Stochastic simulation as a validation tool for econometric models

Calzolari, Giorgio and Corsi, Paolo

IBM Pisa Scientific Center

15 September 1977

Reprinted from:

Models for Regional Planning and Policy-Making

Proceedings of the Joint IBM/IIASA Conference,
Vienna September 1977

UKSC 0097

Co-edited by Professor A. Straszak (IIASA)
Dr. B. V. Wagle (IBM (UK) Scientific Centre)

6.5 STOCHASTIC SIMULATION AS A VALIDATION TOOL FOR ECONOMETRIC MODELS

G. CALZOLARI and P.CORSI, IBM Scientific Centre, Pisa, Italy.

6.5 STOCHASTIC SIMULATION AS A VALIDATION TOOL FOR ECONOMETRIC MODELS

G. CALZOLARI and P.CORSI, IBM Pisa Scientific Centre, Pisa, Italy.

INTRODUCTION

The complete validation of an econometric model is a process which involves a formidable number of activities in the various steps of model building. Several classifications have been presented in econometric literature (References 6 and 10); they individuate a series of problems in the areas of economic structure specification, test of hypothesis and parameter estimation, simulation behaviour and decision making.

We do not want to enter into details of each of the above steps, nor do we want to develop a complete and systematic analysis of the validation process. Our attention will be mainly focussed on some aspects related to simulation and decision making; in particular, the purpose of this paper is to analyze some problems of the validation process for which stochastic simulation (for some technical details, see Appendix) can be profitably used.

In fact, even if deterministic simulation is certainly a powerful tool in this area (References 8,19) nevertheless stochastic simulation seems to be preferable; firstly, because the same information can be drawn with greater methodological correctness, and especially because it can supply additional information.

These techniques prove to be useful mainly in the field of non-linear models; when a model is linear, in fact, analytical procedures are generally preferred.

In this paper, reference will be continuously made to a non-linear model of the Italian economy developed by a research team co-ordinated by the Institute for Economic Planning (ISPE).

The structure of this model is described in detail in another paper presented at this conference (Reference 17); it will be hereafter referred to as ISPE model.

No conclusions will be drawn about the validity of the ISPE model; as already mentioned, the purpose of the paper is exclusively methodological.

What should be pointed out is the fact that, unlike what generally happens, for this model stochastic simultion has been used also in the building phase, giving in this way some indications and suggestions to the model builders.

Some methodological problems of deterministic versus stochastic simulation are discussed in the following section.

The problems related to the presence of heteroschedasticity in the reduced form of the model are analysed on pages 361 to 363 (for the meaning of reduced form in the case of non-linear models, see Reference 11).

The problem of individuation of model subsectors transmitting errors to other sectors is empirically tackled in the section entitled Transmission of Errors through the Model.

The following two sections (see pages 364 and 365) deal with the problem of model reliability respectively in one-step and dynamic forecasting.

In the Appendix some technical details on stochastic simulation are briefly presented.

DETERMINISTIC VERSUS STOCHASTIC SIMULATION

The considerations developed in this section can be more easily understood if reference is made to a small model as an example. Let us suppose that we have the following two equation model (a simplified version of of the example in Reference 11),

$$\log(y_{1,t}) = b \cdot \log(x_{1,t}) + u_t$$

$$y_{2,t} = x_{2,t} + y_{1,t}$$

where the disturbances u_t are normally distributed with mean zero, variance of $\sigma 2 \neq 0$, not autocorrelated, and independent of the predetermined variables. The reduced form is -

$$y_{1,t} = x_{1,t}^{b} \cdot \exp(u_t)$$

$$y_{2,t} = x_{2,t} + x_{1,t}^{b} \cdot \exp(u_t)$$

The random variable $\exp(u_t)$ has log-normal distribution, whose mean is $\exp(\sigma^2/2)$ (Reference 14 p. 400), different from 1 as $\sigma^2 \neq 0$. The conditional expectation of $y_{2,t}$ given the predetermined variables, is therefore:-

$$E(y_{2,t} \mid x_{1,t}, x_{2,t}) = x_{2,t} + x_{1,t}^{b} \cdot \exp(\sigma^2/2)$$

If we solve the model deterministically, that is setting $u_t = 0$, the result for $y_{2,t}$ would be $x_{2,t} + x_{1,t}^{b}$. This example shows the methodological inconsistency of non-stochastic (or deterministic) simulation: the solution values of endogenous variables in non-linear models can be different from the conditional expectations of the same variables. With respect to the historical data, the simulation results "can be expected to diverge systematically from the corresponding elements of y_t (historical values)" (Reference 11). The bias, of course, involves all the validation analysis based on deterministic simulation results, such as goodness of fit measures, multiplier analysis, forecasting etc. Outcomes of stochastic simulation, on the contrary, are generated by the same process as that generating historical values, so that they should be considered methodologically correct.

