Confidence intervals in stationary autocorrelated time series

Halkos, George and Kevork, Ilias

University of Thessaly, Department of Economics

2002

Confidence intervals in stationary autocorrelated time series

George E. Halkos and Ilias S. Kevork
Department of Economics, University of Thessaly

ABSTRACT

In this study we examine in covariance stationary time series the consequences of constructing confidence intervals for the population mean using the classical methodology based on the hypothesis of independence. As criteria we use the actual probability the confidence interval of the classical methodology to include the population mean (actual confidence level), and the ratio of the sampling error of the classical methodology over the corresponding actual one leading to equality between actual and nominal confidence levels. These criteria are computed analytically under different sample sizes, and for different autocorrelation structures. For the AR(1) case, we find significant differentiation in the values taken by the two criteria depending upon the structure and the degree of autocorrelation. In the case of MA(1), and especially for positive autocorrelation, we always find actual confidence levels lower than the corresponding nominal ones, while this differentiation between these two levels is much lower compared to the case of AR(1).

Keywords: Covariance stationary time series, Variance of the sample mean, Actual confidence level

1. INTRODUCTION

The basic assumption required at the stage of constructing confidence intervals for the mean, μ, of normally distributed populations is the observations in the sample to be independent. In a number of cases, however, the validity of this assumption should be seriously taken under consideration, and as a representative example we mention the problem of constructing confidence intervals for the average delay of customers in queuing systems. In such a case, it is very common the delays in a sample of n successive customers to display a certain degree of dependency at different lags, and therefore the application of the classical confidence interval estimator for the steady-state mean, μ,

$$\overline{X} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \tag{1}$$

based on independent, identical, and normal random variables not to be recommended.

Fishman (1978) shows that the variance of the mean of a sample $X_1, X_2, \ldots, X_n$ from a covariance stationary process is

$$\mathrm{Var}\left(\overline{X}_n\right) = \frac{\sigma_X^2}{n} \cdot h\left(\rho_s\right) \tag{2}$$

with

$$h\left(\rho_s\right) = 1 + 2\sum_{s=1}^{n-1}\left(1 - \frac{s}{n}\right)\rho_s \tag{3}$$

and ρ_s to be the s^{th} lag theoretical autocorrelation coefficient between any two variables whose time distance is s. Covariance stationary means that the mean and variance of $\{X_t, t = 1, 2, \ldots\}$ are stationary over time with common finite mean μ and common finite variance σ_X^2. Moreover for a covariance stationary process, the covariance between X_t and X_{t+s} depends only on the lag s and not on their actual values at times t and t+s.

For the last two decades, alternative estimators for (2) have been proposed in the literature in the context of estimating steady-state means in stationary simulation outputs. The reason for developing such variance estimators and not using directly the estimated values of the

autocorrelation coefficients in (2) is that, for s close to n, the estimation of ρ_s (s=1,2,…,n-1) will be not accurate as it will be based on few observations. On the other hand, Kevork (1990) showed that fixed sample size variance estimators, based on a single long replication, have two serious disadvantages. First, in finite samples they are biased. Second, the recommended values for their parameters at the estimation stage differ significantly according to the structure and the degree of the autocorrelation, which characterizes the process under consideration. Taking these two disadvantages into consideration at this stage, we are asked ourselves in what extent the application of these complicated variance estimators of (2) is necessary for covariance stationary processes. In other words can we avoid their use by investigating the consequences of applying the simple confidence interval estimator (1) to covariance stationary processes so that after making appropriate modifications to improve its performance?

Answers to the above questions are given in the current study. More specifically, assuming that the process under consideration follows either the first order autoregressive model, AR(1), or the first order moving average model, MA(1), we investigate the consequences of using (1) for estimating the steady-state mean in the light of the following two criteria: a) the difference between the nominal confidence level and the corresponding actual confidence level which is attained by (1); and b) the ratio of the sampling error of (1) over the corresponding real sampling error which ensures equality among nominal and actual confidence levels. These two criteria are computed analytically for the AR(1) and MA(1) under different values of the parameters φ and θ respectively, and for different sample sizes. The results for the AR(1) verify that the use of the complicated variance estimators for (2) is inevitable, especially when φ is positive and less than one. On the other hand, for the MA(1) the difference between a nominal confidence level of 95% and the achieved actual one is predictable as in low positive autocorrelations it ranges at 5%, while for moderate and high autocorrelations the difference remains almost constant with an average of 10%.

Under the above considerations, the structure of the paper is as follows: In section 2 we review the existing literature concerning the available variance estimators for (2). In section 3, we derive analytic forms for the special function of autocorrelation coefficients, $h(\rho_s)$, for AR(1) and MA(1). In the same section we specify the conditions when this function takes positive values less or greater than one. In section 4, we establish the methodology for computing analytically the actual confidence levels attained by using (1), that is, the actual probability this interval to

include the real steady-state mean of the covariance stationary process. Additionally, we present the actual confidence levels that (1) achieves in AR(1) and MA(1), for different degrees of autocorrelation under different sample sizes. Finally, the last section presents the main findings and conclusions of this research.

2. LITERATURE REVIEW

The presence of autocorrelation in simulation output may be a challenge for Inferential Statistics. This is because the lack of independence in the data becomes a serious problem and the calculation of elementary statistical measures like the standard error of the sample mean is incorrect. In particular, when time series data are positively autocorrelated the use of the classical standard error of the sample mean creates biases, which as a consequence reduces the coverage probabilities of confidence intervals.

