ARIMA (1,0,1) Süreci İçin «ρ_1» in Küçük Örnek Tahminleri

Small Sample Estımates Of «ρ_1» For The ARIMA (1,0,1) Process

Zekâi ŞEN [1]

Imperial College of Science and Technology, London, ENGLAND

Hidrolojik zaman serilerinin en önemli özelliklerinden bir tanesi ardışık gözlemlerin birbiri ile serisel olarak bağıntılı olmasıdır. Bu özelliğe hidrolojide «ısrarlılık» adı verilir. Israrlılık biri birinci mertebeden serisel korelasyon katsayısı, ρ_1, *ile ölçülen kısa süreli ısrarlılık diğeri ise Hurst katsayısı, h, ile ölçülen uzun süreli ısrarlılık olmak üzere iki kısma ayrılır. Verilen bir geçmiş gözlemler dizisinden sadece birer adet* ρ_1 *ve h parametresi hesap edilebilir ki bunlar bir dereceye kadar taraflıdır.*

Bu makalenin esas amacı, yıllık akış serilerinin türetme mekanizmalarının ARIMA (1, 0, 1) modeli olduğu kabulü ile birinci mertebeden serisel korelasyon katsayısı ile ilgili taraflılık etkisinin analitik ifadesini çıkarmaktır.

One of the most important features of hydrological time series is that the successive observations are serially correlated with each other. This property has been termed as the persistence in hydrologic phenomenon. By a further consideration the persistence can be divided into two parts, one of which is the short term persistence which has been so far measured by the first order serial correlation coefficient, ρ_1, *and the other part is the long - term persistence the measure of which is*

---

1) Now at Civil Engineering Faculty, Dept. of Hydraulics and Water Power, Istanbul Technical University, Istanbul.

the Hurst coefficient, h. A given historic sequence yields estimates of both ρ_1 and h which are biased to some extent.

It is the main objective of this paper to derive an analytical expression for the bias amount associated with the serial correlation coefficient if the generating mechanism of hydrologic phenomenon is assumed to be the ARIMA (1, 0, 1) process

1 — INTRODUCTION

The hydrological phenomena evolve along the time axis and the measurements of the phenomena made at equal intervals of time constitute a hydrologic sequence which is referred to as a discrete time series. Some of the examples to such a series are rainfall, evaporation, streamflow sequences etc.

The measured part of an event constitutes historic sequence upon which a mathematical model is constructed which is employed in generating future likely projections of the same phenomenon. Hence, it will be possible in a statistical sense to predict the future events on the basis of which the hydrologist will then be able to assess the benefical and optimal use of the water rescurces systems. Thus, it is obvicus that when hydrologists are provided with a sequence of observations, one of the first steps to be taken is to identify a suitable mathematical model and then comes the estimation of the parameters of the model where the effect of bias becomes effective.

2 — A BRIEF REVIEW OF MATHEMATICAL MODELS IN HYDROLOGY

Up to now, there are various mathematical models each of which is proposed to preserve, in synthetic sequences some of the most meaningful characteristics of observations. In this paper attention will be confined only to annual streamflow sequences which are random in character and due to this randomness it is not possible to represent the event entirely by a deterministic model. The only way be control the random natured events, is to treat them in a statistical sense which requires the probability distribution function of the event.

Some of the aforementioned characteristics are related to this probability distribution function of the event. Among the parameters the mean shows the location of the assumed probability distribution function (pdf), the variance is the measure of dispersion of observations about the mean value. If the pdf of event considered is not symmetrical then it might be a good idea to try and preserve the coefficient of skewness in the proposed model, but the general idea from the parameter estimation point of view is that only low order moments must be employed in the model; otherwise, the estimations obtained from limited historic sequence, of a high order moment will be highly biased and unrepresentative of the underlying parameter. Furthermore, the estimated parameter with a great bias in its structure, will cause the proposed model to generate future sequences of unrealistic values.