Coming to the empirical aspects of this problem, it can be of interest to estimate the magnitude of the above mentioned bias in models of practical use. As far as the authors' experience is concerned (Reference 1), this systematic divergence seems to be always so small as to let us consider the results of deterministic simulation as practically unbiased. With reference, for example, to some variables of the ISPE model, the following results have been obtained (for the meaning of the variables, see Reference 17).

Variable	Deterministic Solution	Mean Stoch. Solution	Std.dev. of mean stoch.
CPNL	92209.1	92244.2	10.29
MT	13587.5	13595.4	3.129
PILL	143111.	143166.	13.91
PVAP	2.11320	2.11377	.0002

TABLE 1

ISPE MODEL. ONE STEP SIMULATION
15000 REPLICATIONS. YEAR 1976.

The values reported in the columns of deterministic and mean stochastic solutions are clearly very close to each other. The existence of a slight bias is, however, rather evident; the mean stochastic solutions, in fact, even if generated by a Monte Carlo method, and consequently not exact, have been computed as sample means of 15000 replicated simulations, so that their standard deviations (which go to zero as the number of replications approaches infinity) are significantly smaller than the estimated biases.

REDUCED FORM HETEROSCHEDASTICITY

Again with reference to the two equation sample model of the previous section, it is clear that if u_t is supposed to be homoschedastic, then the reduced form disturbances, for example referred to $y_{2,t}$,

$$\{y_{2,t} - E\{y_{2,t} \mid x_{1,t}, \quad x_{2,t}\}\} = x_{1,t} b_{1,t} \cdot \{\exp\{u_t\} - \exp\{\sigma^2/2\}\}$$

are clearly heteroschedastic if $x_{1,t}$ is not constant over time (with trend, for example).

Let us now suppose that, in a more complicated model, $y_{2,t}$ appears as explanatory variable in another structural equation, again estimated under homoschedasticity hypothesis. It is clear that the variable defined by this last equation 'the properties of the reduced-form disturbances should not be inferred from those of the structural disturbances' (Reference 11). In other words, the comparison between historical and computed values of endogenous variables should be very carefully used, that is it should take into account this effect. As, however, this comparison is generally performed without any particular care in model validation, stochastic simulation can be profitably used in refining the model (prior to release of the model as in Reference 6).

By means of replicated one-step simulations over the sample period, a series of values of the reduced form standard deviations can be computed for each endogenous variable. If the heteroschedasticity effect seems to be unacceptable, the model builder can intervene modifying the structural form of one or more equations.

In the case of the ISPE model, this effect was found to be particularly strong in a preliminary version of the model. For example, for the variable CS (social security contributions) the reduced form variance increased more than 100 times from the beginning to the end of the sample period. Several functional specifications have been consequently modified, by introducing logarithms and first relative differences instead of levels for several variables. In this way, the heteroschedasticity effect has been significantly reduced.

Year	IAB	PVAI
1961	212.0	.0168
1962	190.4	.0173
1963	170.4	.0200
1964	162.9	.0223
1965	172.9	.0202
1966	178.5	.0195
1967	165.2	.0204
1968	166.2	.0202
1969	156.9	.0189
1970	150.6	.0203
1971	145.5	.0249
1972	134.8	.0253
1973	130.8	.0253
1974	129.9	.0359
1975	120.4	.0444
1976	119.0	.0486

TABLE 2

REDUCED FORM STANDARD DEVIATIONS OVER TIME.
1000 REPLICATIONS

In the present version of the model, among the variables for which this effect seems still to exist, even if much more acceptable than before, we can mention IAB (investment in residential construction) and PVAI (price deflator of gross industrial product).

For these two variables, the reduced form standard deviations over a few years of the sample period are displayed in Table 2.

TRANSMISSION OF ERRORS THROUGH THE MODEL

The importance of individuating blocks of the model transmitting errors to other sectors has been recognized by various authors (see, for example, Reference 7). In Reference 6, this problem is considered as a part of error decomposition in the larger class of non-parametric measures.

In order to analyze this mechanism of transmission, experiments are performed in which stochastic disturbances are introduced only into a single behavioural equation or into a set of behavioural equations of the same sector.

With reference to the ISPE model, experiments of this type performed in a preliminary version clearly showed that the labour market sector was responsible for the transmission of the largest errors through the whole model.

The reduced form standard deviations in 1975, corresponding to insertion of disturbances into the various sectors of that preliminary version, are displayed in Table 3 for the variables CPNL (private consumption expenditures) and CS (social security contributions).

