

## **Some Properties of First Order Sample Autocorrelations of Selected ARMA Models: A Simulation Study**

**A.M. Barry and I.H. Khan**

*Department of Statistics and Operations Research, King Saud University  
P.O. Box 2455, Riyadh 11451, Saudi Arabia*

(Received 22 February 1993; Accepted for publication 24 April 1994)

**Key words:** ARMA models, Autocorrelations, Simulation.

**Abstract.** Since sample autocorrelations play a key role in identification of time series models, the study of their properties has always been of great interest in the literature of time series analysis. The assumption of normality for the distribution of autocorrelations made its mean and variance the main focus of study. Various methods are employed for estimating the variance but results are valid only asymptotically and with approximations sacrificing the accuracy.

The present study explores the empirical distribution using simulation techniques whereby we can test the reliability of these asymptotic results for small samples. Graphical methods are used to test how far the assumptions of normality holds for the distribution of sample autocorrelations. The extent of departure from normality for different models is noted and reported.

### **Introduction**

Autoregressive Moving Average (ARMA) models have been used successfully so far for representing time series ensemble [1; 2]. The most important tool to identify these models is the autocorrelation function which, for a zero mean, second order, stationary series,  $X_t$  is defined as

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

where

$$\gamma_k = E[(X_t)(X_{t+k})], \quad k = 0, 1, 2, \dots$$

In recent statistical literature there have been use of simulation studies in time series analysis for example Bataglia [3]. Also, there are some studies where Monte Carlo methods are employed to investigate some properties of ARMA models for

example [4; 5]. Since most of the theoretical results obtained to determine the parameters of central location and dispersion of sample autocorrelations are valid only asymptotically, this opens up the desirability of simulation studies to validate these results. In this work a simulation procedure is used to investigate these theoretical results. The study also pays attention to the normality assumption for the distribution of sample autocorrelations. The departure from this assumption of normality is noted for different models and reported.

### Location and Dispersion Parameters for Sample ACF

For a given observed time series  $X_1, X_2, \dots, X_n$ , the sample ACF is defined as

$$r_k = \hat{\rho}_k = \frac{\sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X})}{\sum_{t=1}^n (X_t - \bar{X})^2},$$

where  $\bar{X} = \sum_{t=1}^n X_t/n$ , the sample mean of the series.

From the definition of  $r_k$ , a ratio of quadratic functions of  $X$ 's, it should be apparent that the sampling properties of  $r_k$  will not be easily obtained. Even the expected value of  $r_k$  is difficult to derive. We have to accept a general large sample result and consider its implications in special cases.

For a stationary Gaussian process, Bartlett [6] has shown that for  $k > 0$ , and  $k + j > 0$ ,

$$\begin{aligned} \text{Cov}(r_k, r_{k+j}) \cong & \frac{1}{n} \sum_{i=-\infty}^{\infty} (\rho_i \rho_{i+j} + \rho_{i+k+j} \rho_{i-k}) \\ & - 2\rho_k \rho_i \rho_{i-k-j} - 2\rho_{k+j} \rho_i \rho_{i-k} + 2\rho_k \rho_{k+j} \rho_i^2 \end{aligned}$$

For large  $n$  (e.g. [7; 8]),  $r_k$  is approximately normally distributed with mean  $\rho_k$  and variance

$$\text{Var}(r_k) \cong \frac{1}{n} \sum_{i=-\infty}^{\infty} (\rho_i^2 + \rho_{i+k} \rho_{i-k} - 4\rho_k \rho_i \rho_{i-k} + 2\rho_k^2 \rho_i^2)$$

For processes in which  $\rho_k = 0$  for  $k > m$ , Bartlett's approximation becomes  $\text{var}(r_k) \cong \frac{1}{n} (1 + 2\rho_1^2 + \dots + 2\rho_m^2)$ .

If  $X_t$  is a purely random process (white noise), the variance of  $r_k$  reduces to

$$\text{Var}(r_k) = \frac{1}{n}, \quad k = 0, 1, 2, \dots$$

This has been considered as a crude theoretical estimate for the variance of  $r_k$  for any ARMA model.