Looking at the existing literature we may find different methods to overcome the problems of autocorrelation in the construction of confidence intervals for steady-state means. These methods are classified as, sequential, truncation and fixed sample size. Sequential confidence interval methods have as objective to determine the run length (sample size) of realizations of stationary simulation output processes which guarantees both an adequate correspondence between actual and nominal confidence levels and a pre-specified absolute or relative precision, as these terms are defined by Law (1983). Law and Kelton (1982a) distinguish these methods as regenerative and non-regenerative. Fishman's (1977) and Lavenberg and Sauer's (1977) methods belong to regenerative category while the methods developed by Mechanic and McKay (1966), Law and Carson (1978), Adam (1983) and Heidelberger and Welch (1981a) have been characterized as non-regenerative.

For the truncation methods the objective is the elimination of initialization bias effects on the estimation of the steady-state mean. These methods provide estimators for the time point t^* ($1 \leq t^* \leq n$) for which the absolute value of the difference between the expected value of the sample mean from the steady-state mean is greater than a pre-specified very small positive number e for any $t<t^*$. Generating r replications of a simulation output process $\{X_t\}$ under the same initial conditions, some of the truncation methods estimate t^* by applying the truncation rule to each replication (Fishman 1971, 1973b; Schriber, 1974; Heidelberger and Welch, 1983). Some others, however, estimate t^* from a pilot study, which is carried out on a number of

exploratory replications. Then the estimated value of t^* is used as the global truncation point in any other replication for which we use the same initial conditions (Conway, 1963; Gordon, 1969; Gafarian *et al*. 1978; Kelton and Law, 1983).

Fixed sample size confidence intervals methods propose different, asymptotically unbiased, estimators for the variance of the sample mean and these estimators may be used in the construction of confidence intervals. A number of confidence interval methods have been developed in the last decades in order to handle the problem.

The simplest fixed sample size confidence interval method is based on generating, for the process under consideration, k>1 independent replications of size m using independent steams of random numbers. When k is large enough, the variance of the k sample means is defined and used in the construction of confidence intervals, as these means are considered as independent, identical and normal random variables. But this method has practical difficulties, as it requires enormous systems and many hours of working time for the generation of just a single estimate.

Alternatively we may use single replication methods like the non-overlapping batch means (NOBM). This method (Law and Kelton, 1991; Fishman, 1999) divides a single long run into consecutive non-overlapping batches of size m, and from each batch an estimate of the performance measure is obtained. As it becomes obvious, these estimates are considered as equivalent to the corresponding ones, which are taken using independent replications. Specifically, if $\{X_t\}$ is a covariance stationary output process, the non-overlapping batch means method is based on generating a single long replication of $\{X_t\}$. Then, this replication is partitioned into k>1 contiguous and non-overlapping batches of size m. Provided that m is large enough and $\sum_{s=-\infty}^{\infty}|\gamma_s| < \infty$, Law and Garson (1978) showed that the non-overlapping batch means can be considered approximately uncorrelated and normal random variables. But as Song (1996) claims, the approximation of the correct batch size is possible but not trivial. At the same time, the construction of a confidence interval for a steady-state mean requires the satisfaction of normality and independency of the batch means.

Song and Schmeiser (1995) established the overlapping batch means method (OBM), which has smaller mean squared error in the estimation of the sample mean variance. Specifically, if n is the run length (sample size) of a single long replication of a covariance stationary output process $\{X_t\}$, the j^{th} overlapping batch mean of size m $[X_j(m)]$ may be defined

and in this context Welsh (1987) proposed for large m and n/m the following sample mean variance estimator $\hat{\sigma}^2_{OBM} = \frac{m}{n(n-m+1)} \sum_{j=1}^{n-m+1} [X_j(m) - \overline{X}_n]^2$. But Sargent *et al.* (1992) claim that NOBM is preferable to OBM when we construct confidence intervals relying on small samples and probably equivalent in the case of using large samples.

Next, let us consider the standardized time series methods. If $\{X_t\}$ is strictly stationary (the joint distribution of $X_{t_1}, X_{t_2}, ..., X_{t_n}$ is the same as the joint distribution of $X_{t_1+s}, X_{t_2+s}, ..., X_{t_n+s}$ for every t_1, t_2,…, t_n and s) and assuming also that this process is phi-mixing (for large s the correlation of X_t and X_{t+s} becomes negligible; see Law, 1983), the standardized time series methods use a functional central limit theorem to transform the sample X_1, X_2,…, X_n into a process which is asymptotically distributed as a Brownian Bridge process. Dividing a single long replication into k>1 contiguous and non-overlapping batches of size m, for m large and by using Brownian Bridge properties, Schruben (1983) derived four methods for estimating the variance of the sample mean. The area method, the maximum method, the combined area non-overlapping batch means method and the combined maximum non-overlapping batch means method. The standardized time series methods are easy to use and asymptotically have advantages over NOBM, but require long runs.

In these lines and as a parametric time series modeling of simulation output data, we consider the autoregressive method of Fishman (1978). This method assumes that $\{X_t\}$ is covariance stationary and can be represented by a p^{th} order autoregressive process, AR(p). Voss *et al.* (1996) derived good estimates of the steady state average queue delay using data from the transient phase of the simulation using a high-order AR(p) model. But such an autoregressive method is improper for widespread use as general ARIMA models are complex and assumptions for ARIMA modeling may be invalid for some particular simulation models.

