INSTITUTE OF MATHEMATICS
of the
Polish Academy of Sciences

ul. Śniadeckich 8, P.O.B. 21, 00-956 Warszawa 10, Poland http://www.impan.pl

IM PAN Preprint 726 (2010)

Małgorzata Schroeder, Ryszard Zieliński

Estimating AR(1) with Entropy Loss Function

Published as manuscript

Received 15 August 2010

ESTIMATING $AR(1)$ WITH ENTROPY LOSS FUNCTION

Małgorzata Schroeder [1]

Institute of Mathematics
University of Białystok
Akademicka 2, 15-267 Białystok, Poland
mszuli@math.uwb.edu.pl

Ryszard Zieliński

Department of Mathematical Statistics
Institute of Mathematics Polish Acad. Sc.
00-956 Warszawa, POB 21, Poland
R.Zielinski@impan.pl

Key Words: first-order stationary autoregressive process, autocorrelation, entropy loss function, risk

ABSTRACT

There is an abundance of estimators of the autocorrelation coefficient ρ in the $AR(1)$ time series model $X_t = \rho X_{t-1} + \varepsilon_t$. This calls for a criterion to select a suitable one. We provide such a criterion. Typically estimators of an unknown parameter are compared with respect to their mean square error (MSE) (or variance in the case of unbiased estimators), and an estimator with uniformly minimum MSE is considered to be the best one. The symmetric square-error loss is one of the most popular loss functions. It is widely employed in inference, but it is not appropriate for our problem, in which the parameter space is the bounded open interval $(-1, 1)$. The risk based on MSE does not eliminate estimators that may assume values outside the parameter space (for example, the Least Squares Estimator). As a criterion for comparing estimators we propose the Entropy Loss Function (ELF) (or the Kullback-Leibler information number). With that criterion the risk of estimators which may

[1]Address correspondence to Małgorzata Schroeder, Institute of Mathematics, University of Białystok, Akademicka 2, 15-267 Białystok, Poland;
E-mail: mszuli@math.uwb.edu.pl

assume values greater than 1 or smaller than -1 is equal to infinity so they are naturally eliminated. In this paper some well known estimators are compared with respect to their risk under ELF. From among three acceptable estimators that we know, the Maximum Likelihood Estimator (MLE) has the uniformly minimum risk but the estimator constructed by the Method of Moments seems to be preferable. An open problem is if there exists a uniformly best estimator. The problem of constructing a minimax estimator is also open.

1. INTRODUCTION AND NOTATION

Consider a stationary first-order autoregressive $AR(1)$ process of the form

$$X_t = \rho X_{t-1} + \varepsilon_t, \quad t = \ldots, -1, 0, 1, \ldots, \quad |\rho| < 1, \tag{1}$$

where $\varepsilon_t, t = \ldots, -1, 0, 1, \ldots,$ are independent identically distributed normal random variables with expectations equal to zero. The variances of ε_t are assumed to be equal to an unknown constant. All estimators of ρ we consider do not depend on the variance of ε_t hence without loss of generality we assume that $Var(\varepsilon_t) = 1$. For the process (1) we have

$$X_t = \sum_{i=0}^{\infty} \rho^i \varepsilon_{t-i}, \quad E_\rho(X_t) = 0, \quad E_\rho(X_t^2) = \frac{1}{1-\rho^2}, \quad E_\rho(X_t X_{t-1}) = \frac{\rho}{1-\rho^2}. \tag{2}$$

The coefficient ρ is to be estimated. A stationary segment $X_T, X_{T+1}, \ldots, X_{T+n-1}$ of the process is available. Without loss of generality we put $T = 1$. The stationarity of the segment means that X_1 is distributed as $N(0, 1/(1-\rho^2))$.