If  $X_t$  is modelled as  $X_t = \phi X_{t-1} + a_t$ , with  $\rho_k = \phi^k$ ,  $k > 0$ , then [7; 9]

$$\text{Var}(r_k) = \frac{1}{n} \left\{ \frac{(1 + \phi^2)(1 - \phi^{2k})}{1 - \phi^2} - 2k\phi^{2k} \right\}$$

In particular,

$$\text{Var}(r_1) = \frac{1}{n} \{1 - \phi^2\}$$

### Monte Carlo Experiment

A simulation procedure is used to generate realizations from some ARMA models. The disturbances are generated as mutually independent and uncorrelated random normal variates. The IMSL subroutine GGNML is used to generate these numbers for different seed values. The algorithm employed has been rigorously tested by Learmonth and Lewis [10]. The series length selected are 500, 300, 100 and 50 representing large size, medium size and small size series. It will give us an opportunity to compare empirical results of varying series sizes with the corresponding theoretical results. The ARMA models used to generate the series are AR(1), MA(1) and ARMA (1,1).

The first 200 values of each series were discarded to get rid of the transient effect.

### Simulation Results for AR(1) Model

To conduct a Monte Carlo experiment, first we generated series of length 500 for AR(1) model,  $X_t = \phi X_{t-1} + a_t$  with  $\phi = 0.9$ . We obtained 1000 realizations for this model and computed  $r_1$  for each realization. Thus we got a sample size of 1000 for  $r_1$  and estimated the parameters of central location and dispersion for  $r_1$ . This procedure is repeated for the AR(1) models,  $X_t = \phi X_{t-1} + a_t$  with  $\phi = 0.6$  and  $\phi = 0.2$ . The results so obtained are compared with the corresponding theoretical results and are set in Table 1.

Table 1. Results of  $r_1$  for AR(1) models with series length 500

No of realizations for each model = 1000						
Parameter	Empirical results			Asymptotic results		
	Mean ( $r_1$ )	Var ( $r_1$ )	skewness	Mean ( $r_1$ )	Var ( $r_1$ ) = $\frac{1}{n}(1-\phi^2)$	$\frac{1}{n}$
$\phi$						
0.9	0.874	0.0108	-5.794	0.9	0.00038	0.002
0.6	0.584	0.0094	-2.019	0.6	0.00128	0.002
0.2	0.191	0.0026	-0.524	0.2	0.00192	0.002

The experiment is repeated with series length 300, 100 and 50. The results of the experiment are set in Table 2, 3 and 4.

For the AR(1) models with  $\phi = 0.9$  from Tables 1, 2, 3 and 4, we see that the empirical mean of  $r_1$  is departing to the left of the normal curve as compared to the theoretical mean and this deviation to the left is increasing successively for the series length of 500, 300, 100 and 50 which is in accordance with the assumption of asymptoticity for the theoretical mean of  $r_1$ . These findings are confirmed when the distribution of  $r_1$  for these cases are displayed by frequency curves as shown in Figs. 1a1, 1a4, 1a7 and 1a10. The same pattern in the behaviour of the empirical mean of  $r_1$  for AR(1) models with  $\phi = 0.6$  is also emerging as shown in Tables 2, 3 and 4. The graphs in Figs 1a2, 1a5, 1a8 and 1a11 are again confirming this behavior. For AR(1) model with  $\phi = 0.2$ , the behavior of the mean of  $r_1$  is similar to the former models. The mean still deviates from the theoretical mean as for the other models. It still increases with the decrease in sample size. The deviation is minimum for the series of sample size 500 and it is maximum for the series of sample size 50 which is again in confirmation with the assumption that the distribution of  $r_1$  is approximately normal when the sample size tends to infinity with its mean approaching the theoretical autocorrelation. If we look at the AR(1) model with varying parameter values but having the same sample size, we find that the percentage variation of the empirical mean compared to the theoretical mean is decreasing successively for  $\phi = 0.9$ ,  $\phi = 0.6$  and for  $\phi = 0.2$ .