However, all of the proposed models seem to preserve, in common, the three essential parameters, namely, the mean, μ, the standard deviation, σ, and the first order serial correlation coefficient, ρ_1. These are just the three parameters necessary to generate lag-one Markov sequences where the present value of observation is assumed to be dependent on its past and a random shock which is independent of the past. The general expression of this model is given by,

$$x_t - \mu = \rho_1 (x_{t-1} - \mu) + \sigma . (1 - \rho_1{}^2)^{1/2} \varepsilon_t \quad (1)$$

where ε_t is a normally and independently distributed random variable with zero mean and unit variance.

Recently, a new parameter which was first introduced into hydrology by Hurst (1951) in relation with his studies about the long term storage capacity of reservoirs, has given new insights into synthetic hydrology. This parameter referred to as the Hurst coefficient is a measure of the long - term persistance in a hydrological sequence. For natural streamflow sequences which are dependent, Feller (1951) has shown that h is asymptotically equal to 0.5. Moreover, the phenomena which yield h values greater than 0.5 is known among hydrologists as the Hurs phenomenon.

From the Hurst phenomenon point of view, the Markov processes are not capable of preserving h values greater than 0.5. Due to this defect of the Markov processes hydrologists began to search new models that would be adequate to Preserve h as well as μ, σ and ρ_1 simultaneously. It appears that the first set of models proposed for this purpose were the discrete fractional Gaussion noises (dfon) which were presented into hydrology by Mandelbrot and Wallis (1969). Although, these models are capable of preserving any h value between 0 and 1, their use requires tremendous amount of both computer time and memory. Later, a simple model referred to as the ARIMA (1, 0, 1) process was adopted as a potential model for simulating hydrological events by O'-Connel (1971). The advantages of the ARIMA (1, 0, 1) process over the Markovian models are that, for high values of ρ_1 small values of h can be preserved in synthetic sequences and the computer time needed for the generation of an ARIMA (1, 0, 1) sequence is drastically less than the time required for dfGn. The white Markov Process proposed into hydrology by Şen (1974) has the same advantage as that of the ARIMA (1, 0, 1) process.

Finally, the Broken Line process proposed by Mejia (1971) as a potential model, is claimed to preserve the crossing properties of historoc sequence and it is also proven that in its continuous form the preservation of h is possible.

3 — THE SOURCES OF BİAS

After the identification of the model for a specific purpose such as the generation of future streamflows, the parameters of the model have to be estimated on the basis of historic sequence, available. One of the essential properties required from an estimator is that it must yield an unbiased estimate in the long run. Although, both the maximum likelihood and method of moments estimates yield unbiased estimates in the long run, when the length of sequence is short all of the proposed estimators will give biased values of the parameter estimated.

In this study, it is clear that the magnitude of bias will be depending on the nature of the estimator. However, the estimators given for the mean and variance are the same whether it is the maximum likelihood or method of moments estimator. In the estimation of the first order serial correlation coefficient there exists a distinction between the maximum likelihood and the method of moments estimaters. The estimators of the mean and the variance are

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \qquad (2)$$

and

$$S^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (x_i - \bar{x})^2 \qquad (3)$$

respectively. On the other hand, the moment and the makimum likelihood estimators of the first orderserial correlation coefficient are given as

$$r_k = \frac{\sum_{i=1}^{n-k} x_i \, . \, x_{i+k} - \frac{1}{(n-k)} \left(\sum_{i=1}^{n-k} x_i \right) \left(\sum_{i=1}^{n-k} x_{i+k} \right)}{\left[\sum_{i=1}^{n-k} x_i^2 - \frac{1}{(n-k)} \left(\sum_{i=1}^{n-k} x_i \right)^2 \right]^{0.5} \left[\sum_{i=1}^{n-k} x^2_{i+k} - \frac{1}{(n-k)} \left(\sum_{i=1}^{n-k} x_{i+k} \right)^2 \right]^{0.5}} \tag{4}$$

and

$$r_k = \frac{\sum_{i=1}^{n-k} \left(x_i - \frac{1}{n} \sum_{i=1}^{n} x_i \right) \left(x_{i+k} - \frac{1}{n} \sum_{i=1}^{n} x_i \right)}{\sum_{i=1}^{n} \left(x_i - \frac{1}{n} \sum_{i=1}^{n} x_i \right)^2} \tag{5}$$

respectively. There are also various other definitions of the serial correlation coefficient given in the statistics literature and the most important ones are reviewed by Wallis and O'Connel (1972). The circular definition of the serial correlation coefficient proposed by Kendall (1954) yields the same bias correstion factors as the maximum likelihood estimator presented by Jenkins and Watts (1968).

