A program to calculate the empirical bias in autocorrelation estimators

Jaume Amau and Roser Bono
University of Barcelona

The statistical analysis of short time series designs is influenced by the presence of serial dependence. Therefore, it is important to correctly estimate the first-order autocorrelation in behavioral data. The empirical bias is an indicator generally used to evaluate the adequacy degree of an estimator. This paper presents the Bias program, a Monte Carlo simulation program that generates first-order autoregressive processes and calculates the bias in three autocorrelation estimators ($r_1$, $r_{1+}$ and $r_{1'}$) for different values of the lag-one autocorrelation parameter and sample sizes. The program has been designed with MATLAB programming language and it runs on IBM-PC compatible computers with a 486 or later processor.

Programa para el cálculo del sesgo empírico de estimadores de autocorrelación. El análisis estadístico de los diseños de series temporales cortas está influído por la presencia de dependencia serial. De ahí la importancia de estimar correctamente la autocorrelación de primer orden en datos conductuales. El sesgo empírico es un indicador generalmente usado para evaluar el grado de precisión de un estimador. Este artículo presenta el programa Bias, un programa de simulación Monte Carlo que genera procesos autoregresivos de primer orden y calcula el sesgo de tres estimadores de autocorrelación ($r_1$, $r_{1+}$ y $r_{1'}$) para diferentes valores del parámetro de autocorrelación de retardo uno y tamaños muestrales. El programa ha sido diseñado mediante el lenguaje de programación MATLAB y funciona en ordenadores IBM-PC con un procesador 486 o superior.

The study of autocorrelation in time series was considered from the end of the seventies (Bono & Amau, 1996, 2000; Busk & Marrascuilo, 1988; Escudero & Vallejo, 2000; Glass, Willson, & Gottman, 1975; Gorsuch, 1983; Hartman et al., 1980; Huitema, 1985, 1988; Jones, Vaught, & Weinrott, 1977; Rosel, Jara, & Oliver, 1999; Sharpley & Alavosius, 1988; Suen, 1987; Suen & Ary, 1987; Vallejo, 1994, among others). In the nineties, the works were oriented towards the correction of the bias generated in short time-series by the conventional autocorrelation estimator $r_1$. Matyas and Greenwood (1991) verified that the difference between the mean of the conventional estimator $r_1$ and the value of the autocorrelation parameter $\rho$ increases with short time series. These authors concluded that the relationship between $\rho$ and the mean $r_1$ is nonlinear. The greatest coincidence between calculated mean $r_1$ and $\rho$ occurred when $r_1$ was around -0.2 or -0.3. According to these values, when $\rho$ increased positively, the calculated mean $r_1$ underestimated $r_1$ with a negative bias, while when $\rho$ increased negatively, the calculated mean $r_1$ underestimated $\rho$, but with a positive bias. Based on this simulation study, it was deduced that when small sample sizes were used, the $r_1$ values were biased. Huitema and McKeen (1991) proposed the modified estimator $r_{1+}$. This estimator corrects the empirical bias for positive values of the autocorrelation parameter by adding $1/n$ to $r_1$. Amau (1999) and Amau and Bono (2001) proposed the estimator $r_{1'}$ which consists of the correction of $r_1$ by the absolute value of a fitting model. This model is obtained from the polynomial function of the bias for each sample size. The estimator $r_{1'}$ reduces the bias for small sample sizes, both for positive as well as negative values of $\rho$.

This article describes a computer program using MATLAB, Version 5.2 (1998). The program, called Bias, makes it possible to calculate the empirical bias in $r_1$, $r_{1+}$ and $r_{1'}$ by Monte Carlo simulation. This program requires the specification of the parameter $\rho$ and the sample size ($n$). The structure of the program is as follows (see the Appendix): simulation of lag-one autoregressive processes, calculation of the estimators $r_1$ and $r_{1+}$, polynomial models for different sample sizes, calculation of the estimator $r_{1'}$, and estimation of the empirical bias in $r_1$, $r_{1+}$ and $r_{1'}$. Bias runs under the Windows 3.1 or later operating system on IBM-PC compatibles with a 486 or later processor, and 8 MB of extended memory.

In the first place, the program simulates first-order autoregressive processes. The *randn* function (Forsythe, Malcolm, & Moler, 1977) is used to generate observations from a lag-one autoregressive process:

$$Y_t = \rho_1 Y_{t-1} + \epsilon_t$$

where $Y_t$ is the score on the response measure at time $t$, $\rho_1$ is the autocorrelation parameter and $\epsilon_t$ is a random normal variable with a mean of zero and a variance of one. Each time series starts with a normal variate ($Y_0$) having zero mean and variance $1/(1-\rho_1^2)$. For each combination of $n$ and $\rho_1$, 10,000 samples are generated. The first 30 observations of each series are drop-
ped to eliminate any dependency between them (DeCarlo & Tryon, 1993; Huitema & McKean, 1991, 1994a, 1994b). After this, the program calculates the estimators \( r_j \) (equation 2) and \( r_{j+1} \) (equation 3) for each simulated time series. The conventional estimator \( r_1 \) is defined by the following equation:

\[
 r_1 = \frac{\sum_{i=1}^{n-1} (Y_i - \bar{Y}_n) (Y_{i+1} - \bar{Y}_n)}{\sum_{i=1}^{n-1} (Y_i - \bar{Y}_n)^2}
\]

(2)

where \( Y_i \) and \( Y_{i+1} \) are the data obtained in the intervals \( t \) and \( t+1 \) and \( \bar{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i \). The estimator \( r_1 \) is biased with small sample sizes because the numerator only includes \( n-1 \) terms instead of \( n \). Huitema and McKean (1991) proposed the modified estimator \( r_{1+} \) that corrects, to some extent, for the negative bias generated by positive autocorrelations

\[
 r_{1+} = r_1 + (1/n)
\]