The regenerative method was developed for the case in which the simulated process is characterized by the regenerative property and by enough regeneration cycles. This method was developed by Crane and Iglehart (1974a,b,c; 1975). Its principle is based on the identification of random points, where the process probabilistically starts over again. These points are called regeneration points. For instance, studying the delay in queue in the M/M/1 model, the indices of customers who find the system empty can be considered as regeneration points. The amount of data between two regeneration points is called the regeneration cycle. Then, the regeneration

points are used to obtain independent random variables to which inferential methods can be applied. In this context, two methods have been developed for estimating the steady state mean and producing confidence intervals, the classical and the Jacknife. A very good description of these methods is provided in Law and Kelton (1982b). It is worth mentioning here that the main disadvantage of these methods is the identification of regeneration points, especially for complicated simulation models. Specifically, the problem with this method exists when either there are no regeneration points for the output process or when the simulation cannot produce enough cycles.

A new and more recent approach to simulation output analysis relies on resampling methods, such as the Jackknife and the Bootstrap (Quenouille, 1949; Tuckey, 1958; Efron, 1979; Efron and Tibshirani, 1993), which provide non-parametric estimates of bias and standard error. The Bootstrap method relies on pseudo-data created by re-sampling the actual data, but it requires independency, which is not always the case in simulation outputs. The application of this method to time series data may work by re-sampling sets of consecutive observations in order to capture the autocorrelation structure. Various forms of the Bootstrap method appear in the literature. First, the Moving Blocks Bootstrap (MBB), which relies on random re-sampling of fixed size overlapping blocks with replacement (Künsch, 1989; Liu and Singh, 1992; Hall *et al.*, 1995). However, this method requires subjective inputs from the researcher and its estimates vary considerably.

Second, for stationary time series the Stationary Bootstrap (SB) was developed, where the data are re-sampled by contaminated blocks, which have a randomly chosen starting point and with their length geometrically distributed according to some chosen mean (Politis and Romano,1994). Under the same principle, Kim *et al.* (1993a) developed the Binary Bootstrap (BB) to analyze autocorrelated binary data. Kim *et al*. (1993b) introduced the Threshold Bootstrap (TB) extending the BB, and Park and Willemain (1999) modified the TB introducing the Threshold Jackknife (TJ). They claim that for various ARMA models, the TB has a better performance compared to MBB and SB in terms of estimating the standard error of the sample mean, if we optimize each re-sampling scheme with respect to the size of the re-sampling unit. They also show that the MBB has generally a poor performance.

Park *et al.* (2001) test the TB as a non-parametric method of output analysis and show that the TB is an effective alternative to the batch means and relatively easy. They also show that

the TB is more effective in the construction of confidence intervals for the steady state mean and median delay in the M/M/1 model, and establish the asymptotic unbiasedness and consistency of the TB estimators when we refer to the sample mean.

Finally, we have the spectral method where the process $\{X_t\}$ is assumed to be covariance stationary. At zero frequency, the power spectrum f(0) is estimated either by using the Tukey spectral window (Fishman (1973 a,b; Duket and Pritsker, 1978; Law and Kelton, 1984) or by using the periodogram coordinates as presented in Heidelberger and Welch (1981a,b).

3. THE FUNCTION $h(\rho_s)$ IN AR(1) AND MA(1)

3.1 AR(1)

This model is defined by $X_t = \varphi X_{t-1} + \varepsilon_t$, and is stationary when $|\varphi| < 1$. The ε_t's are uncorrelated and normal random variables with mean zero and common variance σ_ε^2. Substituting the s^{th} theoretical autocorrelation coefficient of this model, $\rho_s = \varphi^s$, to (3) we take

$$h(\rho_s) = 1 + 2\left\{\sum_{s=1}^{n-1}\varphi^s - \frac{1}{n}\sum_{s=1}^{n-1}s\varphi^s\right\} \qquad (4)$$

Given

$$\sum_{s=1}^{n-1}\varphi^s = \varphi\frac{1-\varphi^{n-1}}{1-\varphi}$$

and

$$\sum_{s=1}^{n-1}s\varphi^s = \frac{\varphi(1-\varphi^n) - n\varphi^n(1-\varphi)}{(1-\varphi)^2}$$

the function $h(\rho_s)$ takes for the AR(1) the form

$$h(\rho_s) = 1 + 2\left\{\frac{\varphi}{1-\varphi} - \frac{\varphi(1-\varphi^n)}{n(1-\varphi)^2}\right\} = \frac{1+\varphi}{1-\varphi} - \frac{2\varphi(1-\varphi^n)}{n(1-\varphi)^2} \qquad (5)$$

Subtracting -1 from both sides of (5)

$$h(\rho_s) - 1 = \frac{2\varphi}{1-\varphi}\{1 - \psi(n,\varphi)\}$$

where

$$\psi(n,\varphi) = \frac{1-\varphi^n}{n(1-\varphi)}$$

Given $|\varphi| < 1$, for any n≥2, $\psi(n,\varphi)$ takes always values in the interval (0,1), and this is illustrated in figures 1a, 1b, and 1c. Especially, when $-0.50<\varphi<1$, $\psi(n,\varphi)$ converges exponentially to zero. On the contrary, for $-1<\varphi<-0.50$, when n is small $\psi(n,\varphi)$ displays some oscillation which is getting larger and larger as φ approaches –1, while for n large this oscillation vanishes and the function converges again exponentially to zero.

