Monte Carlo methods in econometrics: a package for the stochastic simulation

Bianchi, Carlo and Calzolari, Giorgio and Corsi, Paolo

IBM Scientific Center, Pisa, Italy.

1976

CONGRES EUROPEEN DES STATISTICIENS

MONTE CARLO METHODS IN ECONOMETRICS: A PACKAGE FOR THE STOCHASTIC SIMULATION

C.Bianchi, G.Calzolari, P.Corsi

IBM Scientific Center - Pisa

GRENOBLE 6 - 10 SEPTEMBER 1976

MONTE CARLO METHODS IN ECONOMETRICS: A PACKAGE FOR THE STOCHASTIC SIMULATION

C.Bianchi, G.Calzolari, P.Corsi

IBM Scientific Center
Via S.Maria 67, Pisa, Italy

In this paper, a package implemented at the Scientific Center of IBM Italy in Pisa for the stochastic simulation of linear and non-linear econometric models is presented. After a survey on the adopted methodologies, the input requirements and the produced output are described in some details, using as a sample the Klein model-1. To finish, the performances of the program are analyzed in terms of storage requirements and computation time.

KEYWORDS: Stochastic Simulation, Econometric Models, Computer Programs.

1. Introduction

Once an econometric model has been formulated and its structural coefficients have been estimated on the base of the disposable data, two different types of simulation (or solution) can be performed: the first, called deterministic simulation, consists in solving the model as a system of simultaneous equations setting the structural disturbances equal to their expected values, which are zero; the second, called stochastic simulation, consists in an application of Monte Carlo methodologies to the model, generating pseudo-random disturbances with specified statistical properties, and introducing these generated disturbances into the model during the simultaneous solution. To be more precise, let us formalize an econometric model, as usual [9], in its structural form:

$$Y_t=f(Y_t,Y_{t-1},Y_{t-2},\ldots,X_t,X_{t-1},X_{t-2},\ldots)+U_t$$

where Y_t is the vector of the endogenous (jointly dependent) variables at time t (or lagged of one or more periods), X is the vector of exogenous (predetermined) variables and U_t is the vector of the structural disturbances at time t. During the estimation phase, the coefficients of the structural function f are derived on the base of the historical values (time series) of the involved variables and of the hypotheses about the disturbance terms (usually zero mean, multivariate normal distribution with unknown variance-covariance matrix). Deterministic simulation is performed setting $U_t=0$, so that, for each endogenous variable and for every period one value is obtained as solution. Stochastic simulation, on the contrary, takes into account the disturbance terms, solving the model after adding a vector of pseudo-random numbers E drawn from a prespecified multivariate distribution. "The joint distribution from which these stochastic elements are drawn should, naturally, reflect the true model structure as fully as possible" [16]. For this reason the suggested algorithms are related to the residuals obtained in the estimation phase, in order to generate pseudo-random numbers E_t with the same statistical properties of the structural disturbances U_t. Replicated solutions, as usual in Monte Carlo experiments, generate a distribution of outcomes for each endogenous variable, allowing to draw some statistical inferences.

Describing the purposes of stochastic simulation is not among the aims of this paper (see for example [3], [5] and [16]). We shall therefore confine ourselves to a description of the adopted methodologies and the requirements and performances of the package that has been implemented at the Scientific Center of IBM Italy in Pisa. It must be underlined that this package has been designed for research purposes only and that preference has been given to simplicity of programming and implementation rather than to highly sophisticated performances or facilities.

2. The algorithms

The generation of the vector of pseudo-structural disturbances E_t to be added in the simulation phase, as above mentioned, requires in this package three main steps:

1) Generation of pseudo-random numbers with continuous uniform distribution in the open interval (0,1).
The power residue method is used in the program for this purpose, using as a modulus the prime number $2^{31}-1$ and its primitive root 7^5 as a multiplier, with some minor modifications from the algorithm described in [11]. The numbers generated by this algorithm are then shuffled using a simple 128 cells algorithm, addressed by the lowest order 7 bits of each number [12].

2) Generation of independent pseudo-random numbers with univariate normal distribution, zero mean and unit variance.
Two alternative methods can be used in this phase; they are the Box-Muller (or polar) method [1] and the inverse algorithm [6]. Both the methods, that use the uniform numbers as input, are theoretically exact, not involving asymptotic properties, even if they involve, of course, numerical approximations. The bias arising in the joint use of the power residue method and the Box-Muller transformation [15] is avoided by the above mentioned intermediate phase of shuffling.

3) Generation of pseudo-random numbers with multivariate normal distribution, zero mean and assigned variance-covariance matrix.
Also in this stage two alternative methods are available. The first is based on the triangularization of the covariance matrix (among equations) as it is computed in the estimation phase; it was proposed for application to the stochastic simulation of econometric models by Nagar [14]. The second, proposed by McCarthy [13], is directly based on the use of the estimated residuals. The last method is of more general applicability, being Nagar's algorithm applicable only when the number of stochastic (behavioural) equations is smaller than the sample period length (number of residuals in each equation).

