The Philippine Statistician, 2001
Vol. 50, Nos. 1-4, pp. 17-24.

General Linear Models with Machine-Generated Random Errors

Roberto N. Padua, Ph.D.*

ABSTRACT

The effects of machine-generated random errors on the bias and variance of the estimates of the parameters of a general linear model are investigated in this paper. Stockis and Tong (1999) did the same analysis on an autoregressive AR(1) model and found that machine-generated errors could introduce bias on the parameter estimates as well as inflate their variances. In this study, a similar result is obtained but we found that the resulting estimates remained unbiased with inflated variances.

KEY WORDS: chaos, dynamical functions, asymptotic normality, linear model.

1. INTRODUCTION

Theoretical results in Statistics are often verified empirically through Monte Carlo Simulation. This verification process invariably requires the use of machine-generated errors which are produced by iterating a linear congruential function of the form:

$$X_t = aX_{t-1} + b(\text{mod}\, c) \tag{1}$$

where a, b and c are relatively prime. Typically, c is a large number, say $c = 2^{31}$-1..

Random errors produced by using equation (1) are pseudo-random numbers in the sense that they are not strictly independent. In fact, equation (1) may be generalized by using a dynamical function $\tau(\,)$:

$$X_t = \tau(X_{t-1}), \qquad t = 1, 2, 3, \ldots \tag{2}$$

Since X_t depends on X_{t-1} in some deterministic way, the sequence of numbers generated by a linear congruential generator or a dynamical function cannot be truly independent. It is natural then to ask the effect of using pseudo-random numbers on the theoretical properties of statistical estimates.

Stockis and Tong (1999) investigated this issue in the case of a simple autoregressive AR(1) model:

$$X_t = \varphi X_{t-1} + \varepsilon_n \tag{3}$$

where, theoretically, ε_n are iid $F()$. The Yule-Walker estimate of φ is given by:

$$\hat{\varphi} = \sum X_t X_{t-1} \Big/ \sum X_t^2 \tag{4}$$

---

* Former Commissioner, Commission on Higher Education, currently Vice-President for Research and International Affairs, Mindanao Polytechnic State College, Cagayan de Oro City

Results of this nature are important since most simulation studies utilize machine-generated errors. Recent researches on random number generation have tended to use chaotic functions such as the logistic function and tent maps as sources of random numbers. The ability of chaotic functions to generate 'almost sure" random numbers derive from the act that these numbers have mixing properties, are aperiodic (almost surely) and are dense in [0, 1].

This paper considers the effect of using random numbers generated from the chaotic functions in the case of general linear models. Section 2 discusses the general linear model of interest and the logistic function as a source of the random errors; Section 3 examines the theoretical effects of these numbers on the estimates; Section 4 gives the conclusions and recommendations.

2. THE GENERAL LINEAR MODEL AND CHAOTIC FUNCTIONS

The general linear model is given by:

$$\mathrm{Y} = \mathrm{X}\beta + \varepsilon \tag{5}$$

where Y is an $n \times 1$ random observable vector, X is an $n \times p$ matrix of constants, β is a $p \times 1$ vector parameter and ε is an $n \times 1$ random unobservable errors. In the classical model, ε is assumed to be composed of iid errors $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ with $E(\varepsilon_n)=0$ and $Var(\varepsilon_n)=\sigma^2$ for all i.

Without knowledge of the functional form of the distribution of ε, the least-squares estimate of β is given by:

$$\hat{\beta} = (\mathrm{X}'\mathrm{X})^{-1}\mathrm{X}'\mathrm{Y} \tag{6}$$

The least-squares estimate of β is unbiased in the sense that:

$$E(\hat{\beta}) = \beta \tag{7}$$

with variance:

$$Var(\hat{\beta}) = \sigma^2(\mathrm{X}'\mathrm{X})^{-1} \tag{8}$$

The asymptotic distribution of $\sqrt{n}\left(\hat{\beta} - \beta\right)$ is multivariate normal with mean zero and covariance matrix $\sigma^2\Sigma$ where:

$$\Sigma = \lim_{n \to \infty} \frac{1}{n}(\mathrm{X}'\mathrm{X})^{-1} \tag{9}$$

Such nice, theoretical properties of $\hat{\beta}$ held in the case that $\{\varepsilon_i\}$ are iid with zero mean and constant variance σ^2. However, they may no longer continue to hold if either the independence or identical distribution requirement is violated.