The behaviour of $\psi(n,\varphi)$ leads us to the conclusion that when φ is positive, namely, the autocorrelation function converges exponentially to zero taking only positive values (positive autocorrelation), for any n, the function $h(\rho_s)$ takes values always greater than 1. This means that using the classical confidence interval estimator (1) we underestimate the real sampling error that the interval should have, and as a result we attain actual confidence levels lower that the corresponding nominal ones. On the other hand for $-1<\varphi<0$, that is, the autocorrelation function converges to zero oscillating between negative and positive values (negative autocorrelation), for n≥2 the half width of the classical estimator (1) overestimates the real sampling error, and this results in actual confidence levels greater than the corresponding nominal ones. The size of overestimating (or underestimating) the real sampling error by using (1), which is equal to $[h(\rho_s]^{-0.5}$, is displayed for different n and φ in table 1. When n is large (e.g. n>50), for the case of positive autocorrelation, the half width of the classical estimator (1) is at least 4 times narrower than the real sampling error, whereas for negative autocorrelation the real sampling error is overestimated approximately 3 times.

Figure 1α: $\psi(n,\varphi)$ for $0 < \varphi < 1$

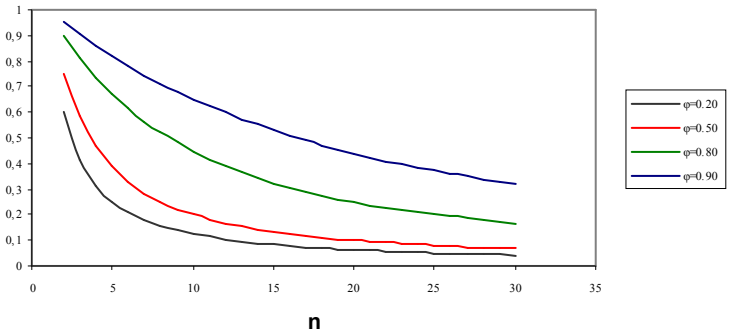

Figure 1b: $\psi(n,\varphi)$ for $-0.50 \le \varphi < 0$

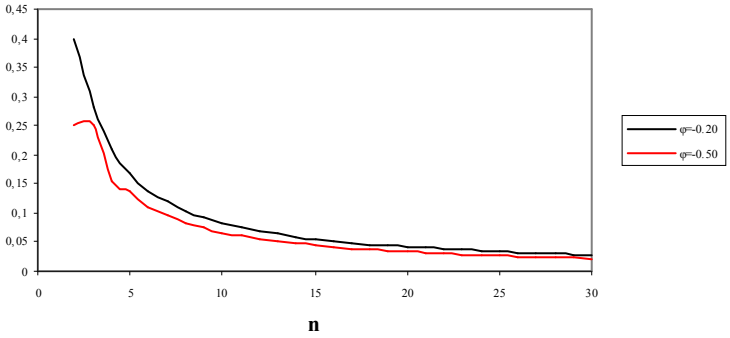

Figure 1c: $\psi(n,\varphi)$ for $-1 < \varphi < -0.50$

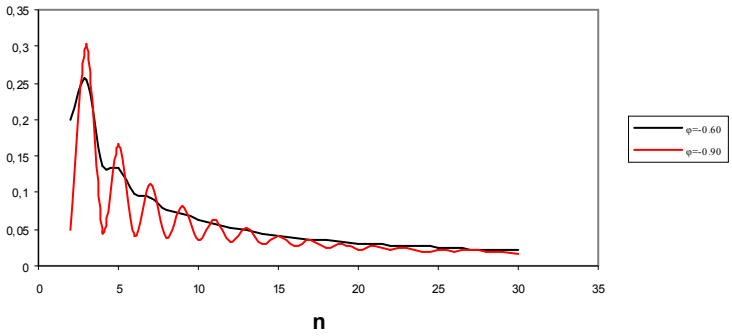

Table 1: Overestimating or underestimating the real sampling error in AR(1)

N	φ = -0.80	φ = -0.50	φ = -0.20	φ = 0.20	φ = 0.50	φ = 0.80	φ = 0.90
2	2.24	1.41	1.12	0.91	0.82	0.75	0.73
3	1.67	1.41	1.15	0.88	0.74	0.63	0.60
4	2.33	1.51	1.17	0.86	0.70	0.57	0.53
5	2.03	1.53	1.18	0.85	0.67	0.53	0.48
6	2.41	1.57	1.18	0.85	0.65	0.50	0.45
7	2.26	1.59	1.19	0.84	0.64	0.47	0.42
8	2.48	1.60	1.19	0.84	0.63	0.45	0.40
9	2.40	1.62	1.20	0.84	0.63	0.44	0.38
10	2.54	1.63	1.20	0.83	0.62	0.43	0.37
11	2.50	1.64	1.20	0.83	0.62	0.42	0.36
12	2.59	1.64	1.20	0.83	0.61	0.41	0.35
13	2.57	1.65	1.21	0.83	0.61	0.41	0.34
14	2.63	1.66	1.21	0.83	0.61	0.40	0.33
15	2.62	1.66	1.21	0.83	0.60	0.39	0.32
16	2.66	1.66	1.21	0.83	0.60	0.39	0.32
17	2.66	1.67	1.21	0.83	0.60	0.39	0.31
18	2.69	1.67	1.21	0.83	0.60	0.38	0.31
19	2.70	1.67	1.21	0.83	0.60	0.38	0.30
20	2.72	1.68	1.21	0.83	0.60	0.38	0.30
50	2.87	1.71	1.22	0.82	0.59	0.35	0.25
100	2.94	1.72	1.22	0.82	0.58	0.34	0.24
200	2.97	1.73	1.22	0.82	0.58	0.34	0.24
500	2.99	1.73	1.22	0.82	0.58	0.33	0.23

