Proceedings of the 1987 Winter Simulation Conference
A. Thesen, H. Grant, W. David Kelton (eds.)

SIMULATION OF STATIONARY TIME SERIES

David A. Burn
IMSL, Inc.
2500 CityWest Boulevard
Houston, TX 77042-3020, U.S.A.

ABSTRACT

The study of parameter estimation, prediction, and hypothesis test methodology in time series analysis frequently involves simulation of a stationary time series from an autoregressive moving average (ARMA) model with nonnormal random shocks. The simulation question is how to generate the dependent and possibly nonnormal initial values of the time series. Usually, the simulation must be warmed up in order to diminish the bias induced by approximate methods used to generate the initial values of the time series. This paper examines current concepts and algorithms for simulating a stationary time series from an ARMA model with either normal or nonnormal random shocks.

1. INTRODUCTION

Nonnormal time series play a fundamental role in the examination of robustness properties of parameter estimates, forecasts, and test statistics. One model of interest is the innovations outlier model which consists of a perfectly observed ARMA process whose random shocks follow a heavy-tailed and possibly nonnormal distribution. An innovations outlier affects both current and future observations of the time series, and is consistent with the forecast system. A second model of interest is the additive outlier model which consists of an imperfectly observed ARMA process whose random shocks follow a normal distribution. An additive outlier affects only the current observation, and is inconsistent with the forecast system. Failure to distinguish an innovations outlier from an additive outlier may result in inefficient parameter estimates, imprecise forecasts, and improper diagnosis of model adequacy (Martin and Zeh 1977). Hence, the simulation of stationary time series from an ARMA model with possibly nonnormal random shocks is essential to an empirical study of the robustness of the forecast system.

1.1. Models of Time Series

Consider the ARMA(p,q) model defined by

$$\phi(B)W_t = \theta_0 + \theta(B)A_t$$

where

$$\begin{aligned}\phi(B) &= 1-\phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p, \qquad p \geq 0\\ \theta(B) &= 1-\theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q, \qquad q \geq 0\end{aligned}$$

and B is the backward shift operator defined by $B^k W_t = W_{t-k}$, for all k. The standard model assumptions are

- The random shocks A_t are independent and identically distributed random variables with mean $\mu_A = 0$ and variance $\sigma_A^2 > 0$.
- The autoregressive operator $\phi(B)$ is stationary. Equivalently, the roots of the equation $\phi(B) = 0$ lie outside the unit circle.
- The moving average operator $\theta(B)$ is invertible. Equivalently, the roots of the equation $\theta(B) = 0$ lie outside the unit circle.

The model includes an overall constant θ_0 to allow for a nonzero series mean μ. Refer to Box and Jenkins (1976, pages 91–93) for further discussion.

The random shock form of the ARMA(p,q) model is given by

$$W_t = \mu + \psi(B)A_t$$

where

$$\begin{aligned}\psi(B) &= \phi^{-1}(B)\theta(B)\\ &= 1+\psi_1 B + \psi_2 B^2 + \cdots\end{aligned}$$

The random shock model is particularly useful since the correlated time series observations W_t are defined in terms of the uncorrelated random shocks A_t. The ψ weights of the infinite order moving average may be determined by equating coefficients of B in $\phi(B)\psi(B) = \theta(B)$ (Box and Jenkins 1976, pages 95–96). The sequence $\{\psi_k\}_{k=0}^{\infty}$ is absolutely convergent, and the series $\sum_{k=0}^{\infty}\psi_k$ is absolutely summable. Note that the ARMA(p,q) model and its random shock form may be equivalently expressed as

$$\phi(B)\tilde{W}_t = \theta(B)A_t$$

and

$$\tilde{W}_t = \psi(B)A_t$$

respectively, where $\tilde{W}_t = W_t - \mu$ corresponds to a time series with zero mean. Without loss of generality, I assume $\mu = 0$ (hence, $\theta_0 = 0$) in the sequel.