The behavior of the empirical variance of  $r_1$  as compared with the theoretical variance for the models considered is as follows. For the AR(1) model with  $\phi = 0.9$ , we find that the difference between the empirical variance and the theoretical approximation is decreasing with decreasing series length. The same behavior is observed for AR(1) model with  $\phi = 0.6$ . For the AR(1) model with  $\phi = 0.2$ , it seems that changes in series length also effect the difference between the empirical variance of  $r_1$  and its theoretical approximation similar to other two cases.

The behavior of these variances is also depicted in Figs 1a1-1a10.

Table 2. Results of  $r_1$  for AR(1) models with series length 300

No of realizations for each model = 1000						
Parameter	Empirical results			Asymptotic results		
	Mean ( $r_1$ )	Var ( $r_1$ )	skewness	Mean ( $r_1$ )	Var ( $r_1$ ) = $\frac{1}{n}(1-\phi^2)$	$\frac{1}{n}$
$\phi$						
0.9	0.859	0.0150	-4.486	0.9	0.000633	0.0033
0.6	0.573	0.0080	-3.012	0.6	0.00213	0.0033
0.2	0.189	0.0037	-0.191	0.2	0.00320	0.0033

Table 3. Results of  $r_1$  for AR(1) models with series length 100

No of realizations for each model = 1000						
Parameter	Empirical results			Asymptotic results		
	Mean ( $r_1$ )	Var ( $r_1$ )	skewness	Mean ( $r_1$ )	Var ( $r_1$ ) = $\frac{1}{n}(1-\phi^2)$	$\frac{1}{n}$
$\phi$						
0.9	0.826	0.0170	-3.496	0.9	0.0019	0.01
0.6	0.554	0.0120	-1.268	0.6	0.0064	0.01
0.2	0.177	0.0098	-0.093	0.2	0.0096	0.01

Table 4. Results of  $r_1$  for AR(1) models with series length 50

No of realizations for each model = 1000						
Parameter	Empirical results			Asymptotic results		
	Mean ( $r_1$ )	Var ( $r_1$ )	skewness	Mean ( $r_1$ )	Var ( $r_1$ ) = $\frac{1}{n}(1-\phi^2)$	$\frac{1}{n}$
$\phi$						
0.9	0.792	0.0092	-1.015	0.9	0.0038	0.02
0.6	0.533	0.0139	-0.368	0.6	0.0128	0.02
0.2	0.165	0.0185	-0.061	0.2	0.0192	0.02

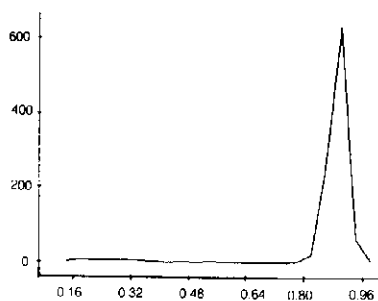


Fig. 1a1. Frequency curve for AR(1) of series length 500 with parameter value 0.9.

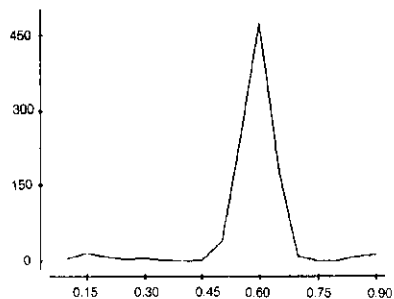


Fig. 1a2. Frequency curve for AR(1) of series length 500 with parameter value 0.6.

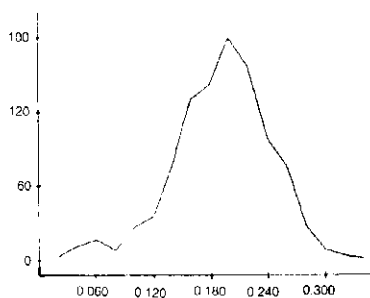


Fig. 1a3. Frequency curve for AR(1) of series length 500 with parameter value 0.2.