Disturbed sector	CPNL	CS
Whole model	2250.	1197.
Internal demand	1246.	264.9
Foreign trade	786.5	344.7
Internal supply	39.65	3.777
Labour market	2013.	1045.
Prices	667.6	99.91
Wages	434.3	211.2
Public sector	85.10	111.9

TABLE 3

REDUCED FORM STANDARD DEVIATIONS AT 1975 INDUCED BY THE VARIOUS SECTORS. (PRELIMINARY ISPE MODEL). 200 REPLICATIONS

The updated version of the model seems to work much better, not only because the values of the reduced form standard deviations are in any experiment smaller than the previous ones, but mostly because the distribution of errors is much more uniform, in the sense that no bad sector can be identified.

ONE-STEP FORECASTING

Other interesting considerations can be drawn from stochastic simulation experiments comparing, in terms of mean values and standard deviations, some target variables of economic policy. It is well known (see, for example Reference 13, p.261) that the forecast error can be decomposed into two terms respectively, due to errors in coefficients estimation and to structural residuals. The computation of the statistical properties of the first term seems to be overwhelming, and it has been completely developed only for linear models, (Reference 9).

In spite of the interest and importance of this effect, stochastic simulation results are generally conditional on estimated coefficients, that is estimated coefficients are considered exact (see, for example, Reference 18).

Where stochastic simulation can help is in the computation of the statistical properties of the second term, in particular the reduced form standard deviation.

Table 4 presents these computed reduced form standard deviations, mean values across replications and their percentual ratio for some selected target endogenous variables of the ISPE model. The values have been obtained by means of one-step simulation for 1977, which is the first year beyond the sample period.

Variable	Comp.mean	Std.dev	Pearson coeff.
IFIT	6994.9	310.4	4.44
LI	7424.3	148.9	2.00
VAP	56068.	922.1	1.64

TABLE 4

ONE-STEP SIMULATION. 1000 REPLICATIONS AT 1977

The ratio in percentual form between standard deviation and computed mean, called Pearson's coefficient of variation (Reference 12 p. 47), should give an idea of the reliability of the model in forecasting.

DYNAMIC FORECASTING

An alternative to one-step is dynamic simulation, which consists in setting lagged endogenous variables equal to the computed values rather than to the historic ones.

In this case, besides the two sources of errors mentioned in the previous section, an additional source of errors is introduced by lagged endogenous variables. Therefore, application of dynamic method induces an accumulation of errors, that can be shown by stochastic simulation. For example for the same variables and the same year 1977 as in Table 4, sample means and standard deviations computed via dynamic simulation are displayed in Table 5; the initial year of this experiment is 1974 (the maximum lag of endogenous variables in this model is 3 years).

Variable	Comp.mean	Std.dev	Pearson coeff.
IFIT	7380.1	434.2	5.88
LI	7523.9	186.5	2.48
VAP	57227.	1236.	2.16

TABLE 5

DYNAMIC SIMULATION. INITIAL YEAR 1974.
1000 REPLICATIONS. VALUES AT 1977

When performing medium and long term forecasting beyond the sample period dynamic simulation is the only available tool, being values of lagged endogenous variables not disposable. Table 6 displays, for the same variables as in Tables 4 and 5, the computed values and standard deviations obtained via dynamic simulation for the period 1978-1980, conditional on values of the exogenous variables pre-assigned by the model builders.

	IFIT		LI		VAP	
Year	Mean	Std.dev.	Mean	Std.dev.	Mean	Std.dev.
1978	7252.3	362.4	7460.8	172.5	57601.	1024.
1979	7525.2	378.7	7506.8	181.1	59547.	1182.
1980	7775.5	420.0	7567.8	185.0	61343.	1328.

TABLE 6

DYNAMIC SIMULATION. INITIAL YEAR 1977. 1000 REPLICATIONS.

APPENDIX

To begin with, we want to recall that an econometric model can generally be represented, from the solution point of view by -

$$y_{i,t} = f_i(y_{j,t-k}, x_{n,t}, u_{i,t})$$

where $y_{i,t}$ are the current endogenous variables, $y_{j,t-k}$ are current and lagged (k is the related lag) values of endogenous variables, $x_{n,t}$ are the exogenous variables, $u_{i,t}$ are the random disturbances which are supposed to have a multivariate normal distribution with zero mean, assigned constant covariances and, in our case, no serial correlation.

Deterministic simulation is the simultaneous solution of an econometric model, obtained by replacing the structural disturbances (inserted in the behavioural equations during the specification and estimation phases) with their mathematical expectations, which are zero.

Stochastic simulation is the simultaneous solution of an econometric model obtained by adding to the intercept of each behavioural equation a pseudo-random shock possessing specified stochastic properties (which are in general related to those of the structural disturbances of the specification phase).

The key feature of stochastic simulation is that pseudo-random shocks can be replicated, so that a distribution of outcomes for endogenous variables in each period can be obtained. For a systematic analysis of stochastic simulation methodology and significance, one could refer to (Reference 18).