Another source of bias comes from the estimation of the mean value of a given sequence. If the population mean value is known then the bias effect will be less than the case where the mean is estimated from a given historic sequence. As it is stated by Marriot and Pope (1954) this kind of bias which is present even when the autocorrelation coefficient is equal to zero, and it is always negative in a long series i.e. the expectation of the serial correlation coefficient calculated using the estimated mean is always algabrically negative and near to 1.

A further source of bias depends on the generation model itself. For the same set of parameters and the estimator the magnitude of bias is a function of the model employed.

As a summary, the magnutide of bias is dependent on there major factors namely, the estimator that is used to estimate the parameter concerned, the type of process such as the Markov process, ARIMA (1, 0, 1) process, the white Markov process and finally whether the mean value is known or not.

One of the most frequently used estimators of the serial correlation coefficient has been proposed by Kendall (1954) which is referred to as the circular definition in which the $(n+1)$ - th observation of the series is assumed to be equal to the i - th observation, the estimator is expressed as,

$$r_k = \frac{\sum_{i=1}^{n}\left(x_i - \frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(x_{i+k} - \frac{1}{n}\sum_{i=1}^{n} x_i\right)}{\sum_{i=1}^{n}\left(x_i - \frac{1}{n}\sum_{i=1}^{n} x_i\right)^2} \tag{6}$$

4 — BIAS CORRECTION ALGORITHMS

So far, in the literature of time series analysis, to the best of author's knowledge, there is not an exact algorithm proposed, to be valid for bias correction of r_k.

However, there are few approximation algorithms currently used in hydrology. One of the earliest of such an algorithm has been presented by Quenouille (1957) but it is a rough method of removing bias in the first order serial correlation coefficient given by

$$\hat{\rho}_1^q = 2\hat{\rho}_1 - \frac{1}{2}\left(\hat{\rho}_{1,1} + \hat{\rho}_{1,2}\right) \tag{7}$$

where $\hat{\rho}_1$ is the estimate of the first order serial correlation coefficient of the time series, considered in full length, then, the whole time series is divided into two halves each having their individual first order correlation coefficients as $\rho_{1,1}$ and $\hat{\rho}_{1,2}$. $\hat{\rho}_1^q$ in Eq. 7 is supposed to be corrected value of the correlation coefficient. As it is stated by Wallis and O'Connel (1972) on average the variance of $\hat{\rho}_1^q$ is greater than the variance of unbiased $\hat{\rho}_1$ estimated by other procedures, consequently, it was concluded that this algorithm cannot be universally recommended, but if many short realizations of the same process are available then the mean of $\hat{\rho}_1^q$ may be satisfactory.

Another very successful algorthm is given by Kendall (1954) in the discussion part of the paper by Marriott and Pope (1954). He assumed the estimates of the lag - one correlation coefficient to be normally distributed and derived the necessary analytical expressions to order n^{-1}. This latter way of finding bias correction factors is adopted in this paper. The review of Kendall's bias correction has been given in Appendix A and the expected value of r_1 is rederived as,

$$E(r_1) = \frac{E(A)}{E(B)} - \frac{\mathrm{Cov}(A,B)}{E^2(B)} + \frac{E(A)\,\mathrm{Var}(B)}{E^3(B)} \tag{8}$$

5 — BIAS CORRECTIONS FOR VARIOUS DEFINITIONS OF «ρ_1»

So far, in this paper only the sources of bias and the methods of finding unbiased estimates have been mentioned. In this section, the expressions for the actual bias correction of the ARIMA (1, 0, 1) process will be given.

