VALIDATION OF COMPUTER SIMULATION MODELS USING PARAMETRIC TIME SERIES ANALYSIS

Der-Ann Hsu and J. Stuart Hunter

Princeton University

ABSTRACT

In this study a statistical testing procedure is suggested to perform the comparison of the time-dependence structure between realistic and simulation-generated data. The time series technique concerned can be used to perform forecasting, which is usually more efficient and cheaper than ordinary simulation methods, under various experimental conditions.

I. INTRODUCTION

Among the statistical methods available for the validation of computer simulation models, parametric time series analysis is the one which can be used to examine the serial dependence of simulation-generated data. Recently, Box and Jenkins [1] have developed a technique for modeling empirical time series employing a family of mixed autoregressive and moving average (ARMA) processes. The mathematical form of these models is

$$z_t - \phi_1 z_{t-1} - \dots - \phi_p z_{t-p} = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

where z_t is the value of the series, deviating from its mean level, at equally spaced time point t, and a_j is an independently distributed random variable. Under certain invertability conditions, the models above can be transformed to be entirely autoregressive, i.e., no θ_i parameters contained in the formula above. Therefore, it is justifiable to discuss only autoregressive models in the sequel without losing the generality implied on the entire family of models.

II. A TESTING SCHEME

In order to test the identity of the serial-dependence structure of realistic and simulation-generated time series (each contains n observations), a testing scheme is briefly described in this section.

Assume that both realistic and simulation-generated time series can be adequately represented by a model

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + a_t$$

where $E(y_t) = 0$, ϕ_i, $i=1,2,\dots,p$, satisfy certain stationarity conditions and a's are independent, identically distributed normal variables, with mean zero and variance σ^2. Letting superscript "(1)" indicating the variable or parameters related to the realistic series and "(2)" that of the simulation-generated series, we have

$$y^{(1)}_{p+1} = \phi^{(1)}_1 y^{(1)}_p + \phi^{(1)}_2 y^{(1)}_{p-1} + \dots + \phi^{(1)}_p y^{(1)}_1 + a^{(1)}_{p+1}$$

....

$$y^{(1)}_n = \phi^{(1)}_1 y^{(1)}_{n-1} + \phi^{(1)}_2 y^{(1)}_{n-2} + \dots + \phi^{(1)}_p y^{(1)}_{n-p} + a^{(1)}_n ,$$

or in matrix notation,

$$\underset{(n-p)\times 1}{\underset{\sim}{y}^{(1)}} = \underset{(n-p)\times p}{X^{(1)}} \; \underset{p\times 1}{\underset{\sim}{\phi}^{(1)}} + \underset{(n-p)\times 1}{\underset{\sim}{a}^{(1)}}$$

The variance of $a^{(1)}$'s is denoted by σ^2_1.

Similarly,

$$\underset{\sim}{y}^{(2)} = X^{(2)} \underset{\sim}{\phi}^{(2)} + \underset{\sim}{a}^{(2)}$$

and the variance of $a^{(2)}$'s is denoted by σ^2_2.

Define $\underset{\sim}{\psi} = \underset{\sim}{\phi}^{(1)} - \underset{\sim}{\phi}^{(2)}$, $\gamma = \sigma^2_2 / \sigma^2_1$ and $\nu = n-2p$.

Adopting Bayesian approach, and using the diffusive prior density for parameters concerned, it can be shown [2] that the joint posterior density function of $\underset{\sim}{\psi}$ and γ is

$$p(\underset{\sim}{\psi}, \gamma \mid \underset{\sim}{y}^{(1)}, \underset{\sim}{y}^{(2)}) \propto \frac{\left| (X^{(1)\prime}X^{(1)})^{-1} + \gamma (X^{(2)\prime}X^{(2)})^{-1} \right|^{-\frac{1}{2}}}{(S(\gamma))^p} \times \left\{ 1 + \frac{(\underset{\sim}{\psi} - \hat{\underset{\sim}{\psi}})' [(X^{(1)\prime}X^{(1)})^{-1} + \gamma (X^{(2)\prime}X^{(2)})^{-1}]^{-1} (\underset{\sim}{\psi} - \hat{\underset{\sim}{\psi}})}{2 \nu S^2(\gamma)} \right\}^{-(\nu + \frac{p}{2})} \gamma^{\frac{\nu}{2}-1} \left(1 + \frac{S_1(\hat{\underset{\sim}{\phi}}^{(1)}) \gamma}{S_2(\hat{\underset{\sim}{\phi}}^{(2)})}\right)^{\nu}$$