If we assume that the errors are generated by some chaotic function $\tau(\ )$, then a connection between chaos and probability has to be made. Birkhoff (as quoted by Hunt (1995) and Berliner (1977)) made such a connection via the ergodic theory. In simple terms, this theory states that for a large enough sample size n, there exists an invariant measure $G(\ )$ such that:

$$\frac{1}{k}\sum_{k}\tau^{k}(X_0)\rightarrow\int_{A}XdG(X)\quad \text{as } k\rightarrow\infty \tag{10}$$

where $\{X_0,\tau(X_0),\tau^2(X_0),\ldots\}$ is the trajectory of $\tau(\cdot)$. In a sense, $G(\cdot)$ describes the probability distribution of the trajectory.

When such an invariant measure exists, then one can think of the chaotic numbers as having been generated from the invariant measure itself. For instance, for the logistic map, $\tau(x)=4x(1-x)$, the invariant measure $G(\cdot)$ is the arcsine distribution. Thus, when n is large (to the magnitude of 10^5), the classical results may continue to hold in the case of a general linear model.

3. MAIN RESULTS

Without loss of generality, assume that $E(\varepsilon_i)=0$ (note that if $E(\varepsilon_i)=\mu$, then let $\varepsilon_i'=\varepsilon_i-\mu$). We can write:

$$\hat{\beta}=\beta+(X'X)^{-1}X'\varepsilon \tag{11}$$

Individually, it is therefore possible to write:

$$\begin{aligned}\hat{\beta}_j &= \beta_j+\sum_{k=1}^{n}w_k\varepsilon_k,\quad j=1,\ldots,p\\ &= \beta_j+w_1\varepsilon_1+w_2\varepsilon_2+\cdots+w_n\varepsilon_n\end{aligned} \tag{12}$$

where $\{w_j\}$ is the jth row of $(X'X)^{-1}X'$

If $\varepsilon_t=\tau(\varepsilon_{t-1})$, where $\tau(\cdot)$ is a chaotic function, we require two definitions that will allow us to represent ε_0 as a sum of iid variables $\{Z_t\}$

Definition 1. A shift is a transformation $\sigma:\Omega\rightarrow\Omega$, where the element of Ω are doubly-infinite sequences of the form $w=(\ldots,w_{-1},w_0,w_1,\ldots)$ of element of a finite set V and σ operates on an element in such a way that the nth coordinate of $\sigma(w)$ is w_{n+1}, i.e. $\sigma(\ldots,w_{-1},w_0,w_1,\ldots)\rightarrow(\ldots,w_0,w_1,w_2,\ldots)$.

Definition 2. A Bernoulli shift (ρ,σ) is shift where probabilities P_i are assigned to the elements i of the finite set V in such a way that $\sum_n P_i$ and $\Pr(w_0=i)=P_i$ independent of each n. Clearly the measure so generated on Ω, say P, is invariant under the shift σ.
Using the logistic map can do a simple illustration of the Bernoulli shift. Suppose that we assign the value 1 if $\tau(x)\geq\frac{1}{2}$ and 0 otherwise. Then, starting from $X_0=0.1$, we obtain the path: $\{0.1,\ 0.36,\ 0.9216,\ 0.2890,\ 0.8219\}$ after four iterations. Our rule then transforms

this into $\{0, 0, 1, 0, 1\}$,which looks like a realization from a binomial distribution with parameters n and $p = \frac{1}{2}$.

When a deterministic dynamical function enjoys the Bernoulli shift property, then the generated sequence $\{\varepsilon_t\}$ admits a linear representation:

$$\varepsilon_t - E(\varepsilon_t) = \sum_{t=-\alpha}^{\alpha} \varphi_t Z_{t-1} \tag{13}$$

where $\{Z_t\}$ is a set of independent and identically distributed (iid) random variables with mean zero and finite variance (Stockis and Tong, 1999). The tent map and the logistic map both possesses the Bernoulli shift property. It now follows that

$$\hat{\beta}_i - \beta_i = \sum_i \sum_j w_i \varphi_{ji} Z_{ji} \tag{14}$$

Equation (14) is the weighted sum of iid random variables. The generalized Central Limit Theorem assures us that the asymptotic distribution of $\hat{\beta}_i$ will be normal. In fact,

Proposition 1. The least-squares estimate of β is unbiased for β even if the errors are generated by a chaotic dynamical function.

Proof.

The proof is straightforward:

$$E(\hat{\beta}) = (X'X)^{-1} X'E(Y) = (X'X)^{-1} X'E(X\beta + \varepsilon) = \beta$$

The unbiasedness of $\hat{\beta}$ holds even in the small-sample case provided that $E(\varepsilon)$ is taken with respect to the invariant measure $G(\cdot)$. However, it is not clear at this point in time if the ergodic theory of Birkhoff holds in the small sample case. Research on the rate of convergence to the invariant measure is needed.

Next, we consider the asymptotic distribution of $\hat{\beta}$.