First of all the ordinary definition is taken into consideration and then the small sample expectation of r_1 has been shown in Appendix B to be

$$E(r_1) = \rho_1 - \frac{1}{\nu}\left[\frac{(1-\phi-2\rho_1)(1-\phi-2\rho_1^2)}{(1-\phi^2)} + \rho_1\right] \tag{9}$$

This last expression is in accordance with the statement made by O'Connell (1973) as 'the bias in ρ_1 observed for small samples of an ARIMA (1, 0, 1) process are functions of the sample size ,n, and the driving parameters $\varnothing$ and θ of the process'.

If the circular or the maximum likelihood definition of the correlation coefficient is adopted the resulting expression for the bias correction is shown in Appendix C,

$$E(r_1) = \rho_1 - \frac{1}{n}\left[\frac{(1-\phi-2\rho_1)}{(1-\phi)}(1-\rho_1) + 3\rho_1 + \frac{4\rho_1^2}{(1-\phi^2)}(\phi-\rho_1)\right] \tag{10}$$

Although bias corrections can be worked out by Eq. 9 or Eq. 10, they do not lend themselves to a simple analytical form for the unbiased population ρ_1 as in the case of the lag - one Markov process. Of course, a cubic equation in ρ_1 form Eq. 9 and another cubic equation from Eq.10 can be written straightforwardly; but to find a solution to these expressions can be difficult and tedious.

One way of finding ϕ and θ values to be employed in the generating process in order to obtain an unbiased value of ρ_1, is to prepare tables which include the number of observations n and the parameters ρ_1, ϕ and θ. One sample of such a table is given in Table 1. The values in Table 1 are obtained by using the maximum likelihood definition of ρ_1. It is obvious that the purpose of tables is to solve Eq. 9 and Eq.10 numerically. Interpolation is required if necessary. To give an example of how to use the tables just an arbitrary set of parameters is picked out as $n = 40$, $\rho_1 = 0.218$, $\phi = 0.85$ and θ automatically turns out to be 0.68. Now, if the given set is employed in the generating process without any bias correction, in the long - term the first order serial correlation coefficient will turn out to be 0.155 which can easily be taken from the Table 1. Now that, it is required the process to yield an unbiased value of ρ_1, some bias correction must be applied to ρ_1 by means of adjusting θ. As a result θ is a control parameter for the correction of bias. The procedure works as follows. The appropriate column with n is found, then, downwards in this column, the value of ρ_1 is sought. Once ρ_1 is spotted then moving horizontally to the left the corresponding values of ρ_1^* and θ^* are found. With these new parameters n, ϕ, ρ_1^* and θ^* ARIMA (1, 0, 1) sequences can be generated so that in the long run an unbiased vaue of ρ_1 will emerge.

To apply the aforementioned procedure, it is necessary to have tables comprising all the values of ϕ and θ. Obviously, if the parameters are corrected for bias then their variances do not remain the same; they might increase. However, the variance of the above found new estimate of ρ_1^* cannot be analytically solved; in such a case the Monte Carlo techniques must be employed. The comparision of the bias corrections of the ARIMA (1, 0, 1) process with that of tha Markov process, on the grounds that both have the same first order serial correlation value will prove that the former model's bias is greater than the latter one.

Another important point is obvious that, as has already been presented by Wallis and O'Connell (1972) for the Markov process there are upper and lower constrains for the bias correction procedure. Outside of these limit values the bias correction procedure is not valid. From hydrological point of view such a situation will hardly be encountered for the streamflow sequences (annual) for which the first order serial correlation coefficient is around 0.2.

As far as the validity of the two equations namely Eq. 9 and Eq. 10 is concerned there is again some restriction which cannot be analytically proved. That is, these two equations will yield untrue bias corrections as ϕ approaches to unity. It has been experimentally found that for values of ϕ greater than 0.9 bias corrections will be violated and they must not be used in the generating process.

6 — APPENDIX A

This appendix presents the review of the bias correction algorithm proposed by Kendall (1954). To simplify the expression, the normal distribution is assumed to be valid for the estimate of the first order serial correlation coefficient.

