right)^2 \frac{1}{\pi} \text{var}(Z) \quad (22)$$

with the CRLB

$$\frac{1}{I(\rho)} = \frac{2(1 - \rho^2)^2}{n}. \quad (23)$$

The ratio of (23) and (22) is given by 0.967, 0.921, and 0.802 for  $|\rho| = 0.0, 0.1,$  and  $0.2,$  respectively. These values are much higher than the corresponding values 0.811, 0.802, and 0.778 for the sample median  $\hat{\rho}_{med}$ . Clearly, focusing exclusively on the variance or the asymptotic variance of an estimator only makes sense when it is unbiased or asymptotically unbiased. In an asymptotic analysis, I must therefore replace  $\tilde{\rho}$  by

$$\hat{\rho}_h = h^{-1}(\bar{Z}). \quad (24)$$

The inverse of the function  $h$  can be determined numerically to any desired precision. However, the discrepancy between  $\tilde{h}^{-1}$  and  $h^{-1}$  may be negligible in practice when  $\rho$  is small. For example,

$$\left( \tilde{h}^{-1}(0.1) - h^{-1}(0.1) \right)^2 = 0.000035,$$

whereas

$$\text{var}(\tilde{\rho}) > 0.002,$$

for  $|\rho| \leq 0.98$  and  $|n| \leq 1000$ .



### 3. Concluding Remarks

This paper is concerned with the estimation of a “stationary” Cauchy distribution, which is obtained by dividing two centered normal random variables under the assumption of equal variances. The location parameter  $\rho$  can be interpreted as the first-order autocorrelation when the variables are successive observations of a stationary process. Applying a suitable transformation to a sample of observations from this distribution allows the use of the sample mean for the estimation of  $\rho$ . The transformed observations follow a truncated Cauchy distribution, which has all its moments because it is defined over a finite interval. However, a mean correction is required. If it is known that  $\rho$  is small, e.g., in financial applications, a simple multiplication by a constant will do the trick. The resulting estimator is more efficient than common estimators for the location parameter of a Cauchy distribution such as the sample median, trimmed means and other estimators based on sample order statistics.

Because of its simplicity and its robustness, this estimator is an ideal tool for the analysis of serial correlation in return series. The robustness against heteroscedasticity is due to the fact that the estimator is based on ratios of successive observations. Robustness against outliers is achieved through the stabilizing transformation of the ratios. Tasks for future research include the extension to higher-order autocorrelation as well as to cross-correlation.

### References

- [1] B. Bloch, Note on the estimation of the location parameter of the Cauchy distribution, *Journal of the American Statistical Association* 61 (1966), 852-855.
- [2] A. Cedilnik, K. Košmelj and A. Blejec, The distribution of the ratio of jointly normal variables, *Metodološki zvezki* 1 (2004), 99-108.
- [3] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (7th edition, edited by A. Jeffrey and D. Zwillinger), Amsterdam: Elsevier/Academic Press, 2007.
- [4] L. Hurwicz, Least-squares bias in time series. In: Koopmans TC (ed.), *Statistical Inference in Dynamic Models*, New York: Wiley (1950), 365-383.

- [5] R. Jamnik, Verjetnostni račun, Ljubljana: Mladinska knjiga, 1971.
- [6] G. Marsaglia, Ratios of normal variables, *Journal of Statistical Software* 16 (2006), 1-10.
- [7] S. Nadarajah and S. Kotz, A truncated Cauchy distribution, *International Journal of Mathematical Education in Science and Technology* 37 (2006), 605-608.
- [8] R Core Team, R: A Language and Environment for Statistical Computing, Vienna: R Foundation for Statistical Computing, 2017.
- [9] E. Reschenhofer, Heteroscedasticity-robust estimation of autocorrelation, Working Paper, 2017a.
- [10] E. Reschenhofer, Using ratios of successive returns for the estimation of serial correlation in return series, Working Paper, 2017b.
- [11] T. J. Rothenberg, F. M. Fisher and C. B. Tilanus, A note on estimation from a Cauchy sample, *Journal of the American Statistical Association* 59 (1964), 460-463.
- [12] R. Zieliński, A median-unbiased estimator of the AR(1) coefficient, *Journal of Time Series Analysis* 20 (1999), 477-481.