3.2 MA(1)

It is given by $X_t = \varepsilon_t + \theta\varepsilon_{t-1}$, and although the model is stationary for any θ, the invertibility condition restricts θ in the interval (-1,1). Substituting the autocorrelation function

$$\rho_s = \begin{cases} \dfrac{\theta}{1+\theta^2} & , \ s = 1 \\ 0 & , \ s > 1 \end{cases}$$

into (2) we take

$$h(\rho_s) = 1 + 2\frac{n-1}{n} \cdot \frac{\theta}{1+\theta^2} \qquad (6)$$

It is obvious that when θ is positive (negative), the function $h(\rho_s)$ takes values greater (positive and smaller) than one. So, as in the case of AR(1), under a positive (negative) autocorrelation the real sampling error is underestimated (overestimated) by using (3), attaining actual confidence levels lower (greater) than the nominal ones. Table 2, similar to table 1, illustrates for positive and negative autocorrelations the size of underestimating and overestimating respectively the real sampling error when we use the classical confidence interval estimator. Comparing the two tables we observe that the size of underestimation is smaller in MA(1) under positive autocorrelation especially in large samples, but for negative autocorrelation, the real sampling error is much more overestimated in MA(1) compared with the AR(1).

Table 2: Overestimating or underestimating the real sampling error in MA(1)

n	θ = -0.80	θ = -0.50	θ = -0.20	θ = 0.20	θ = 0.50	θ = 0.80	θ = 0.90
2	1.40	1.29	1.11	0.92	0.85	0.82	0.82
3	1.69	1.46	1.16	0.89	0.81	0.78	0.78
4	1.93	1.58	1.19	0.88	0.79	0.76	0.76
5	2.13	1.67	1.20	0.87	0.78	0.75	0.75
6	2.31	1.73	1.21	0.87	0.77	0.74	0.74
7	2.47	1.78	1.22	0.87	0.77	0.74	0.73
8	2.61	1.83	1.23	0.86	0.77	0.73	0.73
9	2.74	1.86	1.23	0.86	0.76	0.73	0.73
10	2.86	1.89	1.24	0.86	0.76	0.73	0.73
11	2.97	1.91	1.24	0.86	0.76	0.73	0.72
12	3.08	1.94	1.24	0.86	0.76	0.73	0.72
13	3.17	1.96	1.25	0.86	0.76	0.73	0.72
14	3.26	1.97	1.25	0.86	0.76	0.72	0.72
15	3.34	1.99	1.25	0.86	0.76	0.72	0.72
16	3.42	2.00	1.25	0.86	0.76	0.72	0.72
17	3.50	2.01	1.25	0.86	0.76	0.72	0.72
18	3.57	2.02	1.25	0.86	0.75	0.72	0.72
19	3.63	2.03	1.25	0.86	0.75	0.72	0.72
20	3.70	2.04	1.26	0.86	0.75	0.72	0.72
50	4.77	2.15	1.27	0.85	0.75	0.71	0.71
100	5.41	2.19	1.27	0.85	0.75	0.71	0.71
200	5.85	2.21	1.27	0.85	0.75	0.71	0.71
500	6.16	2.23	1.27	0.85	0.75	0.71	0.71

4. ACTUAL CONFIDENCE LEVELS ATTAINED BY THE CLASSICAL INTERVAL ESTIMATOR IN AR(1) AND MA(1)

Given that the random variables $X_1, X_2, \ldots, X_n$ from a covariance stationary process are normally distributed with steady-state mean μ and common standard deviation σ_X, the actual confidence interval for μ is derived from

$$\Pr\left\{-z_{\alpha_N/2} \le \frac{\overline{X}-\mu}{\frac{\sigma_X}{\sqrt{n}}\left[h(\rho_s)\right]^{1/2}} \le z_{\alpha_N/2}\right\} = 1-\alpha_N$$

as

$$\overline{X} - z_{\alpha_N/2}\frac{\sigma_X}{\sqrt{n}}\left[h(\rho_s)\right]^{1/2} \le \mu \le \overline{X} + z_{\alpha_N/2}\frac{\sigma_X}{\sqrt{n}}\left[h(\rho_s)\right]^{1/2}$$

where $1-\alpha_N$ is the nominal confidence level. Assuming, therefore, that X_t's are independent, and using the classical interval estimator (1), we ignore the function $h(\rho_s)$ of the theoretical autocorrelation coefficients. The omittion of $h(\rho_s)$ from (1) has as a result that although (1) is aimed at a nominal confidence level of $1-\alpha_N$, the attained actual probability to include μ is different from $1-\alpha_N$. We shall call this probability actual confidence level of the interval. This probability is analytically computed by