2. ENTROPY LOSS FUNCTION

Recall that if $f_\theta(x)$ is a probability density function, where θ is an unknown parameter to be estimated and $\hat{\theta}$ is an estimator of θ, then the *Entropy Loss Function* is given by the formula (see Kullback (1959), Rényi (1962), Sakamoto et al. (1986), Nematollahi and Motamed-Shariati (2009))

$$L(\theta, \hat{\theta}) = E_\theta \log\left(\frac{f_\theta(x)}{f_{\hat{\theta}}(x)}\right). \tag{3}$$

For a vector observation $(X_1, X_2, \dots, X_n)$ of the process we have

$$f_\rho(X_1, X_2, \dots, X_n) =$$

$$= (2\pi)^{-n/2}\sqrt{1-\rho^2}\exp\Big\{-\frac{1}{2}(1-\rho^2)X_1^2\Big\}\prod_{i=2}^{n}\exp\Big\{-\frac{1}{2}(X_i-\rho X_{i-1})^2\Big\}, \tag{4}$$

so that for $\rho, \hat{\rho} \in (-1, 1)$ we have

$$\log\Big(\frac{f_\rho(X_1, X_2, \dots, X_n)}{f_{\hat{\rho}}(X_1, X_2, \dots, X_n)}\Big) = \frac{1}{2}\log\frac{1-\rho^2}{1-\hat{\rho}^2} + (\rho-\hat{\rho})\sum_{i=2}^{n}X_iX_{i-1} - \frac{1}{2}(\rho^2-\hat{\rho}^2)\sum_{i=2}^{n}X_{i-1}^2.$$

Now we assume

$$L(\theta, \hat{\theta}) = \frac{1}{n}\int_{-\infty}^{\infty}\log\Big(\frac{f_\rho(x_1, x_2, \dots, x_n)}{f_{\hat{\rho}}(x_1, x_2, \dots, x_n)}\Big)f_\rho(x_1, x_2, \dots, x_n)dx_1dx_2\cdots dx_n. \tag{5}$$

For $|\hat{\rho}| \geqslant 1$ we set $L(\theta, \hat{\theta}) = \infty$. Eventually

$$L(\rho, \hat{\rho}) = \begin{cases} \frac{1}{n}\Big(\frac{1}{2}\log\frac{1-\rho^2}{1-\hat{\rho}^2} + \frac{\rho-\hat{\rho}}{2(1-\rho^2)}\big(\rho - (1-\frac{2}{n})\hat{\rho}\big)\Big), & \text{if } |\hat{\rho}| < 1, \\ +\infty, & \text{if } |\hat{\rho}| \geqslant 1. \end{cases} \tag{6}$$

The coefficient $1/n$ in formula (5) comes from the fact that if $(X_1, X_2, \dots, X_n)$ are i.i.d. then the formula is identical with (3) for real x.

The function (6) of argument $\hat{\rho} \in (-1, 1)$, for $n = 10$, $\rho = 0$, and $\rho = 0.5$, is presented in Fig. 1. The function $L(\rho, \hat{\rho})$ is convex, equals zero if $\hat{\rho} = \rho$ and tends to infinity as $\hat{\rho}$ approaches 1 or -1. The Entropy Loss Function has been successfully applied in estimation of parameters when the standard Mean Square Error criterion appeared to be unsatisfactory. Interesting applications can be found in Parsian et al. (1996) and Singh et al. (2008). The idea of the criterion is strictly connected with the Kullback-Leibler divergence between probability distributions: a distribution indexed by an unknown parameter and that indexed by an estimator of the parameter.

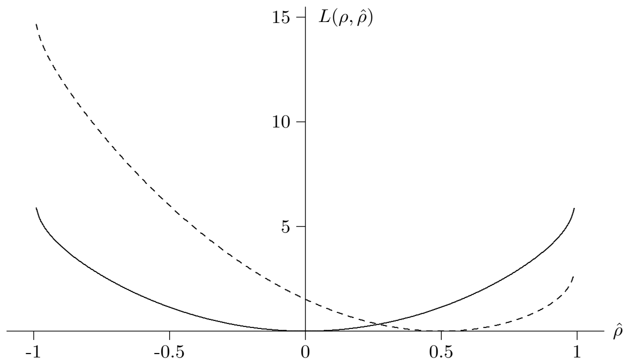


Fig.1. Entropy Loss Functions for $\rho = 0$ (solid) and $\rho = 0.5$ (dashed)