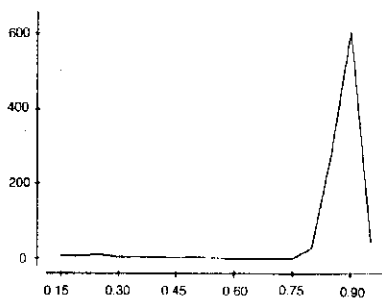


Fig. 1a4. Frequency curve for AR(1) of series length 300 with parameter value 0.9.

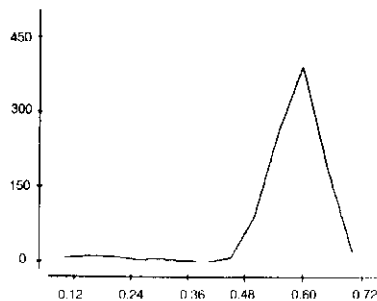


Fig. 1a5. Frequency curve for AR(1) of series length 300 with parameter value 0.6

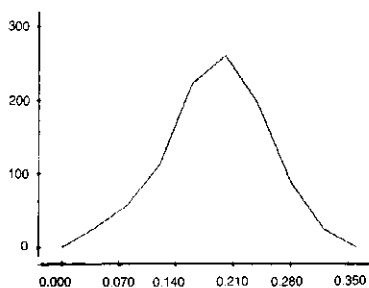


Fig. 1a6. Frequency curve for AR(1) of series length 300 with parameter value 0.2

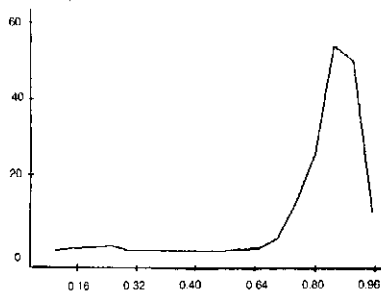


Fig. 1a7. Frequency curve for AR(1) of series length 100 with parameter value 0.9

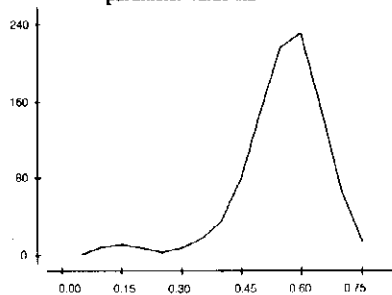


Fig. 1a8. Frequency curve for AR(1) of series length 100 with parameter value 0.2

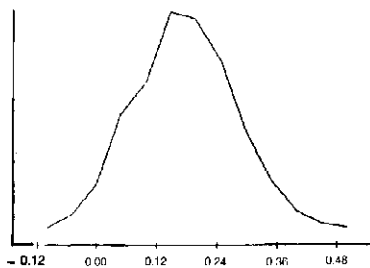


Fig. 1a9. Frequency curve for AR(1) of series length 100 with parameter value 0.6

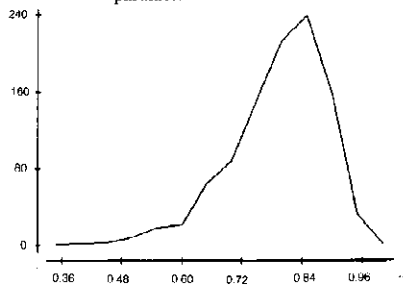


Fig. 1a10. Frequency curve for AR(1) of series length 50 with parameter value 0.9

### Simulation Results for MA(1) Model

For MA(1) model,  $X_t = a_t - \theta a_{t-1}$ , we generated time series of lengths 500, 300, 100 and 50. The values for the parameter  $\theta$  are taken as 0.9, 0.6 and 0.2 for each series length. For each combination of parameter values and series length 1000 realizations are obtained. The autocorrelation at lag 1 ( $r_1$ ) is computed for each realization of the model. Similar to AR(1) model, the empirical and theoretical results are compared. These results are set in Table 5, 6, 7 and 8 as follows:

For MA(1) models with  $\theta = 0.9$ , the results indicate that the empirical means deviate from the theoretical means to the same extent for the series length of 500 and 300. The amount of deviation for the series length 100 and series length 50 is greater for the empirical mean as compared to the theoretical mean. The similar pattern is repeated for MA(1) model with  $\theta = 0.6$ . In case of MA(1) model with  $\theta = 0.2$ , the difference between empirical mean and theoretical mean is fluctuating as series length varies from 500 to 50.