Proposition 2. Let $\{\varepsilon_t\}$ be represented as a linear combination of iid random variables $\{Z_t\}$:

$$\varepsilon_t = \sum_{j=1}^{n} \varphi_{tj} Z_j$$

and suppose that:

(i) $\sum_{j=1}^{n} \varphi_{tj}^2 < \infty$

(ii) $\lim \frac{1}{n}\sum_{i=1}^{n} w_i^2 = \theta_k$

then:

$$\sqrt{n}(\hat{\beta}_k - \beta_k) \xrightarrow{L} N(0, \sigma^2\theta_k \sum_{j=1}^{\infty} \varphi_{ij}^2 \quad \text{as } n \to \infty$$

for $k = 1,\ 2, \ldots, p$.

Proof.

The proof follows from Lindeberg's Central Limit theorem by noting that $(\hat{\beta}_k - \beta_k)$ can be expressed as a linear combination of iid random variables $\{Z_t\}$.

Moreover,

$$\begin{aligned} Var(\hat{\beta}_i) &= Var(\sum_i \sum_{j=1}^{\infty} w_i \varphi_{ij} z_j) \\ &= \sigma^2 (\sum_i \sum_j w_j^2 \varphi_{ij}^2) \\ &= \sigma^2 (\sum_i w_i^2 \sum_{i=1}^{\infty} \varphi_{ij}^2) \end{aligned}$$

which gives the asymptotic variance as $n \to \infty$.

Individually, therefore, the components of $\hat{\beta}$ are asymptotically univariate normal which are unbiased for β and inflated variances (by a factor $\sum_{i=1}^{\infty} \varphi_{ij}^2$).

Proposition 3. Let $\{\varepsilon_t\}$ be represented as a linear combination of iid random variables $\{Z_t\}$:

$$\varepsilon_t = \sum_{j=1}^{\infty} \varphi_{tj} z_{tj}, \quad t = 1,\ 2, \ldots, n$$

where $E(z_{tj}) = 0,\ \ Var(z_{tj}) = \sigma^2$ for all j, and

$$\begin{cases} Cov(z_{tj}, z_{kj}) = \gamma_{tk}, & t \neq k \\ Cov(\varepsilon_t, \varepsilon_k) = \lambda_{tk} \end{cases}$$

Suppose that:

(i) $\sigma^2 \sum_{j=1}^{\infty} \varphi_{tj} = \lambda_{tt} < \infty, \quad t = 1,\ 2, \ldots, n$

(ii) $P_n = (X_n' X_n)^{-1} X_n' \to P \quad \text{as } n \to \infty$

(iii)
$$\Lambda_n = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1n} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{n1} & \lambda_{n2} & \cdots & \lambda_{nn} \end{pmatrix} \to \Lambda \text{ as } n \to \infty$$

Then, $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{L} MVN(\mathbf{0}, P'\Lambda P)$ as $n \to \infty$.

Proof.

It suffices to compute the variance-covariance matrix of $\{Z_t\}$. By Proposition 2,

$$Var(\varepsilon_t) = \sigma^2 \sum_{j=1}^{\infty} \varphi_{tj}^2 = \lambda_{tt}, \quad t = 1, 2, \ldots, n$$

and

$$Cov(\varepsilon_t, \varepsilon_k) = (\sum_j \sum_m \varphi_{tj}\varphi_{km})\gamma_{tk}, \quad t \neq k$$

The result follows by an application of the multivariate central limit theorem.

Stockis and Tong (1999) provided conditions under which the dynamical function $\tau(\cdot)$ admits for a Bernoulli shift representation. We interpret some of these conditions statistically.

The diagonal elements of Λ_n are given by:

$$\sigma^2 \sum_{j=1}^{\infty} \varphi_{kj}^2 = \tau_{kk}, \quad k = 1, 2, \ldots, nn$$

If

$$\sum_{j=1}^{\infty} \varphi_{kj}^2 < \infty,$$

then the diagonal elements are assured of convergence. On the other hand, the off-diagonal elements of Λ_n are

$$\lambda_{kj} = \gamma_{kj} \sum_i \sum_m \varphi_{ki}\varphi_{jm}, \quad k \neq j.$$

Since

$$\left| \sum_i \sum_m \varphi_{ki}\varphi_{jm} \right| \le \left| \sum_i \varphi_{ki}^2 \right| \left| \sum_m \varphi_{jm}^2 \right|,$$

we need only investigate $\gamma_{kj} = Cov(z_{ki}, z_{jm})$. For a fixed k,

$Cov(z_{ki}, z_{km}) = 0$ by the iid assumption.