The risk function of an estimator $\hat{\rho} = \hat{\rho}(X_1, X_2, \ldots, X_n)$, to be denoted by $R_{\hat{\rho}}(\rho)$, $\rho \in (-1, 1)$, is given by the formula

(7)

$$R_{\hat{\rho}}(\rho) = \begin{cases} \int_{R^n} L\Big(\rho, \hat{\rho}(x_1, x_2, \ldots, x_n)\Big) f_\rho(x_1, x_2, \ldots, x_n) dx_1 dx_2 \cdots dx_n & \\ \qquad \text{for } \rho \text{ such that } P_\rho\{|\hat{\rho}(X_1, X_2, \ldots, X_n)| < 1\} = 1, & \\ \infty \qquad \text{for } \rho \text{ such that } P_\rho\{|\hat{\rho}(X_1, X_2, \ldots, X_n)| \geqslant 1\} > 0. & \end{cases}$$

Analytic or numerical calculation of the risk $R_{\hat{\rho}}(\rho)$ is rather difficult but Monte Carlo simulations can be easily applied.

2. ESTIMATORS

To demonstrate our idea the following six estimators have been chosen.

Maximum Likelihood Estimator:

$$\hat{\rho}_{MLE} = \arg\max_{\rho} \mathcal{L}(\rho; X_1, X_2, \ldots, X_n)$$

where

$$\mathcal{L}(\rho; x_1, x_2, \ldots, x_n) = \log(1-\rho^2) - x_1^2(1-\rho^2) - \sum_{i=2}^{n} (x_i - \rho x_{i-1})^2.$$

Observe that for every $x_1, x_2, \ldots, x_n$, the function $\mathcal{L}(\rho; x_1, x_2, \ldots, x_n)$ of $\rho \in (-1, 1)$ is concave (the second derivative is negative), $\mathcal{L}(\rho; x_1, x_2, \ldots, x_n) \to -\infty$ as $\rho \to \pm 1$ so that

$\hat{\rho}_{MLE} \in (-1, 1)$ is uniquely defined. A disadvantage of this estimator is that in order to calculate the value of $\hat{\rho}_{MLE}$ one has to solve numerically an algebraic equation of the third order. A more serious problem is that the lack of a simple closed formula makes it difficult to study the properties of the estimator.

Least Squares Estimator:

$$\hat{\rho}_{LSE} = \arg\min_{\rho} \sum_{i=2}^{n} (X_i - \rho X_{i-1})^2 = \frac{\sum_{i=2}^{n} X_i X_{i-1}}{\sum_{i=2}^{n} X_{i-1}^2}.$$

Estimator constructed by the Method of Moments: The sample counterpart of the correlation coefficient between X_t and X_{t-1}, $t = 2, \ldots, n$,

$$\rho = \frac{Cov(X_t, X_{t-1})}{\sqrt{Var(X_t) Var(X_{t-1})}},$$

if $EX_t = 0,\ t = 1, 2, \ldots, n$, is given by the formula

$$\hat{\rho}_{MM} = \frac{\sum_{i=2}^{n} X_i X_{i-1}}{\sqrt{\sum_{i=1}^{n-1} X_i^2 \sum_{i=2}^{n} X_i^2}}.$$

It should be noted that the support of the estimator ρ_{MM} is in the interval $(-1, 1)$.

Hurwicz estimator [Hurwicz (1950), Zieliński (1999)]:

$$\rho_{HUR} = Med\left(\frac{X_2}{X_1}, \frac{X_3}{X_2}, \ldots, \frac{X_n}{X_{n-1}}\right),$$

where $Med(\xi_1, \xi_2, \ldots, \xi_m)$ denotes a median of $\xi_1, \xi_2, \ldots, \xi_m$. A nice property of the estimator is that it is median-unbiased, which means that

$$P_\rho\{\hat{\rho}_{HUR} \leqslant \rho\} = P_\rho\{\hat{\rho}_{HUR} \geqslant \rho\} = \frac{1}{2} \quad \text{for all} \quad \rho \in (-1, 1).$$

This property holds under very general distributional assumptions, without assuming statistical independence (Luger 2005).