1.3. Moments of Time Series

Departures from normality are often characterized in terms of skewness (symmetry of distribution) and kurtosis (peakedness of distribution). In particular, the normal distribution has skewness $\sqrt{\beta_1} = 0$ and kurtosis $\beta_2 = 3$. Values of these moments for a variety of *univariate* distributions are given in Patel, Kapadia, and Owen (1976).

The mean, standard deviation, skewness, and kurtosis of the marginal distribution of the ARMA(p,q) process may be expressed in terms of the corresponding moments of the random shock process. The relationship between the central moments of the ARMA(p,q) process and the central moments of the random shock process is given by

$$\begin{aligned}
\mu_1(W) &= \mu_1(A)\sum_{j=0}^{\infty}\psi_j \\
\mu_2(W) &= \mu_2(A)\sum_{j=0}^{\infty}\psi_j^2 \\
\mu_3(W) &= \mu_3(A)\sum_{j=0}^{\infty}\psi_j^3 \\
\mu_4(W) &= \mu_4(A)\sum_{j=0}^{\infty}\psi_j^4 + 3[\mu_2(A)]^2\sum_{j\neq k}\psi_j^2\psi_k^2 \\
&= \left\{\mu_4(A) - 3[\mu_2(A)]^2\right\}\sum_{j=0}^{\infty}\psi_j^4 + 3[\mu_2(A)]^2\left(\sum_{j=0}^{\infty}\psi_j^2\right)^2
\end{aligned}$$

where $\mu_i(W)$ and $\mu_i(A)$ denote the i-th central moment of W_t and A_t respectively, for $i = 1,2,3,4$. These moments are easily determined based on the random shock form of the ARMA(p,q) model. The mean, standard deviation, skewness, and kurtosis of the distribution of W_t are defined by

$$\begin{aligned}
\mu &= \mu_1(W) \\
\sigma &= [\mu_2(W)]^{1/2} \\
\sqrt{\beta_1} &= \mu_3(W)/[\mu_2(W)]^{3/2} \\
\beta_2 &= \mu_4(W)/[\mu_2(W)]^2.
\end{aligned}$$

For independent and identically distributed (0, σ_A^2) random shocks,

$$\begin{aligned}
\mu &= 0 \\
\sigma &= \sigma_A\left(\sum_{j=0}^{\infty}\psi_j^2\right)^{1/2} \\
\sqrt{\beta_1} &= \sqrt{\beta_1(A)}\frac{\sum_{j=0}^{\infty}\psi_j^3}{\left(\sum_{j=0}^{\infty}\psi_j^2\right)^{3/2}} \\
\beta_2 &= [\beta_2(A) - 3]\frac{\sum_{j=0}^{\infty}\psi_j^4}{\left(\sum_{j=0}^{\infty}\psi_j^2\right)^2} + 3.
\end{aligned}$$

The stationarity and invertibility assumptions imply that the infinite sums of the ψ_j are absolutely convergent. Hence, the above moments exist.

2. OBJECTIVES

My primary objective is to simulate a *stationary* time series W_t of length n according to a specified ARMA(p,q) model. My secondary objective is to generate p initial observations of a stationary time series W_t defined by a specified ARMA(p,q) model.

2.1. Simulation of Time Series

The following algorithms are equivalent—selection of either is a matter of perspective.

SA 1: Simulation Algorithm 1

1. Generate p initial ARMA(p,q) observations W_t for $t = 1-p, 2-p, \ldots, 0$.
2. Generate $n+q$ random shocks A_t for $t = 1-q, 2-q, \ldots, n$.
3. Generate n ARMA(p,q) observations W_t for $t = 1,2,\ldots,n$ using the initial ARMA(p,q) observations from Step 1 and the random shocks from Step 2.