For all MA(1) models under consideration, empirical variances of  $r_1$  are smaller than the corresponding theoretical variances except for series length 500.

### Simulation Results for ARMA(1,1) Models

We discuss simulation results for ARM(1,1) models  $X_t = \phi X_{t-1} + a_t - \theta a_{t-1}$ , for different values of  $\phi$  and  $\theta$ . These results are set in Tables 9, 10, 11 and 12. The mean of sample autocorrelation  $r_1$  for series length 500 is very much affected by the variation in parameter values as shown in Table 9. The value of mean in all cases is not far from the theoretical mean. This fact is further established if we look Figs 2a1-2a2. These figures illustrate that the distribution of sample autocorrelation  $r_1$  is very close to normal distribution. Therefore the results validate theoretical assumption of approximate normal distribution for sample autocorrelation  $r_1$ .

In case of series length 300 the behavior of sample autocorrelation  $r_1$  is very similar to the previous cases shown in Table 10 and Figs 2a3-2a4. The empirical variances are bigger than before but graphical representation display a very close proximity to normal distribution and therefore validate theoretical assumptions. If we look at Table 11 we find that the difference between empirical means and theoretical means is slightly larger than in the two previous cases and empirical variances are also larger than the case of series length 300. This pattern continues to series length 50. The graphical representation as shown in Figs 2a5-2a8 again validate the theoretical assumptions.

**Table 5. Results of  $r_1$  for MA(1) models for series length 500**

No of realizations for each model = 1000						
Parameter	Empirical results			Asymptotic results		
	Mean ( $r_1$ )	Var ( $r_1$ )	skewness	Mean ( $r_1$ )	$(1 + 2\rho_1^2)/n$	$\frac{1}{n}$
$\theta$						
0.9	-0.485	0.006	3.676	-0.497	0.00524	0.002
0.6	-0.435	0.005	3.272	-0.441	0.00344	0.002
0.2	-0.186	0.003	0.695	-0.192	0.00216	0.002

**Table 6. Results of  $r_1$  for MA(1) models for series length 300**

No of realizations for each model = 1000						
Parameter	Empirical results			Asymptotic results		
	Mean ( $r_1$ )	Var ( $r_1$ )	skewness	Mean ( $r_1$ )	$(1 + 2\rho_1^2)/n$	$\frac{1}{n}$
$\theta$						
0.9	-0.486	0.006	2.966	-0.497	0.00873	0.0033
0.6	-0.434	0.005	2.468	-0.441	0.00573	0.0033
0.2	-0.187	0.004	0.242	-0.192	0.00360	0.0033

**Table 7. Results of  $r_1$  for MA(1) models for series length 100**

No of realizations for each model = 1000						
Parameter	Empirical results			Asymptotic results		
	Mean ( $r_1$ )	Var ( $r_1$ )	skewness	Mean ( $r_1$ )	$(1 + 2\rho_1^2)/n$	$\frac{1}{n}$
$\theta$						
0.9	-0.481	0.009	1.521	-0.497	0.0262	0.01
0.6	-0.427	0.009	1.103	-0.441	0.0172	0.01
0.2	-0.186	0.009	0.001	-0.192	0.0108	0.01

Table 8. Results of  $r_1$  for MA(1) model for series length 50

No of realizations for each model = 1000						
Parameter		Empirical results			Asymptotic results	
$\phi$	Mean ( $r_1$ )	Var ( $r_1$ )	skewness	Mean ( $r_1$ )	$(1 + 2\rho_1^2)/n$	$\frac{1}{n}$
0.9	-0.481	.010	0.255	-0.497	0.0524	0.02
0.6	-0.428	.011	0.216	-0.441	0.0344	0.02
0.2	-0.197	.017	0.074	-0.192	0.0216	0.02