M-estimator with Huber loss function [Lehmann (1998)]:

$$\rho_{MHU} = \arg\min_{\rho} \sum_{i=1}^{n-1} L(X_{i+1} - \rho X_i).$$

with

$$L(x) = \begin{cases} \frac{1}{2}x^2 & \text{if } |x| \leqslant k, \\ k|x| - \frac{1}{2}k^2 & \text{if } |x| > k. \end{cases}$$

Following Lehmann we assume $k = 3/2$. Here also no simple explicit formula for $\hat{\rho}_{MHU}$ is known.

Burg's estimator [Provost (2005), Brockwell, Davis (2002)]:

This estimator has been constructed as that minimizing the forward and backward prediction errors:

$$\rho_{BUR} = \arg\min_{\rho} \sum_{i=2}^{n} ((X_i - \rho X_{i-1})^2 + (X_{i-1} - \rho X_i)^2).$$

Then

$$\rho_{BUR} = \frac{2\sum_{i=2}^{n} X_i X_{i-1}}{\sum_{i=2}^{n}(X_i^2 + X_{i-1}^2)}.$$

It should be noted that the support of the estimator ρ_{BUR} is in the interval $(-1, 1)$.

3. DISTRIBUTIONS OF ESTIMATORS

To assess basic properties of the estimators a simulation study has been performed. Some results (histograms) for $\rho = 0.8$ and $n = 10$, based on $10,000$ simulation runs, are exhibited in Fig. 2. We can observe that only the Maximum Likelihood Estimator $\hat{\rho}_{MLE}$, Burg's estimator $\hat{\rho}_{BUR}$, and the estimator constructed by the Method of Moments $\hat{\rho}_{MM}$ do not assume values outside the interval $(-1, 1)$.

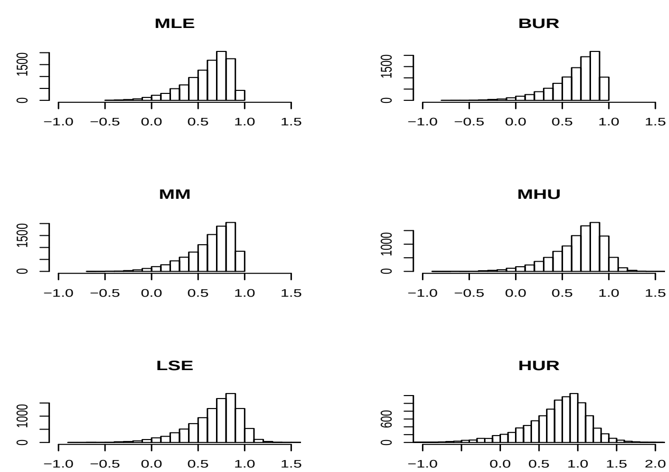
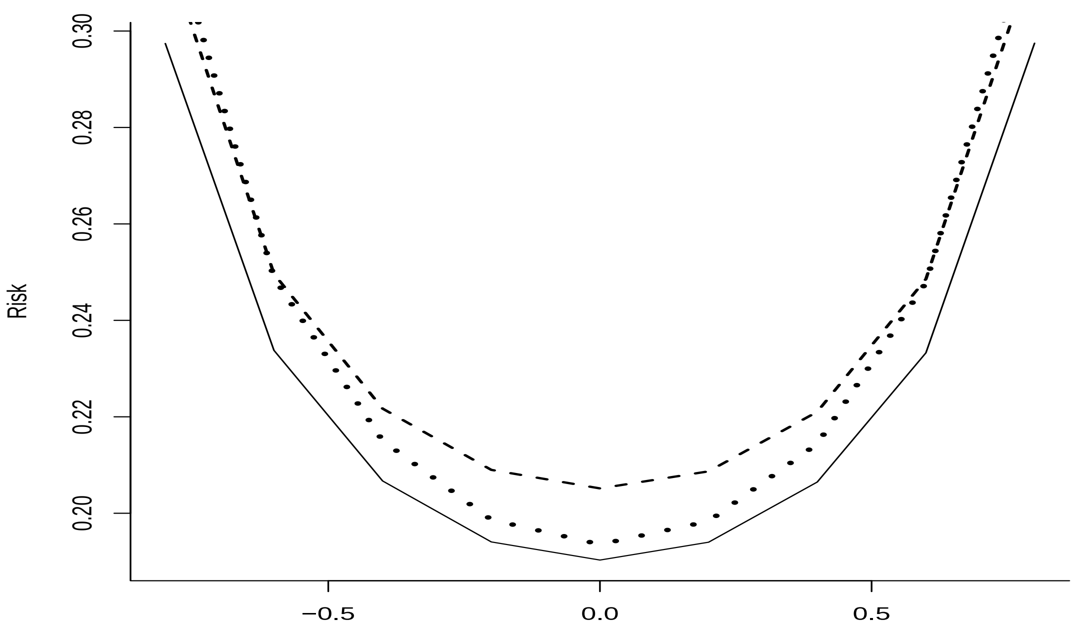