where $S_1(\hat{\phi}^{(1)})$ and $S_2(\hat{\phi}^{(2)})$ are sum of squares of residuals calculated based on the least squares estimates of $\phi^{(1)}$ and $\phi^{(2)}$ respectively, $\hat{\psi} = \hat{\phi}^{(1)} - \hat{\phi}^{(2)}$, $S^2(\gamma) = \frac{1}{2\nu}(S_1(\hat{\phi}^{(1)}) + \frac{1}{\gamma} S_2(\hat{\phi}^{(2)}))$, $\gamma > 0$ and $-\infty < \psi_i < \infty$, $i=1,2,\ldots,p$. It can be further proved that

$$G(\psi,\gamma) = \log \left\{ \frac{p(\hat{\psi},\hat{\gamma} \mid y^{(1)}, y^{(2)})}{p(\psi,\gamma \mid y^{(1)}, y^{(2)})} \right\} \overset{\rightarrow}{\sim} \frac{1}{2} \chi^2_{p+1}$$

where "$\overset{\rightarrow}{\sim}$" denotes "is asymptoticly distributed as". Thus, the approximate $(1-\alpha)100\%$ Highest Posterior Density (HPD) region of ψ and γ can be found based on this result.

It is apparent that one may not be satisfied with just rejecting the hypothesis that the simulated time series has ϕ and σ^2 which are identical to the realistic data. Rather, one may ask which of the following four possible conclusions (i) $\psi = 0$, $\gamma = 1$; (ii) $\psi \neq 0$, $\gamma = 1$; (iii) $\psi = 0$, $\gamma \neq 1$; and (iv) $\psi \neq 0$, $\gamma \neq 1$ is true such that the improvement of the simulation models can be further pursued. This is so called simultaneous inference. A convenient procedure, which involves two stages, for distinguishing these four cases is reported below.

Stage 1: Examine $G(0,1)$. If $G(0,1) < \frac{1}{2}\chi^2_{p+1}(0.05)$, where 0.05 is the significance level, case (i) is recommended. Otherwise, go to the second stage.

Stage 2: If $G(\hat{\psi},1) < \frac{1}{2}\chi^2_{p+1}(0.05)$, case (ii) will be accepted. Otherwise, examine $G(0,\hat{\gamma})$. If $G(0,\hat{\gamma}) < \frac{1}{2}\chi^2_{p+1}(0.05)$, case (iii) is recommended. Otherwise, select case (iv).

III. FORECASTING

An ordinary method of performing simulation forecasting under experimental conditions is to repeat the simulation process for a certain number of times. A rough confidence interval can thus be constructed. But this can be very inefficient and expensive when the simulation model is huge and complicated. As soon as the simulation is validated as reasonably good for its purpose, through the procedure suggested above, a single run of the simulation under the condition desired will produce a time series enough for the modeling using Box-Jenkins' technique. The forecasting from the model identified is a straightforward matter as demonstrated in [1].

IV. REMARKS

(1) A philosophical discussion of the problems of validation of computer simulation models has been presented by Naylor and Finger [3]. In addition to their suggestion of a three-stage verification of simulation (examination of postulates, the statistical estimation, and predictive power of the simulation), we would like to propose a recursive feedback process as a fundation of the evolution of computer simulation. The evolutionary process consists of (i) theoretical formulation of the simulation system; (ii) statistical construction of the models; and (iii) diagnostic checking. The results secured from the third stage will provide information necessary for a reformulation at the next replication of the process. The methodology suggested in the previous sections is just one sort of the dynamic diagnostic instruments which can be employed in the third stage of the evolutionary procedure. An emphasis we have to make is that the dynamic behaviour observed from the major simulated variables will reveal some of the most important aspects of the adequacy of a simulation system.

(2) The discussion presented in the previous sections has aimed at the comparison of realistic versus simulated data. The comparison can also be done for the data generated from different simulation experiments. For instance, to compare the effects of two sets of economic policies employed by the government on the business cycles, the values of ϕ and σ^2 estimated from the experiments may have direct economic interpretations.

REFERENCES

[1] Box, G.E.P. and Jenkins, G.M. (1970), Time Series Analysis: Forecasting and Control, San Francisco; Holden-Day.

[2] Hsu, D.-A. (1973), Stochastic Instability and the Behaviour of Stock Prices, Ph. D. Dissertation, Department of Statistics, University of Wisconsin - Madison.

[3] Naylor, T.H. and Finger, J.M. (1967), "Verification of Computer Simulation Models", Management Science, Vol. 14, No. 2, Oct. 1967, pp. 92-101.