SA 2: Simulation Algorithm 2

1. Generate p initial AR(p) observations Z_t for $t = 1-p-q, 2-p-q, \ldots, 0-q$.
2. Generate $n+q$ random shocks A_t for $t = 1-q, 2-q, \ldots, n$.
3. Generate $n+q$ AR(p) observations Z_t for $t = 1-q, 2-q, \ldots, n$ using the initial AR(p) observations from Step 1 and the random shocks from Step 2.
4. Generate n ARMA(p,q) observations W_t for $t = 1,2,\ldots,n$ by applying the MA(q) operator to the AR(p) observations from Step 3.

By definition, SA 1 generates observations of an ARMA(p,q) process. To verify that SA 2 generates observations of an ARMA(p,q) process, consider the AR(p) model

$$\phi(B)Z_t = A_t$$

and the MA(q) model

$$W_t = \theta(B)Z_t.$$

Then

$$\phi(B)W_t = \phi(B)\theta(B)Z_t = \theta(B)A_t$$

which implies that the W_t process is defined by an ARMA(p,q) model.

2.2. Initialization of Time Series

Since the random shocks are assumed to be independent and identically distributed, all random shocks may be easily generated from the specified distribution of A_t. However, the observations of the time series are dependent and may be non-normal, and the joint distribution of the time series is usually unknown. Hence, the generation of the initial observations is the most difficult step in the simulation of stationary time series.

Furthermore, if an approximate method is used to generate the initial observations of the time series, then these observations represent transients which may disturb the stochastic equilibrium (stationarity) of the system. The *induction period* is defined as the length of time M required to minimize the transient bias induced by the initial observations of the time series (Anderson 1975). The total number of observations of the time series to be generated is $M + n$ with the first M observations used to warm up the simulation. Often the induction period M is chosen arbitrarily, and the simulation may be warmed up longer than necessary.

Random Shock Method. Generate initial observations of the time series from an approximation of the random shock form of the ARMA model.

Linear Prediction Method. Generate initial observations of the time series from the linear prediction equations based on the ARMA model.

In particular, any method of generating the initial observations of the time series should satisfy the following criteria:

1. The *joint distribution* of the p initial observations should correspond to the joint distribution of p consecutive observations of the time series.
2. The *induction period* should be of minimal length.

Examination of the Random Shock Method, Linear Prediction Method, and my proposed Adjusted Linear Prediction Method is performed in light of these criteria.

3. METHODOLOGY

The description of the methodology consists of theory, algorithm, and implications. Implementation of the algorithm is not accounted for in the implications.

3.1 Random Shock Method

Suppose the MA(m) model

$$W_t = \sum_{k=0}^{m} \psi_k A_{t-k}$$

provides a satisfactory approximation of the random shock form of the ARMA(p, q) model for sufficiently large m. Then the following algorithm may be used to perform Step 1 of SA 1.

IA 1: Initialization Algorithm 1

1. Generate $m + p$ random shocks A_t for $t = 1 - m, 2 - m, \ldots, p$.
2. Generate p series values W_t for $t = 1, 2, \ldots, p$ using the MA(m) model and the random shocks from Step 1.

For *normal* and *nonnormal* random shocks

- The joint distribution of $W_1, \ldots, W_p$ is an approximation to the joint distribution of p consecutive observations of the time series.
- An induction period may be required.

Anderson (1979) proposed a 'precise' method for determining the optimal induction period M for the Random Shock Method. The induction period M depends upon

- n, the length of the simulated series
- $\phi_1, \ldots, \phi_p$, the AR parameters

However, the form of this dependence is quite complicated for ARMA models with $p > 1$.

3.2 Linear Prediction Method

Let W_t be defined by the AR(p) model

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \cdots + \phi_p W_{t-p} + A_t$$

with theoretical autocovariance function

$$\gamma_k = E[W_t W_{t-k}].$$

The recursive relationship of Durbin (1960) is rewritten by Wilson (1978) as

$$\phi_{k+1,k+1} = (\gamma_{k+1} - \phi_{k,1}\gamma_k - \cdots - \phi_{k,k}\gamma_1)/\sigma_k^2$$

$$\phi_{k+1,j} = \phi_{k,j} - \phi_{k+1,k+1}\phi_{k,k+1-j}, \qquad j = 1, \ldots, k$$

$$\sigma_{k+1}^2 = \sigma_k^2(1 - \phi_{k+1,k+1}^2)$$

where $\sigma_0^2 = \gamma_0$.