Fig. 2. Histograms

4. RISK OF ESTIMATORS

The Maximum Likelihood Estimator $\hat{\rho}_{MLE}$, Burg's estimator $\hat{\rho}_{BUR}$, and the estimator constructed by the Method of Moments $\hat{\rho}_{MM}$ assume all their values in the open interval $(-1, 1)$, so that the risk of these estimators is finite. The risk functions for $n = 10$ and for 10^5 simulation runs for all three estimators are presented in Fig. 3. Numerical values of the risk functions are presented in the Table. Two numbers in each entry are results of two independent simulations of 10^5 runs. Small differences indicate that the accuracy of the simulation results is satisfactory.



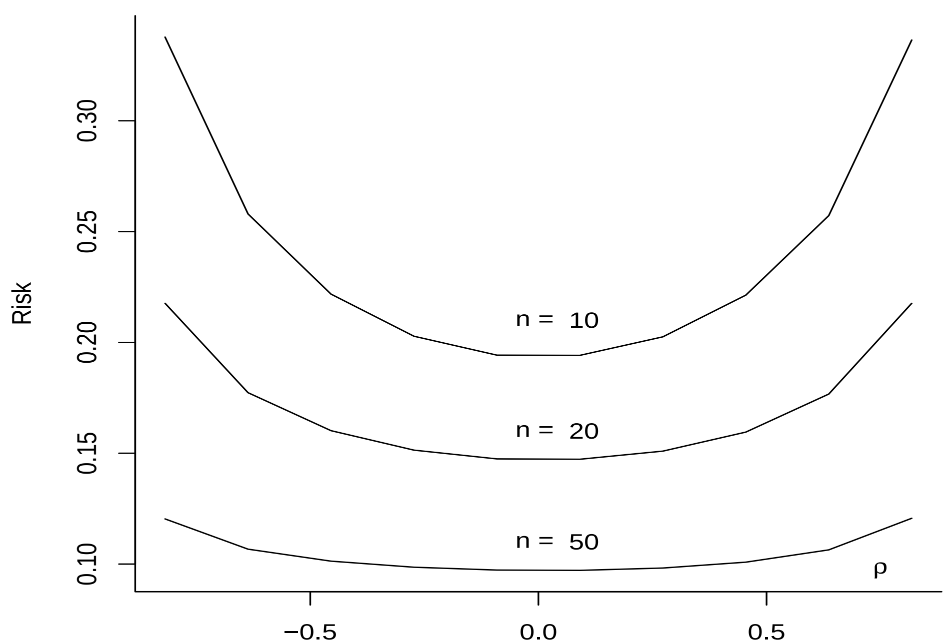
Fig, 3. Risk functions of estimators:
MLE(solid), BURG (dashes),MM (dotted)

It turns out that the Maximum Likelihood Estimator $\hat{\rho}_{MLE}$ has the uniformly smallest risk.