(3)

The empirical bias function of the conventional estimator \( r_1 \) mildly drifts from the linearity (Arnau, 1999; Arnau & Bono, 2001; DeCarlo & Tryon, 1993; Huitema & McKean, 1991, 1994a, 1994b). Models with significant coefficients of the empirical bias function for \( n = 6, 10, 20 \) and \( 30 \) were derived with polynomial fitting. Within the MATLAB environment, another program that calculated the empirical bias of the estimator \( r_1 \) in function of \( \rho_1 \) and \( n \) by Monte Carlo simulation was designed (Arnau, 1999; Arnau & Bono, 2001). The polytool function of the Statistics Toolbox of the MATLAB was used for polynomial fitting. With the significant polynomial coefficients at 5% (Table 1) the following fitting models based on \( n \) were designed and were introduced into the Bias program.

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficients of the polynomial functions of the bias</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>Coefficients:</th>
<th>Linear</th>
<th>Quadratic</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>-0.1648</td>
<td>-0.5643</td>
<td>-0.0916</td>
</tr>
<tr>
<td></td>
<td>CI (0.05)</td>
<td>-0.1703</td>
<td>-0.5725</td>
</tr>
<tr>
<td>10</td>
<td>-0.0972</td>
<td>-0.3760</td>
<td>-0.0676</td>
</tr>
<tr>
<td></td>
<td>CI (0.05)</td>
<td>-0.1039</td>
<td>-0.3860</td>
</tr>
<tr>
<td>20</td>
<td>-0.0482</td>
<td>-0.2028</td>
<td>-0.0333</td>
</tr>
<tr>
<td></td>
<td>CI (0.05)</td>
<td>-0.0530</td>
<td>-0.2100</td>
</tr>
<tr>
<td>30</td>
<td>-0.0373</td>
<td>-0.1360</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CI (0.05)</td>
<td>-0.0406</td>
<td>-0.1470</td>
</tr>
</tbody>
</table>

For sample sizes other than those studied in this paper, it is previously necessary to calculate, by simulation, the empirical biases of the conventional estimator. Once the corresponding empirical biases have been found, the polynomial fitting is carried out with polytool function specifying the \( \rho_1 \) values, empirical biases previously obtained and the degree of the polynomial fit. In this way, the significant coefficients are obtained. These must be included in the polynomial models for different sample sizes statement indicating the value of \( n \).

Using the equations 4-7, the Bias program calculates the estimator \( r_{1+} \) by adding the corresponding polynomial model in absolute values to \( r_1 \).

\[
r_{1+} = r_1 + \text{Fitting Model}
\]

(8)

Finally, for each simulated sample, the program calculates the empirical bias \( r_1, r_{1+} \) and \( r_{1+} \), which is the difference between the mean value of the estimator and the value of the parameter \( \rho_1 \). The Bias program only calculates \( r_{1+} \) for four sample sizes (\( n = 6, 10, 20 \) and \( 30 \)). For other \( n \) values, the polynomial fitting of the bias must be performed first in order to determine the corresponding correction model and to introduce it into the program.

Running the program

The Bias program is a function M-file and the input variables in the MATLAB workspace, autoregressive parameter (rho) and sample size (n), must be introduced into it. As can be seen in the Appendix, the first line of the function M-file specifies the file name (bias), the input variables names (rho and n) and output variables names (r1_bias, r1plus_bias and r1prime_bias). The following expression must be specified in the MATLAB workspace in order to run the program: \( [r1\_bias, r1\_plus\_bias, r1\_prime\_bias]=\text{bias(rho,n)} \). The MATLAB executes the commands in Bias program and the results are shown in the MATLAB workspace.

Availability

Interested users who prefer not to type the program can request a file of the program by e-mailing the second author (rbono@psi.ub.es). It has been designed with the Version 5.2 of MATLAB but it can also work with previous versions.

Acknowledgements

The research was supported by Grant PS95-0228 from the General Council of Scientific and Technical Investigation, resolution of the Secretary of State of Universities and Investigation of the Ministry of Education and Culture (Spain).
A PROGRAM TO CALCULATE THE EMPIRICAL BIAS IN AUTOCORRELATION ESTIMATORS

APPENDIX
Program Listing

function [r1_bias, r1_plus_bias, r1_prime_bias]=bias(rho,n)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% LAG-ONE AUTOREGRESSIVE PROCESSES SIMULATION
% %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Number of simulations
for s=1:10000
% Number of observations
obs=30+n;
% Variance
variances=1/(1-rho^2);
% Standard deviation
sd=sqrt(variance);
% Series with mean 0 and fixed standard deviation
y0=normrnd(0,sd,obs,1);
% Random series with mean 0 and variance 1
y=randn(obs,1);
% First value of the series
y(1,1)=y0(1)+e(1);
% Next values of the series
for t=2:obs
 y(t,1)=rho*y(t-1,1)+e(t);
 t=t+1;
end
% Eliminations of the first 30 observations
y=y(31:obs);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% CALCULATION OF r1
% %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% y_mean=mean(y);
denominator=y-y_mean).^2;
denominator_sum=sum(denominator);
y_lag=y(2:n);
y=y(1:n-1);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% EMPIRICAL BIAS IN r1, r1+ AND r1'
% %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
r1_bias=mean(r1)-rho;
r1_plus_bias=mean(r1_plus)-rho;
r1_prime_bias=mean(r1_prime)-rho;

References