Wilson (1978) states that initial observations of W_t may be generated by the sequence of AR(k) models for $k = 0, 1, \ldots, p - 1$

$$W_{k+1} = \phi_{k,1} W_k + \cdots + \phi_{k,k} W_1 + A_{k+1}$$

where

- $A_{k+1} = \sigma_k \epsilon_{k+1}$
- ϵ_{k+1} are iid standard normal variates

The following algorithm may be used to perform Step 1 of SA 2.

IA 2: Initialization Algorithm 2

1. Generate p iid $(0, 1)$ random shocks ϵ_t for $t = 1, 2, \ldots, p$.
2. Generate p initial observations W_t for $t = 1, 2, \ldots, p$ using the sequence of AR(k) models for $k = 0, \ldots, p - 1$ and the random shocks from Step 1.

For *normal* random shocks

- The joint distribution of $W_1, \ldots, W_p$ is equivalent to the joint distribution of p consecutive observations of the time series.
- An induction period is not required.

For *nonnormal* random shocks

- The second-order moments of the joint distribution of $W_1, \ldots, W_p$ are equivalent to the second-order moments of the joint distribution of p consecutive observations of the time series.
- An induction period may be required.

3.3 Adjusted Linear Prediction Method

Define

$$\phi_k(B) = \begin{cases} 1 & k = 0 \\ 1 - \phi_{k,1}B - \cdots - \phi_{k,k}B^k & k = 1, \ldots, p-1 \end{cases}$$

and

$$\begin{aligned} \psi_k(B) &= \phi_k^{-1}(B) \\ &= \begin{cases} 1 & k = 0 \\ 1 + \psi_{k,1}B + \psi_{k,2}B^2 + \cdots & k = 1, \ldots, p-1 \end{cases} \end{aligned}$$

Note that

- Stationarity of $\phi(B)$ ensures stationarity of $\phi_k(B)$ for $k = 1, \ldots, p-1$
- $\{\psi_{k,j}\}_{j=0}^{\infty}$ is absolutely convergent
- $\sum_{j=0}^{\infty} \psi_{k,j}$ is absolutely summable

Consider the AR(k) model

$$\phi_k(B)W_{k+1} = A_{k+1}$$

and its corresponding random shock form

$$W_{k+1} = \psi_k(B)A_{k+1}$$

for $k = 0, 1, \ldots, p-1$. Stationarity of W_t implies the following relationship between the moments of the time series and the moments of the random shock process for each linear prediction model:

$$\begin{aligned} \sigma(A_{k+1}) &= \sigma(W)\left(\sum_{j=0}^{\infty} \psi_{k,j}^2\right)^{-1/2} \\ \sqrt{\beta_1(A_{k+1})} &= \sqrt{\beta_1(W)}\frac{\left(\sum_{j=0}^{\infty} \psi_{k,j}^2\right)^{3/2}}{\sum_{j=0}^{\infty} \psi_{k,j}^3} \\ \beta_2(A_{k+1}) &= [\beta_2(W) - 3]\frac{\left(\sum_{j=0}^{\infty} \psi_{k,j}^2\right)^2}{\sum_{j=0}^{\infty} \psi_{k,j}^4} + 3 \end{aligned}$$

The following algorithm may be used to perform Step 1 of SA 2.

IA 3: Initialization Algorithm 3

1. Generate p **independent** $(0,1)$ random shocks A_t for $t = 1, 2, \ldots, p$ from successive Johnson curves based on skewness and kurtosis of A_t.
2. Generate p initial observations W_t for $t = 1, 2, \ldots, p$ using the sequence of AR(k) models for $k = 0, \ldots, p-1$ and the random shocks from Step 1.