Table 9. Results for ARMA(1,1) models for series length 500

No. of realizations for each model = 1000							
Parameter		Empirical results			Asymptotic results		
$\theta$	$\phi$	Mean ( $r_1$ )	Var ( $r_1$ )	skewness	Mean ( $r_1$ )	$(1 + 2\rho_1^2)/n$	$\frac{1}{n}$
0.9	0.2	-0.394	0.00102	0.120	-0.396	0.0026	0.002
0.2	0.9	0.832	0.00078	-0.510	0.844	0.0048	0.002

Table 10. Results for ARMA(1,1) models for series length 300

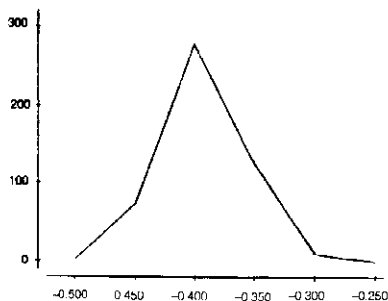
No. of realizations for each model = 1000							
Parameter		Empirical results			Asymptotic results		
$\theta$	$\phi$	Mean ( $r_1$ )	Var ( $r_1$ )	skewness	Mean ( $r_1$ )	$(1 + 2\rho_1^2)/n$	$\frac{1}{n}$
0.9	0.2	-0.396	0.0021	-0.052	-0.396	0.0044	0.0033
0.2	0.9	0.823	0.0017	-0.552	0.844	0.0081	0.0033

**Table 11. Results for ARMA(1,1) models for series length 100**

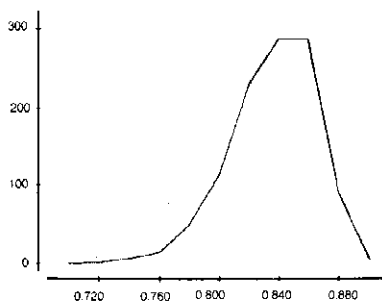
No. of realizations for each model = 1000							
Parameter		Empirical results			Asymptotic results		
$\theta$	$\phi$	Mean ( $r_1$ )	Var ( $r_1$ )	skewness	Mean ( $r_1$ )	$(1 + 2\rho_1^2)/n$	$\frac{1}{n}$
0.9	0.2	-0.391	0.0058	0.190	-0.396	0.013	0.01
0.2	0.9	0.780	0.0071	-0.903	0.844	0.024	0.01

**Table 12. Results for ARMA(1,1) models for series length 50**

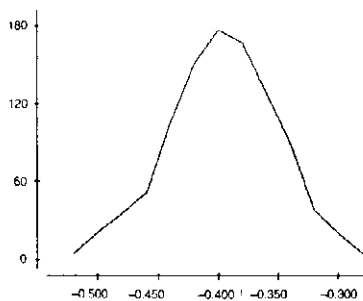
No. of realizations for each model = 1000							
Parameter		Empirical results			Asymptotic results		
$\theta$	$\phi$	Mean ( $r_1$ )	Var ( $r_1$ )	skewness	Mean ( $r_1$ )	$(1 + 2\rho_1^2)/n$	$\frac{1}{n}$
0.9	0.2	-0.383	0.013	0.446	-0.396	0.026	0.02
0.2	0.9	0.711	0.015	-0.866	0.844	0.048	0.02



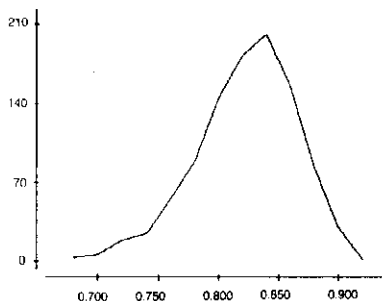
**Fig. 2a1.** Frequency curve for ARMA (1,1) of series length 500 with AR parameter 0.2 and MA parameter 0.9



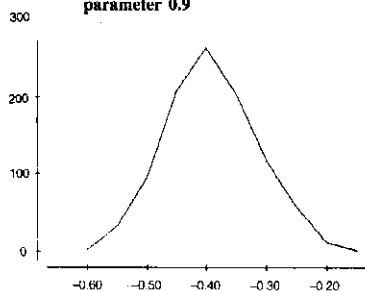
**Fig. 2a2.** Frequency curve for ARMA (1,1) of series length 500 with AR parameter 0.9 and MA parameter 0.2