All of the estimators used for the serial correlation coefficient can, in general, be written in the following form,

$$r_k = \frac{A}{(B.C)^{1/2}} \tag{A.1}$$

where A is the k — th order autocovariance function, B and C are variances of the two overlapping subsectors of the actual series. If it is the maximum likelihood estimator them B equals C. B and C will have their own probability distribution fuctions (pdf) for any given length of series i.c. they are all random variables (r. v.) and in turn have their own moment. Let the first order moments of A, B and C be represented by $E(A)$, $E(B)$ and $E(C)$, respectively. The terms on the right hand side of Eq. A. 1 can be rewritten as follows,

$$r_k = \frac{E(A)+a}{\{[E(B)+b][E(C)+c]\}^{0.5}} \tag{A.2}$$

where a, b and c are the deviations from their respective expectations and again the same sort of $r.v$'s are valid where the only difference is that the new set of $r.v$'s has expectation equal to zero.

However, the right hand side of Eq. A. 2 can be expanded into an infinite summation by applying the Binomial expansion formula to each one of the terms both in the nominator and in the denominator, so that the expectation of sides will be easy and straightforward. First of all Eq. A. 2 can be rewritten as,

$$r_k = [E(A)+a] \cdot [E(B)+b]^{-1/2} \cdot [E(C)+c]^{-1/2}$$

or,

$$r_k = \frac{E(A)}{E(B).E(C)} \left[1 + \frac{a}{E(A)} \right] \cdot \left[1 + \frac{b}{E(B)} \right]^{-1/2} \left[1 + \frac{c}{E(C)} \right]^{-1/2}$$

After the application of the Binomial expansion one can obtain,

$$r_k = 1 - \frac{1}{2} \cdot \frac{c}{E(C)} + \frac{3}{8} \cdot \frac{c^2}{E^2(C)} + \cdots - \frac{1}{2} \cdot \frac{b}{E(B)} + \frac{1}{4} \cdot \frac{b.c}{E(B).E(C)}$$

$$- \frac{3}{16} \cdot \frac{b.c^2}{E(B).E^2(C)} + \frac{3}{8} \cdot \frac{b^2}{E^2(B)} - \frac{3}{16} \cdot \frac{b^2.c}{E^2(B).E(C)} + \frac{9}{64} \cdot \frac{b^2.c^2}{E^2(B).E^2(C)} + \frac{a}{E(A)}$$

$$- \frac{1}{2} \cdot \frac{a.c}{E(A).E(C)} + \frac{3}{8} \cdot \frac{a.c^2}{E(A).E^2(C)} - \frac{1}{2} \cdot \frac{a.b}{E(A).E(B)} + \frac{1}{4} \cdot \frac{a.b.c}{E(A).E(B).E(C)}$$

$$- \frac{3}{16} \cdot \frac{a.b.c^2}{E(A).E(B).E^2(C)} + \frac{3}{8} \cdot \frac{a.b^2}{E(A).E^2(B)} - \frac{3}{16} \cdot \frac{a.b^2.c}{E(A).E^2(B).E(C)} +$$

$$+ \frac{9}{64} \cdot \frac{a.b^2.c^2}{E(A).E^2(B).E^2(C)} - \cdots\cdots \qquad \text{(A.3)}$$

A further simplification of the above expression is obtained by asuuming that $B = C$. In the case of the maximum likelihood estimator this equivalence is perfectly satisfied, but it is an approximation when the estimation is worked out by the method of moments. So,

$$E(B) = E(C) \; ; \quad E(b^2) = E(c^2) = E(b.c) = \text{Varb}$$

Thus, by taking the expectation of both sides of Eq. A. 3,

$$E(r_k) = \frac{E(A)}{E(B)} \left[1 + \frac{E(b^2)}{E^2(B)} - \frac{E(a.b)}{E(A).E(B)} \right]$$

or,

$$E(r_k) = \frac{E(A)}{E(B)} - \frac{\text{Cov}(a,b)}{E^2(B)} + \frac{E(A).\text{Varb}}{E^3(B)} \qquad \text{(A.4)}$$

This last expression gives the general bias correction algorithm for any kind of process and it was first applied to various processes such as the Markov,moving average and white noise processes.