For *normal* and *nonnormal* random shocks, this method is equivalent to the Linear Prediction Method. In addition, for *nonnormal* random shocks the Adjusted Linear Prediction Method ensures that the third- and fourth-order moments of the marginal distributions are equivalent.

4. EXAMPLE

Consider the stationary time series W_t defined by the AR(2) model

$$W_t = 0.5W_{t-1} + 0.3W_{t-2} + A_t$$

where the A_t are iid (0,1) random variables with skewness $\sqrt{\beta_1(A)} = 1$ and kurtosis $\beta_2(A) = 9$. A simulation experiment was conducted to examine the third- and fourth-order moments of the two initial observations of the time series. Each of the three initialization algorithms was used to generate NREP $= 10^4$ pairs of initial observations (W_1, W_2). Summary statistics examining the marginal (univariate) and joint (bivariate) distribution of the replicates were computed. This two-stage procedure was *super*-replicated NSUPER=16 times to allow assessment of the variability of the summary statistics. The random shocks were generated using the Johnson curve algorithm of Hill, Hill, and Holder (1985).

The graphical analysis of the simulation output is illustrated in part by Figure 1 and Figure 2. The boxplots in these displays describe the distribution of the 16 super-replicates of the bivariate skewness and bivariate kurtosis, respectively. (See Mardia 1970 for definitions of these measures.) Each boxplot is notched to illustrate an 84% confidence interval for the median. This allows pairwise comparison of boxplots to be performed at a family wide confidence level of 95% (Hettmansperger 1984). The dotted line in each figure represents the theoretical value of the bivariate skewness (0.9653) and bivariate kurtosis (14.7289).

In general, the Random Shock Method and Adjusted Linear Prediction Method appear to describe the bivariate distribution of the initial observations rather well. The order m of the moving average approximation of the Random Shock Method was equivalent to the order of the ψ polynomial used to compute the moments of the Adjusted Linear Prediction Method. In this case, it appears that neither of these methods requires an induction period. For the Linear Prediction Method, the confidence interval for the bivariate skewness fails to include the true value, and the confidence interval for the bivariate kurtosis barely contains the true value. The distribution of these sample measures is more variable and more skewed for the Linear Prediction Method than either of the other methods. However, based on these data, there is no pairwise differ-

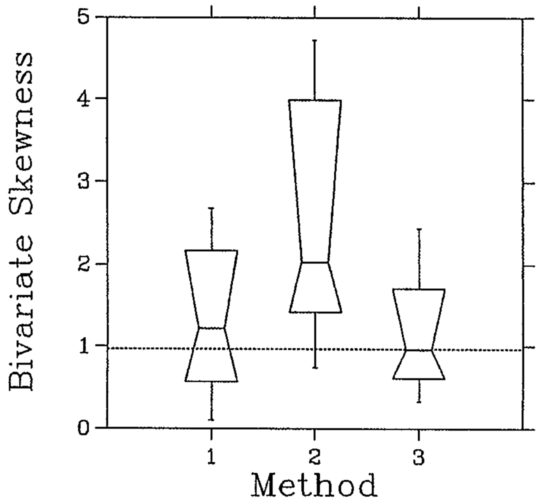

Figure 1: Boxplots of the Estimated Bivariate Skewness of the Initial Observations of the AR(2) Example. (Method: 1=Random Shock, 2=Linear Prediction, 3=Adjusted Linear Prediction)

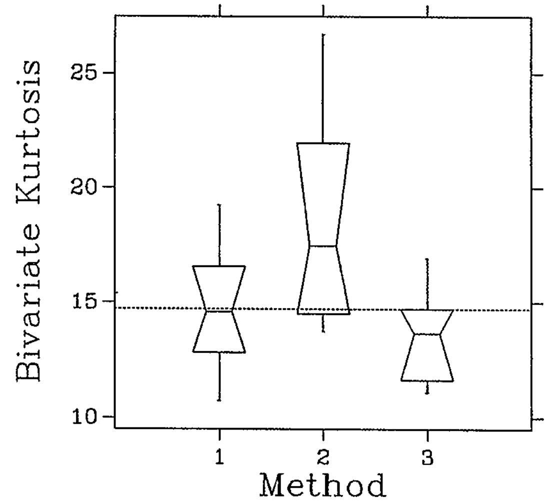

Figure 2: Boxplots of the Estimated Bivariate Kurtosis of the Initial Observations of the AR(2) Example. (Method: 1=Random Shock, 2=Linear Prediction, 3=Adjusted Linear Prediction)

ence between the methods at the 95% confidence level. Further graphical and statistical analysis of the simulation output is in progress.