$$1-\alpha_A = \Pr\left\{-z_{\alpha_N/2} \le \frac{\overline{X}-\mu}{\frac{\sigma_X}{\sqrt{n}}} \le z_{\alpha_N/2}\right\} =$$

$$= \Pr\left\{-\frac{z_{\alpha_N/2}}{\left[h(\rho_s)\right]^{1/2}} \le \frac{\overline{X}-\mu}{\frac{\sigma_X}{\sqrt{n}}\left[h(\rho_s)\right]^{1/2}} \le \frac{z_{\alpha_N/2}}{\left[h(\rho_s)\right]^{1/2}}\right\} =$$

$$= \Pr\left\{-z^*_{\alpha_N/2} \le Z \le z^*_{\alpha_N/2}\right\} = 1-2\Phi\left(-z^*_{\alpha_N/2}\right) \qquad (7)$$

where $\Phi\left(-z^{*}_{\alpha_N/2}\right)$ is the cumulative distribution function of the standard normal evaluated at

$$-z^{*}_{\alpha_N/2} = -\frac{z_{\alpha_N/2}}{\left[h(\rho_s)\right]^{1/2}}$$

for a nominal confidence level $1-\alpha_N$.

With reference to AR(1) and MA(1), at nominal confidence level 0.95, tables 3 and 4 present the actual confidence levels attained by the classical interval estimator (1) under different values of φ and θ respectively. These actual confidence levels have been computed analytically after substituting the exact values of $\left[h(\rho_s\right]$, obtained by using the values of φ and θ in (5) and (6) respectively, into (7). Regarding AR(1), for $0<\varphi<1$, the actual confidence levels not only are lower than 0.95, but also are declining as the sample size increases. The same holds for a given n where, as φ approaches one, the actual confidence levels are decreasing again. The last two remarks make obvious that for large n and heavy autocorrelations, using (1) we attain actual confidence levels which are far away from the corresponding nominals. On the other hand, with φ taking values on the interval (-1,0), the actual confidence levels, being always greater than the nominal one, are increasing by drawing larger and larger samples.

Similar pattern of changes for the actual confidence levels are observed in the MA(1). However, for θ close to one, the differences between the actual and nominal levels are not so great as these differences were in the case of AR(1). Additionally, given n, the attained confidence levels for the MA(1) display some stability at certain intervals of θ. So, for low values of θ and large samples ($n\geq 50$) the difference between the nominal and the actual confidence level is approximately at 5%, while for moderate and large values of θ ($\theta>0.50$) this difference ranges on average at 10%. On the contrary, for θ negative, in large sample the actual confidence level is very close to 100%.

Table 3: AR(1): Actual confidence levels of the classical confidence interval estimator for the stationary mean at nominal confidence level 95%

n	φ = -0.80	φ = -0.50	φ = -0.20	φ = 0.20	φ = 0.50	φ = 0.80	φ = 0.90
2	1.00	0.99	0.97	0.93	0.89	0.86	0.84
3	1.00	0.99	0.98	0.92	0.85	0.79	0.76
4	1.00	1.00	0.98	0.91	0.83	0.73	0.70
5	1.00	1.00	0.98	0.91	0.81	0.70	0.66
6	1.00	1.00	0.98	0.90	0.80	0.67	0.62
7	1.00	1.00	0.98	0.90	0.79	0.65	0.59
8	1.00	1.00	0.98	0.90	0.78	0.63	0.57
9	1.00	1.00	0.98	0.90	0.78	0.61	0.55
10	1.00	1.00	0.98	0.90	0.78	0.60	0.53
11	1.00	1.00	0.98	0.90	0.77	0.59	0.52
12	1.00	1.00	0.98	0.90	0.77	0.58	0.51
13	1.00	1.00	0.98	0.90	0.77	0.57	0.49
14	1.00	1.00	0.98	0.90	0.77	0.57	0.48
15	1.00	1.00	0.98	0.90	0.76	0.56	0.48
16	1.00	1.00	0.98	0.90	0.76	0.56	0.47
17	1.00	1.00	0.98	0.89	0.76	0.55	0.46
18	1.00	1.00	0.98	0.89	0.76	0.55	0.45
19	1.00	1.00	0.98	0.89	0.76	0.54	0.45
20	1.00	1.00	0.98	0.89	0.76	0.54	0.44
50	1.00	1.00	0.98	0.89	0.75	0.51	0.38
100	1.00	1.00	0.98	0.89	0.75	0.50	0.36
200	1.00	1.00	0.98	0.89	0.74	0.49	0.35
500	1.00	1.00	0.98	0.89	0.74	0.49	0.35