7 — APPENDIX B

In this part of the appendix, bias correction formula for the ARIMA (1, 0, 1) process will be exposed. The definition used for the first order serial correlation coefficient is that of the one given in Eq. 4. The corresponding A, B and C values are,

$$A = \frac{1}{(n-k)} \sum_{t=1}^{n-k} x_t \, x_{t-k} - \frac{1}{(n-k)^2} \left(\sum_{t=1}^{n-k} x_t \right) \left(\sum_{i=1}^{n-k} x_{t+k} \right) \qquad (B.1)$$

$$B = \frac{1}{(n-k)} \sum_{t=1}^{n-k} x_t^2 - \frac{1}{(n-k)^2} \left(\sum_{t=1}^{n-k} x_t \right)^2 \qquad (B.2)$$

and

$$C = \frac{1}{(n-k)} \sum_{t=1} x^2_{t+k} - \frac{1}{(n-k)^2} \left(\sum_{t=1} x_{t+k} \right)^2 \qquad (B.3)$$

As is obvious from these last two expressions, it is a fair assumption to say that $B \approx C$, so there are two variables left to be treated which are A and B. Eq. A. 3 requires only the expected values and the covariance of these two variables which can be evaluated as follows. Let us put $n - k = \nu$ in the above expression then by taking the expectation of B,

$$E(B) = \frac{1}{\nu} . E\left(\sum_{t=1}^{\nu} x_t^2 \right) - \frac{1}{\nu^2} E \left(\sum_{t=1}^{\nu} x_t \right)^2$$

After a tedious and long algabric calculations $E\ (B)$ has been simplified by Şen (1974) as,

$$E(B) = \frac{1}{\nu^2} \left| \nu - 1 - \frac{2}{\nu} \sum_{j=1}^{\nu-1} (\nu - j) \rho_j \right| \qquad (B.4)$$

and

$$E(A) = \frac{1}{\nu} \left| \nu \rho_k - \rho_k - \frac{2}{\nu} \sum_{j=1}^{\nu-1} (\nu - j) . (\rho_{k-j} + \rho_{k-j}) \right| \qquad (B.5)$$

On the basis of these two last expressions the expected value of the serial correlation coefficient of any order can be calculated provided that the particular correlation structure of the process concerned is given. In this paper, the process considered is the ARIMA (1, 0, 1) process of which the correlation structure is given as,

$$\rho_1 = \frac{(1-\phi.\theta).(\phi-\theta)}{1+\theta^2-2.\phi\theta}$$

$$\rho_k = \phi\ \rho^{k-1} \quad \text{for } k \geq 2 \tag{B.6}$$

So if the necessary calculations are performed we get,

$$E(B) - \frac{1}{\nu}\left\{\nu - 1 - \frac{2}{\nu}\left[\nu\rho_1 \frac{1-\phi^{\nu-1}}{1-\phi} - \frac{\rho_1}{1-\phi}\left[\frac{1-\phi^{-1}}{1-\phi} - (\nu-1).0^{\nu-1}\right]\right]\right\}$$

when the higher terms than n^{-1} are ignored the expression shrinks to

$$E(B) = 1 - \frac{1}{\nu} - \frac{2}{\nu(1-\phi)} \tag{B.7}$$

The expected value of A for any kind of process is given by Şen (1974) as,

$$E(A) = \frac{1}{\nu}\left[\nu\rho_k - \frac{1}{\nu}\sum_{j=0}^{\nu-1}(\nu-j)\rho_{k+j} - \frac{1}{\nu}\sum_{j=1}^{k}(\nu-j)\rho_{k-j} - \frac{1}{\nu}\sum_{j=1}^{\nu-k-1}(\nu-k-j)\rho_j\right] \quad \text{for } k>0$$

For the ARIMA (1, 0, 1) process to the order of n^{-1} $E(A)$ becomes

$$E(A) = \rho_1 . \phi^{k-1} - \frac{2\rho_1}{\nu(1-\phi)} - \frac{1}{\nu} \tag{B.8}$$