5. DISCUSSION

Theoretical and empirical comparison of the three methods indicates that for *normal* random shocks, the Linear Prediction Method provides initial observations from the correct joint distribution and does not require an induction period. For *nonnormal* random shocks, if moments of order less than or equal to four are of interest, the Adjusted Linear Prediction Method may yield adequate results. For a specified ARMA model, the computation time of IA 1 is linearly related to the order m of the moving average approximation while the computation times of the other algorithms are constant. Nevertheless, the Random Shock Method may be the optimal initialization method for ARMA models with nonnormal random shocks since the time required to obtain the initial values should ensure the correct joint distribution and eliminate the need for an induction period.

ACKNOWLEDGEMENTS

The author wishes to thank Steven Isham for assistance in producing the graphical displays of the simulation results.

REFERENCES

Anderson, O. D. (1975). On the collection of time series data. *Operations Research Quarterly* **26**, 331–335.

Anderson, O. D. (1979). On warming-up time series simulations generated by Box-Jenkins models. *Journal of the Operations Research Society* **30**, 587–589.

Bacon-Shone, J. (1985). Algorithm AS 210: Fitting five parameter Johnson S_B curves by moments. *Applied Statistics* **34**, 95–101.

Durbin, J. (1960). The fitting of time series models. *Review of the International Institute of Statistics* **28**, 233–244.

Elderton, W. P. and Johnson, N. L. (1969). *Systems of Frequency Curves*. Cambridge University Press, Cambridge.

Hettmansperger, T. P. (1984). Two-sample inference based on one-sample sign statistics. *Applied Statistics* **33**, 45–51.

Hill, I. D. (1985). Algorithm AS 100: Normal-Johnson and Johnson-Normal transformations. In *Applied Statistics Algorithms* (P. Griffiths and I. D. Hill, eds.). Ellis Horwood, Chichester, 183–187.

Hill, I. D., Hill, R., and Holder, R. L. (1985). Algorithm AS 99: Fitting Johnson curves by moments. In *Applied Statistics Algorithms* (P. Griffiths and I. D. Hill, eds.). Ellis Horwood, Chichester, 171–182.

Johnson, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika* **36**, 149–176.

Johnson, N. L. (1965). Tables to facilitate fitting S_U frequency curves. *Biometrika* **52**, 547–550.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika* **57**, 519–530.

Martin, R. D., and Zeh, J. E. (1977). Determining the character of time series outliers. *Proceedings of the Business and Economic Statistics Section of the ASA*, 818–823.

Ord, J. K. (1972). *Families of Frequency Distributions.* Hafner, New York.

Patel, J. K., Kapadia, C. H., and Owen, D. B. (1976). *Handbook of Statistical Distributions.* Marcel Dekker, New York.

Wilson, G. T. (1979). Some efficient computational procedures for high order ARMA models. *Journal of Statistical Computation and Simulation* **8**, 301–309.

AUTHOR BIOGRAPHY

DAVID A. BURN is a statistical software designer with IMSL, Inc. He received a B.A. in mathematics from American International College in 1980, and M.A. and Ph.D. degrees in statistics from The Pennsylvania State University in 1982 and 1985, respectively. His current research interests include time series analysis, design and analysis of simulation experiments, and quantitative graphics. He is a member of ASA, ACM, IIF, and TIMS.

David A. Burn
IMSL, Inc.
2500 CityWest Boulevard
Houston, TX 77042-3020, U.S.A.
(713) 782-6060