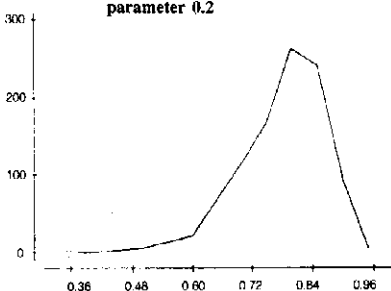
**Fig. 2a3** Frequency curve for ARMA (1,1) of series length 300 with AR parameter 0.2 and MA parameter 0.9



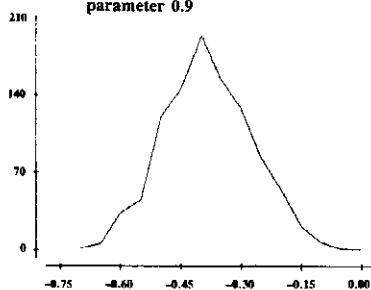
**Fig. 2a4** Frequency curve for ARMA (1,1) of series length 300 with AR parameter 0.9 and MA parameter 0.2



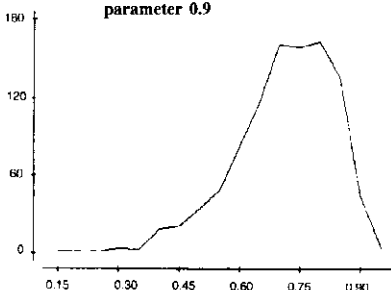
**Fig. 2a5** Frequency curve for ARMA (1,1) of series length 100 with AR parameter 0.2 and MA parameter 0.9



**Fig. 2a6** Frequency curve for ARMA (1,1) of series length 100 with AR parameter 0.2 and MA parameter 0.9



**Fig. 2a7** Frequency curve for ARMA (1,1) of series length 50 with AR parameter 0.2 and MA parameter 0.9



**Fig. 2a8** Frequency curve for ARMA (1,1) of series length 50 with AR parameter 0.9 and MA parameter 0.2

### Concluding Remarks

The study reveals some interesting results for the sample autocorrelation  $r_1$  in case of AR(1), MA(1) and ARMA(1,1) models. For AR(1) models the empirical means are smaller than the theoretical means for parameter value  $\phi = 0.9, 0.6$  and  $0.2$ .

For AR(1) models, Kendall and Stuart [11] give an approximate formula for computing mean of  $r_1$  as follows:

$$\text{mean}(r_1) = \phi - \frac{1 + 3\phi}{n - 1}$$

In our case the simulation results correspond closely with this approximation for parameter values  $\phi = 0.9, 0.6$  and  $0.2$ . The results are as follows:

Sample size	Parameter value	Mean ( $r_1$ )	Simulation result for Mean ( $r_1$ )
500	0.9	0.893	0.874
500	0.6	0.594	0.584
500	0.2	0.197	0.191
300	0.9	0.888	0.859
300	0.6	0.591	0.573
300	0.2	0.195	0.189
100	0.9	0.863	0.826
100	0.6	0.572	0.554
100	0.2	0.184	0.177
50	0.9	0.824	0.792
50	0.6	0.543	0.533
50	0.2	0.167	0.165

The simulated results are invariably smaller than Kendall and Stuart results.

In contrast to AR(1) models, the empirical means in MA(1) models are larger than theoretical means and the distributions of  $r_1$  for these models are skewed to the right. In general the empirical variances are larger than theoretical variances for series length 500 and 300 but in most cases the opposite is true for series length 100 and 50.

For ARMA (1,1) models, empirical variances are invariably smaller than theoretical variances. The empirical means are bigger than theoretical means for all cases for parameter values  $\theta = 0.9$  and  $\phi = 0.2$ . The opposite is true for parameter values  $\theta = 0.2$  and  $\phi = 0.9$ . Figures 2a1-2a8 show that for ARMA (1,1) with bigger AR parameter the frequency curves are skewed to the left and the opposite is true with bigger MA parameter. We also observe that distributions of  $r_1$  for ARMA (1,1) models bear a closer proximity to normal distributions than the other models under consideration.