The other two terms required for Eq. A. 4 are given by Kendall (1954) as,

$$\text{Cov}(A,B) = \frac{2}{\nu}\sum_{j=\infty}^{+\infty}\rho_j . \rho_{j+k}$$

In the case of the ARIMA (1, 0, 1) process this last expression turns out to be

$$\text{Cov}(A,B) = \frac{2}{\nu}\left(\rho_1\phi^{k-1} + k.\rho_1^2.\phi^{k-2} + 2.\ \rho_1^2\frac{\phi^k}{1-\phi^2}\right) \tag{B.9}$$

and finally,

$$\mathrm{Var}(B)=\frac{2}{\nu}\sum_{j=-\infty}^{+\infty}\rho^2_j$$

Transformation of this expression through the ARIMA (1, 0, 1) process

$$\mathrm{Var}(B)=\frac{2}{\nu}\left(2+\frac{2.\rho_1^2}{1-\phi^2}\right) \tag{B.10}$$

In this way all of the necessary terms to work out Eq. A. 4 are calculated the substitution of which into Eq. A. 4 yields,

$$B(r_k)=\rho_1\,\phi^{k-1}-\frac{1}{\nu}\left\{\frac{1}{(1-\phi^2)}\left[1-\phi+2.\rho_1).(1+\phi-2.\rho_1^2.\phi^{k-1})-\right.\right.$$
$$\left.\left.-(1-\phi^2)\rho_1\,\phi^{k-1}\right]+2.k.\,\rho_1^2.\phi^{k-2}\right\} \tag{B.11}$$

This is the general expression which gives the biased value of the k—th order autocorrelation ocefficient or, the amount of bias which is defined by $E(r_k)-\rho_k$ can be found as,

$$E(r_k)-\rho_1.\phi^{k-1}=-\frac{1}{\nu}\left\{\frac{1}{(1-\phi^2)}\left[(1-\phi+2.\rho_1).(1+\phi-2.\rho_1^2.\phi^{k-1})-\right.\right.$$
$$\left.\left.-(1-\phi^2)\rho_1\,.\,\phi^{k-1}\right]-2\,k.\,\rho^2.\phi^{k-2}\right\} \tag{B.12}$$

Hydrologists are interested in the first order serial correlation coefficient, hence by substituting $k-1$ in the above expressions, the bias values for the first order serial correlation coeffiicent becomes,

$$E(r_1)=\rho-\frac{1}{\nu}\left\{\frac{1}{(1-\phi^2)}\left[(1-\phi^2)+\rho\,(1+2.\phi-4.\rho)+\phi.\,\rho_1(2\,\rho_1-\phi)\right]+2\,\frac{\rho_1}{\phi}\right\} \tag{B.13}$$

and subsequently the amount of bias for r_1 is,

$$E(r_1)-\rho_1=-\frac{1}{\nu}\left\{\frac{1}{(1-\phi^2)}\,(1-\phi+2.\rho_1).(2+\phi-2.\rho_1^2)+\rho_1\right\} \tag{B.14}$$

In order to verify the correctness of the above derived expressions they can be checked against the corresponding formula that have already appeared in the literature for the Markov process. To obtain the Markov case it is necessary to substitute $\rho_1 = \phi$ in the aforementioned expressions. Thus,

$$E(r_1) = \rho_1 - \frac{1}{\nu}(3 + \rho_1) \tag{B.15}$$

where Eq. B. 15 is exactly equivalent to the one presented by Kendall (1954) for the Markov process case.