The generation of the structural disturbances $u_{i,t}$ involves three different steps -

A Generation of independent pseudo-random numbers uniformly distributed in the open interval (0,1). In this application the power residue method has been used, with prime modulus $2^{31}-1$ and its primitive root 7^5 as a multiplier (Reference 15).

B Transformation of the previously generated numbers into standard normal deviates. The logarithmic-trignometric procedure by Box and Muller (Reference 3) has been applied, after an intermediate phase of shuffling (Reference 4) to avoid "biases".

C Transformation of the standard normal deviates into the required pseudo-random disturbances, whose covariance matrix must be "equal" to "the sample covariance matrix estimated from regression residuals" (Reference 5, p.124). The algorithm by McCarthy (Reference 16) has been chosen in this application.

Finally, coming to some computational aspects, the displayed results have been obtained by means of a package developed at the IBM Scientific Centre of Pisa, (Reference 2). The package has been written in FORTRAN-G language and works under the operating system VM-370/CMS on an IBM/370 model 168.

REFERENCES

1 Bianchi,C., Calzolari,G., Corsi,P.,
'Divergences in the Results of Stochastic and Deterministic Simulation of an Italian Non-Linear Econometric Model',
L.Dekker,Ed., 'Simulation of Systems, 8th AICA Congress', North Holland, Amsterdam, (1976).

2 Bianchi,C., Calzolari,G., Corsi,P.,
'A Program for Stochastic Simulation of Econometric Models',
Econometrica, (forthcoming in January or March 1978 issue).

3 Box,G.E.P., Muller, M.E.,
'A Note on the Generation of Random Normal Deviates',
Ann.Math.Stat., 29, (1958).

4 Calzolari,G., Ciriani,T.A., Corsi,P.,
'Generation and Testing of Pseudo-Random Numbers with Assigned Statistical Properties to be Used in the Stochastic Simulation of Econometric Models',
IBM Technical Report CSP034/513-3544, (1976).

5. Cooper,J.P.,
'Development of the Monetary Sector, Prediction and Policy Analysis in the FRB-MIT-Penn Model',
Lexington Books, London, (1974).

6 Dhrymes,P.J., Howrey,E.P., Hymans,S.H., Kmenta,J., Leamer,E.E., Quandt,R.E., Ramsey,J.B., Shapiro,H.T., Zarnowitz,V.,
'Criteria for Evaluation of Econometric Models',
Annals of Economic and Social Measurement, 1/3 (1972).

7 Fromm,G., Taubman,P.,
'Policy Simulations with an Econometric Model',
North Holland, (1968).

8 Fromm,G., Klein,L.R.,
'A Comparison of Eleven Econometric Models of the United States',
The American Economic Review, v.LXIII, n.2, (1973).

9 Goldberger,A.S., Nagar,A.L., Odeh,H.S.,
'The Covariance Matrix of Reduced-Form Coefficients and of Forecasts for a Structural Econometric Model',
Econometrica, v.29, n.4, (1961).

10 Howrey,E.P.,
'Selection and Evaluation of Econometric Models',
Goldberg,W., Ed., 'Simulation versus Analytical Solution for Business and Economic Models', Gothenburg,(1973).

11 Howrey,E.P., Kelejian,H.H.,
'Simulation Versus Analytical Solutions; The Case of Econometric Models',

T.H.Naylor, Ed., 'Computer Simulation Experiments with Models of Economic Systems', John Wiley, New York (1971).

12 Kendall,M.G., Stuart,A., 'The Advanced Theory of Statistics', v.1, Charles Griffin, London, (1969).

13 Klein,L.R., 'A Textbook of Econometrics', Prentice Hall, Englewood Cliffs, (1974).

14 Kmenta,J., 'Elements of Econometrics', Macmillian, New York, (1971).

15 Lewis,P., Goodman,A., Miller, J., 'A Pseudo-Random Number Generator for the System/360', IBM System Journal,2 (1969).

16 McCarthy,M.D., 'Some Notes on the Generation of Pseudo-structural Errors for Use in Stochastic Simulation Studies', B.G. Hickman, Ed., 'Econometric Models of Cyclical Behavior', Studies in Income and Wealth n.36, NBER, New York, (1972).

17 Sartori,F., 'An Economic Model of Italy', IBM UK Scientific Centre / IIASA Seminar on Models for Regional Planning and Policy Making, Vienna, (1977).

18 Sowey,E.R., 'Stochastic Simulation of Macroeconometric Models; Methodology and Interpretation', A.A. Powell, and R.A. Williams, Eds., 'Econometric Studies of Macro and Monetary Relations', North Holland, Amsterdam, (1973).

19 Zellner,A., Peck, S.C., 'Simulation Experiments with a Quarterly Macroeconometric Model of the U.S. Economy', A.A. Powell, and R.A. Williams, Eds., 'Econometric Studies of Macro and Monetary Relations', North Holland, Amsterdam, (1973).