The overall picture which emerge from this study conveys the message that the theoretical assumptions for autocorrelation  $r_1$  are valid and the accuracy of the results increases with increasing sample size. Also similar studies for higher order ARMA models and for higher order autocorrelations are plausible and could be suggested for the future.

**Acknowledgements.** We should like to thank Mr. M. Sarwar Khan for help in plotting frequency polygons.

### References

- [1] Box, G.E.P. and Pierce, D.A. "Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models." *Journal of American Statistical Association*, 65, No. 332 (1970), 1509-26.
- [2] Chatfield, C. *The Analysis of Time Series: An Introduction, 4th edition*. London: Chapman and Hall, 1988.
- [3] Batagalia, F. and Carlucci, F. "A Simulation Study of Causality Test." In: O.D. Anderson (Ed.), *Time Series Analysis Theory and Practice I*. Amsterdam: North Holland Publishing Company, 1982.
- [4] Bhansali, R.J. "A Simulation Study of Autoregressive and Window Estimation of Inverse Correlation Function." *Applied Statistics*, 32, No. 2 (1983), 141-149.
- [5] Batts, T. and McNow, F. "The Predictive Performance of Three Autoregressive Moving Average Models: A Monte Carlo Investigation." *Journal of Time Series Analysis*, 10, No. 4 (1988), 301-314.
- [6] Bartlett, M.S. "On Theoretical Specification of Sampling Properties of Autocorrelated Time Series." *Journal of Royal Statistical Society, Series B*, 8, No. 1 (1946), 27-41.
- [7] Priestley, Maurice B. *Spectral Analysis and Time Series*. London: Academic Press, 1981.
- [8] Wei, W.S. *Time Series Analysis Univariate and Multivariate Methods*. USA, California, Redwood City: Addison-Wesley, 1990.
- [9] Cryer, Jonathan D. *Time Series Analysis*. Boston: Duxbury Press, 1986.
- [10] Learmonth, G.P. and Lewis, P.A.W. *Statistical Tests of Some Widely Used and Recently Proposed Uniform Random Number Generators*. Report NPS55LW73111A, C.A. Monterey: Naval Postgraduate School, 1973.
- [11] Kendal, M.G.; Stuart, A. and Ord, J.K. *The Advanced Theory of Statistics*, Vol. 3, 4th ed. London: C. Griffin Company Ltd., 1983.

## بعض خواص الترابطات الذاتية للعينة من الدرجة الأولى لبعض نماذج ARMA: دراسة باستخدام المحاكاة

عدنان ماجد بري وإسرار حسين خان

قسم الإحصاء وبحوث العمليات، كلية العلوم، جامعة الملك سعود،

ص.ب. ٢٤٥٥، الرياض، المملكة العربية السعودية

(استلم في ١٤١٣/٩/١هـ؛ قبل للنشر في ١٤١٤/١١/١٣هـ)

ملخص البحث. يلعب الترابط الذاتي للعينة دور رئيسي في التعرف على نموذج المتسلسلة الزمنية. ودراسة خواص الترابط الذاتي للعينة حصل على كثير من الاهتمام في الأبحاث المنشورة لتحليل المتسلسلات الزمنية كما أن الفرضية الطبيعية لتوزيع الترابط الذاتي جعل المتوسط والتباين هو بؤرة الاهتمام في هذه الدراسات وقد استخدمت الكثير من الطرق لتقدير التباين ولكن هذه الطرق صحيحة فقط للعينات الكبيرة جداً وتقريب يضحي بالدقة المنشودة.

تقوم الدراسة الحالية باستكشاف التوزيع التجريبي للترابط الذاتي باستخدام تقنية المحاكاة حيث يمكننا اختبار موثوقية هذه النتائج التي تعتمد على العينات الكبيرة وذلك لعينات صغيرة. وقد استخدمنا الطرق البيانية لاختبار بعد هذه الفرضيات حول الفرضية الطبيعية لتوزيع دالة الترابط الذاتي للعينة. وقد لوحظ وتوون مقدار ابتعاد مختلف النماذج عن الفرضية الطبيعية.