Table 4: MA(1): Actual confidence levels of the classical confidence interval estimator for the stationary mean at nominal confidence level 95%

n	θ = -0.80	θ = -0.50	θ = -0.20	θ = 0.20	θ = 0.50	θ = 0.80	θ = 0.90
2	0.99	0.99	0.97	0.93	0.90	0.89	0.89
3	1.00	1.00	0.98	0.92	0.89	0.87	0.87
4	1.00	1.00	0.98	0.92	0.88	0.86	0.86
5	1.00	1.00	0.98	0.91	0.87	0.86	0.86
6	1.00	1.00	0.98	0.91	0.87	0.85	0.85
7	1.00	1.00	0.98	0.91	0.87	0.85	0.85
8	1.00	1.00	0.98	0.91	0.87	0.85	0.85
9	1.00	1.00	0.98	0.91	0.87	0.85	0.85
10	1.00	1.00	0.98	0.91	0.86	0.85	0.85
11	1.00	1.00	0.98	0.91	0.86	0.85	0.84
12	1.00	1.00	0.99	0.91	0.86	0.85	0.84
13	1.00	1.00	0.99	0.91	0.86	0.84	0.84
14	1.00	1.00	0.99	0.91	0.86	0.84	0.84
15	1.00	1.00	0.99	0.91	0.86	0.84	0.84
16	1.00	1.00	0.99	0.91	0.86	0.84	0.84
17	1.00	1.00	0.99	0.91	0.86	0.84	0.84
18	1.00	1.00	0.99	0.91	0.86	0.84	0.84
19	1.00	1.00	0.99	0.91	0.86	0.84	0.84
20	1.00	1.00	0.99	0.91	0.86	0.84	0.84
50	1.00	1.00	0.99	0.91	0.86	0.84	0.84
100	1.00	1.00	0.99	0.90	0.86	0.84	0.84
200	1.00	1.00	0.99	0.90	0.86	0.84	0.84
500	1.00	1.00	0.99	0.90	0.86	0.84	0.84

5. CONCLUSIONS

In this study, we examined in covariance stationary processes the performance of the classical confidence interval estimator for the steady-state mean. One of the assumptions for deriving this estimator refers to the independence of random variables in the sample. The following two criteria were used: a) The actual probability, called as actual confidence level, the classical confidence interval estimator to include the steady-state mean, given the nominal confidence level; and b) the ratio of the sampling error of the classical confidence interval estimator over the corresponding true one which ensures equality between actual and nominal confidence levels. These criteria are computed analytically for the stationary AR(1) and MA(1) models, for different values of φ and θ respectively.

For the AR(1), when the autocorrelation converges exponentially to zero taking on positive values, the actual confidence levels attaining by the classical estimator, being always lower than the corresponding nominal confidence levels, are decreasing as the sample is getting larger and larger. Especially, for the case of heavy autocorrelation and large samples, the actual confidence levels are dramatically low as they range even less than 40%. In such cases the classical confidence interval estimator underestimates the true sampling error over four times. On the contrary, when the autocorrelation function converges to zero oscillating between positive and negative values, the classical estimator overestimates the true sampling error, and as a result, we always attain actual confidence levels greater than the corresponding nominal ones. As a concluding remark for the AR(1), therefore, we can say that the behaviour of the two criteria under consideration is differentiated substantially according to the structure and the level of autocorrelation.

Regarding MA(1), we always observe for positive autocorrelation actual confidence levels lower than the corresponding nominal ones. However, the discrepancies between these two levels are much smaller and more predictable compared to the case of AR(1). Particularly, for large samples, when the autocorrelation is light, these discrepancies range at 5%, while for moderate or heavy autocorrelations the discrepancies display very little differentiation at an average level of 10%. It is also worthwhile to mention that in MA(1), for negative autocorrelations the actual confidence levels are almost 100%, and this is due the fact that the true sampling error is highly overestimated. Especially in large samples the half-width of the classical confidence interval estimator overestimates the true sampling error by more than five times.

REFERENCES

Adam, N.R., 1983. Achieving a confidence interval for parameters estimated by simulation. Management Science, Vol.29, pp. 856-866.

Conway, R.W., 1963. Some tactical problems in digital simulation. Management Science, Vol. 10, pp. 47-61.

Crane, M.A., and D.L. Iglehart, 1974a. Simulating stable stochastic systems, I: General multi-server queues. Journal of the Association for Computing Machinery, Vol. 21, pp.103-113.

Crane, M.A., and D.L. Iglehart, 1974b. Simulating stable stochastic systems, II: Markov chains. Journal of the Association for Computing Machinery, Vol. 21, pp.114-123.

Crane, M.A., and D.L. Iglehart, 1974c Simulating stable stochastic systems, III: Regenerative processes and discrete event simulations. Operations Research, Vol. 23, pp.33-45.

Crane, M.A., and D.L. Iglehart, 1975. Simulating stable stochastic systems, IV: Approximation techniques. Management Science, Vol. 21, pp.1215-1224.

Ducket, S.D., and A.A.B. Pritsker, 1978. Examination of simulation output using spectral methods. Mathematical Computing Simulation, Vol. 20, pp. 53-60.

Efron, B., 1979. Bootstrap methods: Another look at the jackknife. The Annals of Statistics 7, 1-26

Efron, B. and Tibshirani, R., 1993. An Introduction to the Bootstrap. Chapman & Hall, New York.

Fishman, G.S., 1971. Estimating the sample size in computing simulation experiments. Management Science, Vol. 18, pp. 21-38.

Fishman, G.S., 1973a. Statistical analysis for queuing simulations. Management Science, Vol. 20, pp. 363-369.

Fishman, G.S., 1973b. Concepts and methods in discrete event digital simulation. John Wiley and Sons, New York.

Fishman, G.S., 1977. Achieving specific accuracy in simulation output analysis. Communication of the Association for computing Machinery, Vol. 20, pp. 310-315.