The numerical solution of the system of equations representing the econometric model is performed by means of the Gauss-Seidel iterative procedure. It has been preferred to other methods because of the simplicity of use and the applicability to linear and non-linear models without differences. In default of other specifications, the initial values for the iterative procedure are chosen equal to the historical values of the endogenous variables at the same year.

3. Input requirements

First of all it must be underlined that the package runs on an IBM/370 model 168, under the operating system VM-370/CMS [7], and its use is strictly based on the facilities of this system. To perform stochastic simulation of a model, two files must be prepared on the work minidisk of the user's virtual machine: one containing the model, the second containing the time series values, the estimated residuals and some indications about the simulation to be performed. The characteristics of the two files will be clarified by means of the sample model.

3.1. The Klein model-1 as a sample

Let us suppose to perform stochastic simulation on the classical Klein model-1 [8]. This model is too well known to need a comment and, at the same time, sufficiently small to be used as a sample for illustrating requirements and performances of a package for econometric applications. The model consists of 6 equations, 3 of which behavioural (stochastic). The sample period is from 1920 to 1941. The parameters of the model here presented have been obtained by means of the Three Stage Least Squares method [10].

3.2. The input file for the model

This file must be written with the format of FORTRAN programs and, for our model, it could be as in table 1.
The order in which equations are written is free, while the names Y and YL respectively for current and lagged endogenous variables, X for the exogenous variables, I for the current period of time and E for the disturbances are fixed. This file must have FORTRAN as a filetype (for CMS [7]) while the choice of its filename is free. For example KLEIN1 FORTRAN could be its complete CMS identifier. The file can be entered into the user's minidisk via card-deck or, more commonly, using the conversational EDIT facilities.

Table 1

```
C...........................................................
C   LIST OF ENDOGENOUS VARIABLES
C
C  Y(1)    = C        CONSUMPTION.
C  Y(2)    = I        INVESTMENT.
C  Y(3)    = W1       PRIVATE WAGES.
C  Y(4)    = Y        NATIONAL PRODUCT.
C  Y(5)    = P        PROFITS.
C  Y(6)    = K        END-OF-YEAR CAPITAL STOCK.
C
C   LIST OF EXOGENOUS VARIABLES
C
C  X(1,I) = W2        GOVERNMENT WAGES.
C  X(2,I) = T         INDIRECT TAXES.
C  X(3,I) = TIME      1931=0.
C  X(4,I) = G         GOVERNMENT EXPENDITURES.
C
C   LIST OF EQUATIONS
C
C     CONSUMPTION.
      Y(1) = 16.441 + .79008*(Y(3)+X(1,I)) + .12489*Y(5) +
     +       .16314*YL(5,I-1) + E(1)
C     INVESTMENT
      Y(2) = 28.178 - .013079*Y(5) + .75572*YL(5,I-1) -
     -       .19485*YL(6,I-1) + E(2)
C     PRIVATE WAGES
      Y(3) = 1.7972 + .40049*(Y(4)+X(2,I)-X(1,I)) + .14967*X(3,I)+
     +       .18129*(YL(4,I-1)+X(2,I-1)-X(1,I-1)) + E(3)
C     NATIONAL PRODUCT
      Y(4) = Y(1) + Y(2) + X(4,I) - X(2,I)
C     PROFITS
      Y(5) = Y(4) - Y(3) - X(1,I)
C     CAPITAL
      Y(6) = YL(6,I-1) + Y(2)
C...........................................................
```

3.3. The input file for the time series

This file must be in card format (80 characters per record). The first two records are purely comment lines, which will be printed as a header of the final printout. The third record contains twelve integer numbers, each of five digits, specifying, respectively:

1) Number of replications for each year of the simulation period.
2) Number of endogenous variables (or of equations).
3) Number of stochastic or behavioural equations; its value is the dimension of the vector of the disturbances (E in the input file for the model) for each year.
4) Number of exogenous variables.
5) Initial year of the time series; the first of all, if not all the time series begin in the same period.
6) Final year of the time series; the last among all, if the time series do not end all at the same period.
7) Initial period for the simulation.
8) Final period for the simulation.
9) Flag (0 or 1) to perform (if 1) some descriptive statistics on the generated disturbances (mean and variance-covariance matrix).
10) Number of replications desired for the above statistics (not to be confused with that of point 1).
11) Flag (0 or 1) to perform (if 1) the verify procedure (check if data, coefficients and residuals in the input files have been correctly entered; see partial method in [4]).
12) Flag (0 or 1) to perform one step simulation (if 0: lagged endogenous variables are set to their historical values), or dynamic simulation (if 1: lagged endogenous variables are set to the values computed at the same replication for the previous periods; see total and final method respectively in [4]).