We note that

$$\varepsilon_k = \tau^k(x_0)$$

for some initial value $x_0 \in U(0,1)$.

Also, $\varepsilon_{k+j} = \tau^{k+j}(x_0)$. Since both $\tau^k(\varepsilon_0) = \varepsilon_k$ and $\tau^{k+j}(\varepsilon_0) = \varepsilon_{k+j}$ depend on ε_0, the correlations between $\tau^k(\varepsilon_0)$ and $\tau^{k+j}(\varepsilon_0)$ cannot be zero.

Machine limitations pose some serious problems in actual simulation studies. In particular, the choice of the initial value ε_0 will always be a rational number. If the chaotic map $\tau(.)$ is the logistic map, then the initial value ε_0 will always be a periodic point of $\tau(.)$. (Padua, 2000)

Theoretically, if ε_0 is a periodic point of period P, the sequence $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_P\}$ will repeat after P iterations so that

$$Cov(\varepsilon_1, \varepsilon_{P+1}) = 1.$$

However, due to machine truncations, the value

$$\varepsilon_1 = \tau(\varepsilon_0)$$

will fall on the orbit of yet another periodic point ε_0^* (also a rational number), the value ε_0^{**} and so on.

For example, starting with $\varepsilon_0 = 0.1$, the logistic map $\tau(x) = 4x(1-x)$ gets trapped to the orbit of the periodic point $\varepsilon_0{}^* = 0.0$ after the 515th iteration, yet $\varepsilon_0 = 0.1$ is a periodic point whose period is not P = 515. Consequently, $Cov(\varepsilon_1, \varepsilon_{P+1}) < 1$, and in fact, it will be close to zero.

Table 1 shows this phenomenon in the case of the logistic map

$$\tau(x) = 4x(1-x)$$

with $x_0 = 0.3$.

Table 1

Autocorrelation Values for the Logistic Map with $x_0 = 0.3$, n = 1000 At Various Lag (Absolute Values)

Lag	1	2	3	4	5	6	7	8	9	10	11
Autocorrelation	.074	.076	.005	.035	.029	.017	.014	.022	.004	.047	.048
Lag	12	13	14	15	16	17	18	19	20	21	22
Autocorrelation	.046	.029	.048	.069	.001	.0004	.029	.049	.048	.047	.071

The correlation values at Lag 1 and Lag 2 are significant at $\alpha = 0.05$, however, the lag correlation up to Lag 9 are not significant (and may be considered equal to zero). The pattern is somewhat repeated from Lag 10 to Lag 18, and so on.

The set of all periodic points of the logistic map forms a dense subset of [0, 1] and is a countable set. The Lebesgue measure of the set of all periodic points of the logistic map is therefore zero, i.e., there are more non-periodic points than periodic points. If it were possible to choose a non-rational starting value on a theoretical computer, then we could theoretically generate a purely iid set of random numbers.

4. CONCLUSION

The study proved that if the error components of a general linear model were generated from a chaotic dynamical function admitting a Bernoulli shift representation, then the estimates of the parameters remain unbiased with inflated variances. However, for small sample sizes, the reference to Birkhoff's ergodic theory need further study.

We recommend a study on the rate at which certain dynamical functions numbers approaching the Frobenius-Perron invariant measure. In particular, such an investigation will put substance on the claim that the estimates of the parameters of the linear model will be unbiased even in the small sample case.

Further research on the connection between chaos and probability needs to be undertaken in view of the current resurgence of interest on the subject mater of dynamical systems. Dynamical systems are seen as good models for explaining biological phenomena (Brown et al (1997)), fluid turbulence, and others.

ACKNOWLEDGMENTS

I would like to thank the editors of the Philippine Statistician and the anonymous referee for their helpful comments and suggestions.

References

BERLINER, J. (1977). "Statistical Aspects of Chaos Theory and Invariant Means" (London: Royal Statistical Society, Vol. XII, pp. 3 – 15)

BROWN, P. and ROTHERY, R. (1997). Models in Biology. (New York, John-Wiley Publishing, USA)

HUNT, R. (1995). Estimating Invariant Measures and the Lyapunov Exponents." (London: Royal Statistical Society, pp. 45 – 67)

PADUA, R. (2000). Statistical Viewpoint of Chaos" (Philippines: Journal of Mathematical Sciences, MPSC pp. 33 – 50)

STOCKIS, Y. and TONG, R. (1999). "Effects on Machine-Generated Errors on AR(1) Models (London: Royal Statistical Society, pp. 97 – 101)

PAULICAN, A. (2000). "Simulation Approach on the Effects of Machine-Generated Errors on General Linear Models" (unpublished doctoral dissertation, Mindanao Polytechnic State College, Cagayan de Oro City)