8 — APPENDIX C

The aim of this last appendix is to provide similar bias correction formula for the circular definition of the autocorrelation function given in Eq. 6. The only difference of Eq. 6 from that of Eq. 5 is the nominator which can be written as,

$$A = \sum_{t=1}^{n} x_t \,.\, x_{t+k} - \frac{1}{n^2}\left(\sum_{t=1}^{n} x_t\right)^2$$

After performing the necessary algabric calculations $E(A)$ becomes,

$$E(A) = \frac{1}{n}\,[(n-k)\,\rho_1 \,.\, \phi^{k-1} + k.\phi^{n-k-1}] - \frac{1}{n^2}\left\{ n + 2\left[n\,\rho_1 \frac{1-\phi^{n-1}}{1-\phi} - \frac{\rho_1}{1-\phi}\left[\frac{1-\phi^{n-k-1}}{1-\phi} - (n-k-1).\phi^{n-k-1}\right]\right]\right\}$$

By ignoring the higher order terms the expression can be shortened to,

$$E(A) = \frac{1}{n}\,[n-k)\,\rho\,\phi^{k-1} + k.\phi^{n-k-1}] - \frac{1}{n} - \frac{2\rho_1}{n(1-\phi)} \tag{C.1}$$

Hence, again by substituting the four terms, namely, $E(A)$, $E(B)$, Var (B) and Cov (A, B) in Eq. A. 4 the following expressions are obtained. In general,

$$E(r_k) = \rho.\phi^{k-1} - \frac{1}{n}\left\{\frac{(1-\phi+2.\rho_1)}{1-\phi}(1-\rho_1 \,.\, \phi^{k-1}) + k\,\rho_1(\phi + 2.\rho_1).\phi^{k-2} - k.\rho_1 \,.\, \phi^{n-k-1} + \frac{4.\rho_1^2.\phi^{k-1}}{1-\phi^2}(\phi - \rho_1)\right\} \tag{C.2}$$

When $k = 1$ is inserted in the above equation then the small sample expectation of r_1 is obtained as,

$$E(r_1)=\rho_1-\frac{1}{n}\frac{(1-\phi+2\rho_1)}{(1-\phi)}(1-\rho_1)+3.\rho_1-\frac{4.\rho_1^2}{(1-\phi^2}(\phi-\rho_1) \qquad (C.3)$$

Table 1 — Variation of Parameters, ϕ=0.85.

P_1	θ	n — 20	n — 30	n — 40	n — 50	n — 100
0.203	0.700	0.005	0.071	0.104	0.124	0.163
0.218	0.680	0.011	0.080	0.115	0.135	0.177
0.250	0.670	0.023	0.099	0.137	0.159	0.205
0.298	0.640	0.043	0.128	0.171	0.196	0.247
0.329	0.620	0.058	0.139	0.198	0.221	0.275
0.360	0.600	0.075	0.170	0.218	0.246	0.303
0.406	0.570	0.102	0.203	0.254	0.284	0.345
0.503	0.500	0.184	0.295	0.350	0.384	0.450
0.619	0.400	0.286	0.397	0.452	0.486	0.552
0.706	0.300	0.396	0.499	0.551	0.582	0.644

REFERENCES

(1) Hurst, H.E., (1951), Long-Term Storage Capacity of Reservoirs, Trans. Am. Soc. Civ. Engrs., 116, 770-808.

(2) Jenkins, G.M., Watts, D.G., (1968), Spectral Analysis and Its Applications, San Francisco, Holden-Day, pp. 525.

(3) Kendall, M.G., (1954), Note on the Bias in the Estimation of Autocorrelation. Biometrika, 42, 403-404.

(4) Mandelbrot, B.B., Wallis, J.R., (1969), Some Long-Run properties of Geophysical Records, Water Resour. Res., 5(2), 321-340.

(5) Marriott, F.H.C., Pope, J.A., (1954), Bias in the Estimation of Autocorrelations, Biometrika, 42, 390-402.

(6) O'Connell, P.E., (1973), Choice of Generating Mechanism in Synthetic Hydrology with Inadequate Data, Int. Assoc. Hydrol. Sci., Madrid Symposium, June, 1973.

(7) Ouenouille, M.H., (1956), Notes in Bias in Estimation, Biometrika, 43, 353-360.

(9) Wallis, J.R., O'Connell, P.E., (1973), The Small Sample Estimation of, Water Resour. Res., 8, 707.

(8) Şen, Z, (1974), Small Sample Properties of Stationary Stochastic Processes and the Hurst Phenomenon in Hydrology, Ph. D. Thesis, London, Imperial College of Science and Technology, pp. 284.