Table. Risk of estimators

ρ	R_{MLE}	R_{MM}	R_{MM}/R_{MLE}	R_{BRU}	R_{BRU}/R_{MLE}
-0.8	0.2974	0.3249	1.09	0.3175	1.07
	0.2992	0.3261	1.09	0.3188	1.07
-0.6	0.2338	0.2489	1.06	0.2496	1.07
	0.2357	0.2500	1.06	0.2508	1.06
-0.4	0.2067	0.2151	1.04	0.2217	1.07
	0.1953	0.2158	1.04	0.2225	1.07
-0.2	0.1941	0.1987	1.02	0.2090	1.08
	0.1953	0.1991	1.02	0.2094	1.07
0	0.1903	0.1935	1.02	0.2052	1.08
	0.1913	0.1939	1.01	0.2054	1.07
0.2	0.1940	0.1983	1.02	0.2087	1.08
	0.1949	0.1988	1.02	0.2091	1.07
0.4	0.2065	0.2144	1.04	0.2210	1.07
	0.2074	0.2154	1.04	0.2220	1.07
0.6	0.2333	0.2478	1.06	0.2486	1.07
	0.2342	0.2493	1.06	0.2502	1.07
0.8	0.2975	0.3239	1.09	0.3165	1.06
	0.2976	0.3248	1.09	0.3175	1.07

The risk functions for $n = 10, 20, 50$ and for 10^5 simulation runs for the estimator constructed by the Method of Moments $\hat{\rho}_{MM}$ are presented in Fig. 4. It is obvious that the risk is smaller if the number of observations is greater.



Fig, 4. Risk functions of estimators constructed by the Method of Moments

5. CONCLUDING REMARKS

The risk function based on the Entropy Loss Function enables one to eliminate estimators which may assume values outside the parameter space. We call such estimators unacceptable. We have considered three acceptable estimators. Among them, the Maximum Likelihood Estimator $\hat{\rho}_{MLE}$ has the uniformly smallest risk but the estimator $\hat{\rho}_{MM}$ constructed by the Method of Moments seems to be preferable in practice: its risk slightly exceeds that of $\hat{\rho}_{MLE}$ (not more than by 10 percent) but it is much easier to apply and it is easier to analyze its theoretical properties (ratio of two quadratic forms). It would be interesting to know if there exists an estimator with the uniformly smallest risk in the class of all acceptable estimators. It is also of interest to construct a minimax estimator under the Entropy Loss Function.

BIBLIOGRAPHY

Brockwell, J.P., Davis, A.R. (2002) *Introduction to Time Series and Forecasting*, Springer-Verlag, New York.

Hurwicz, L.(1950) Least-Squares Bias in Time Series. In *Statistical Inference in Dynamic Economic Models*, ed. T.C.Koopmans, Wiley, New York 365-383.

Kullback, S. (1959) *Information theory and statistics*, Wiley, New York.

Lehmann, E.L. (1998) *Theory of point estimation*, Springer-Verlag, New York.

Luger, R. (2005) Median-Unbiased estimation and exact inference methods for first-order autoregressive models with conditional heteroscedasticity of unknown form, *Journal of Time Series Analysis,* 27, 1, 119-128.

Nematollahi, N., Motamed-Shariati, F. (2009) Estimation of the selected gamma population under the entropy loss function. *Communications in Statistics - Theory and Methods*, *38*, 208-221.

Parsian, A., Nematollahi, N. (1996) Estimation of scale parameter under entropy loss function. *Journal of Statistical Planning and Inference*, 52, 77-91.

Provost, S.B., Sanjel, D. (2005) Inference about the first-order autoregressive coefficient. *Communications in Statistics-Thery and Methods,34*, 1183-1201.

Rényi, A. (1962) *Wahrscheinlichkeitsrechnung mit einem Anhang über Informationstheorie*, Deutscher Verlag der Wissenschaften, Berlin.

Sakamoto, Y., Ishiguro, M., Kitagawa, G. (1986) *Akaike information criterion statistics*, KTK Scientific Publishers Tokyo and D.Reidel Publishing Company.

Singh, P.K., Singh,S.K. and Singh, U. (2008): Bayes estimator of inverse Gaussian parameters under general Entropy Loss Function using Lindley's approximation. *Communications in Statistics-Simulation and Computation, 37,* 1750-1762.

Zieliński, R. (1999) A median-unbiased estimator of the AR(1) coefficient. *Journal of Time Series Analysis*, *20*, 477-481.