Fishman, G., 1978. Principles of Discrete Event Simulation. Wiley, New York.

Fishman, G., 1999. Monte Carlo: Concepts, Algorithms, and Applications. Springer, New York.

Gafarian, A.V., Ancker, C.J., JR, and T. Morisaku, 1978. Evaluation of commonly used rules for detecting steady state in computer simulation. Naval Research Logistics Quarterly, Vol. 25, pp. 511-529.

Gordon, G., 1969. System simulation. Prentice-Hall, Englewood Cliffs N.j.

Hall, P., Horowitz, J. and Jing, B.-Y., 1995. On blocking rules for the bootstrap with dependent data. Biometrika 82, 561-574.

Heidelberger, P., and P.D. Welch, 1981a. A spectral method for confidence interval generation and run length control in simulations. Communications of the Association for Computing Machinery, Vol. 24, pp. 233-245.

Heidelberger, P., and P.D. Welch, 1981b. Adaptive spectral methods for simulation output analysis. IBM Journal of Research and Development, Vol. 25, pp. 860-876.

Heidelberger, P., and P.D. Welch, 1983. Simulation run length control in the presence of an initial transient. Operations Research, Vol. 31, pp. 1109-1144.

Kevork, I.S, 1990. Confidence Interval Methods for Discrete Event Computer Simulation: Theoretical Properties and Practical Recommendations. Unpublished Ph.D. Thesis, University of London, London

Kim, Y., Haddock, J. and Willemain, T., 1993a. The binary bootstrap: Inference with autocorrelated binary data. Communications in Statistics: Simulation and Computation 22, 205-216.

Kim, Y., Willemain, T., Haddock, J. and Runger, G., 1993b. The threshold bootstrap: A new approach to simulation output analysis. In: Evans, G.W., Mollaghasemi, M., Russell, E.C., Biles, W.E. (Eds.), Proceedings: 1993 Winter Simulation Conference, pp. 498-502.

Kelton, D.W. and A.M. Law, 1983. A new approach for dealing with the startup problem in discrete event simulation. Naval Research Logistics Quarterly, Vol. 30, pp. 6410658.

Künsch, H., 1989. The jackknife and the bootstrap for general stationary observations. The Annals of Statistics 17, 1217-1241.

Lavenberg, S., S., and C. H. Sauer, 1977. Sequential stopping rules for the regenerative method of simulation. IBM Journal of Research and Development, Vol. 21, pp. 545-558.

Law, A.M., 1983. Statistical analysis of simulation output data. Operations Research, Vol. 31, pp. 983-1029.

Law, A.M., and J.S. Carson, 1978. A sequential procedure for determining the length of a steady state simulation. Operation Research, Vol. 27, pp. 1011-1025.

Law, A.M., and W.D. Kelton, 1982a. Confidence interval for steady state simulations: II. A survey of sequential procedures. Management Science, Vol. 28, pp. 560-562.

Law, A.M., and W.D. Kelton, 1982b. Simulation modelling and analysis. McGraw Hill, New York.

Law, A.M., and W.D. Kelton, 1984. Confidence intervals for steady state simulations: I. A survey of fixed sample size procedures. Operation Research, Vol. 32, pp. 1221-1239.

Law, A. and Kelton, W., 1991. Simulation Modeling and Analysis, second ed. McGraw-Hill, New York.

Liu, R. and Singh, K., 1992. Moving blocks jackknife and bootstrap capture weak dependence. In: Le Page, R., Billard, L., (Eds.), Exploring the Limits of Bootstrap. Wiley, New York, pp.225-248.

Mechanic, H., and W. McKay, 1966. Confidence intervals for averages of dependent data in simulation II. Technical report 17-202 IBM, Advanced Systems Development Division.

Park, D. and Willemain, T., 1999. The threshold bootstrap and threshold jackknife. Computational Statistics and Data Analysis 31, 187-202.

Park, D.S., Kim, Y.B., Shin, K.I. and Willemain, T.R., 2001. Simulation output Analysis using the threshold bootstrap, European Journal of Operational Research 134, 17-28.

Quenouille, M., 1949. Approximation tests of correlation in time series. Journal of Royal Statistical Society Series B 11, 68-84.

Schriber, T.J., 1974. Simulation using GPSS. John Wiley and Sons, New York.

Sargent, R.G., Kang, K. and Goldsman, D., 1992. An investigation of finite-sample behavior of confidence interval estimators. Operation Research 40, 898-913.

Schruben, L., 1983. Confidence interval estimation using standardized time series. Operations Research 31, 1090-1108.

Song, W.T., 1996. On the estimation of optimal batch sizes in the analysis of simulation output. European Journal of Operational Research 88, 304-319.

Song, W.T. and Schmeiser, B.W., 1995. Optimal mean-squared-error batch sizes. Management Science 41, 111-123.

Tukey, J., 1958. Bias and confidence interval in not quite large samples (Abstract). The Annals of Mathematical Statistics 29, 614.

Voss, P., Haddock, J. and Willemain, T., 1996. Estimating steady state mean from short transient simulations. In: Charnes, J.M., Morrice, D.M., Brunner, D.T.

Welch, P.D., 1987. On the relationship between batch means, overlapping batch means and spectral estimation. Proceedings of the 1987 Winter Simulation Conference, pp. 320-323.