Table 2

```
KLEIN MODEL-1. STOCHASTIC SIMULATION
THREE-STAGE  LEAST SQUARES ESTIMATES
  100    6    3    4 1919 1941 1921 1941    0 1000    0    1
 1920 1941  ENDOGENOUS VARIABLES: CONSUMPTION
     39.800003       41.899994       45.000000       49.199997
     50.600006       52.600006       55.100006       56.199997
     57.300003       57.800003       55.000000       50.899994
     45.600006       46.500000       48.699997       51.300003
     57.699997       58.699997       57.500000       61.600006
     65.000000       69.699997
   ..................(other endogenous variables)................
 1920 1941  EXOGENOUS VARIABLES: GOVERNMENT WAGES
      2.200000        2.700000        2.900000        2.900000
      3.100000        3.200000        3.300000        3.600000
      3.700000        4.000000        4.200000        4.800000
      5.300000        5.600000        6.000000        6.100000
      7.400000        6.700000        7.700000        7.800000
      8.000000        8.500000
   ..................(other exogenous variables)..................
 1921 1941  RESIDUALS: FIRST STOCHASTIC EQUATION
     -0.441652       -1.015035       -1.528914       -0.498503
     -0.013201        0.775871        1.300392        1.099296
     -0.583924       -0.191713       -0.559775       -0.674632
      0.576672       -0.021124        0.053897        1.855538
     -0.459595        0.061281        1.260191        0.949993
     -1.945064
   ..................(other residuals)...........................
```

From the fourth record on, the time series values must be specified. Each time series is written with a standard format 4F15.6 and must be preceded by a record indicating the initial and final period for that series (format 2I5 for yearly data). The order in which time series must be written is the following: at first, the time series of all the endogenous variables, in the same order in which they are indicated in the input file for the model (as variables, not as equations) that is the series of Y(1) (C in our case) followed by that of Y(2) (I) etc.; then the series of all the exogenous variables (X(1,I) in our case); to finish all the series of the residuals produced by the adopted estimation method (the series of E(1) must be the first, etc.). Table 2 presents the time series input file for the sample model.

The CMS identifier of this file must be KLEIN1 DATA, if KLEIN1 FORTRAN is the name of the input file for the model, that is DATA is the fixed filetype, while the filename must be the same as that chosen for the FORTRAN input file. This file could be entered into the user's minidisk via card-deck or using the EDIT facilities, but it is more commonly set up by means of the facilities of the time series processor, implemented at the Pisa Scientific Center of IBM Italy [2].

4. The execution procedure

Once the two files (model and time series) have been built up, the stochastic simulation procedure is activated by the following command issued at the interactive terminal in the CMS environment:
SS KLEIN1
First of all the procedure checks for the existence of the two files KLEIN1 FORTRAN and KLEIN1 DATA. Then it takes KLEIN1 FORTRAN and combines it with a standard header (SUBROUTINE statement and list of arguments) and closing statements. The obtained subroutine is then compiled, loaded (if correct) and repeatedly called by the Gauss-Seidel procedure until convergence is reached. The values of the vector E are supplied, in this case, by the Box-Muller and McCarthy procedures (default options) while the choice of Nagar and inverse algorithms should be specified in the starting command (SS KLEIN1 NAGAR, or SS KLEIN1 INVERSE, or SS KLEIN1 NAGAR INVERSE). The resulting values for each endogenous variable, each period and each replication are stored into a work matrix, that is used to compute the statistics for the printout.

5. The final printout

For every year of the simulation period a printout of all the endogenous variables is displayed. It presents, for each variable, the actual (observed) value, the deterministic solution, the estimated mean, minimum and maximum of the stochastic solutions. The same information are presented on outputs per variable, together with the first relative differences of the actual, deterministic and mean stochastic values. Some other statistics and empirical indicators of goodness of fit are then computed. They are:
1) The mean over the simulation period of the actual, deterministic and mean stochastic values.
2) The Root Mean Square Error (RMSE) of the deterministic and mean stochastic solutions [9], [16].
3) The Mean Absolute Percentage Error (MAPE) of the deterministic and mean stochastic solutions [9].
4) Theil's inequality coefficient (U) of the deterministic and mean stochastic solutions [17].

5) The coefficients and standard errors of the regression (with intercept) of the observed values on the deterministic or mean stochastic solutions [17].
6) The coefficients and standard errors of the regression (without intercept) of the first relative differences of the observed values over those of the deterministic or mean stochastic solutions [17].
Tables 3 , 4 and 5 present some of these results for the simulation of the test model.

Table 3

KLEIN MODEL-1. STOCHASTIC SIMULATION
THREE-STAGE LEAST SQUARES ESTIMATES

MCCARTHY ALGORITHM. DYNAMIC SIMULATION. YEAR 1921. 100 REPLICATIONS

VAR	OBSERV	DETERM	MEAN.STOC	MIN	MAX	RANGE	STD.DEV
Y(1)	41.900	45.332	45.426	39.937	49.261	9.3237	2.0161
Y(2)	-.20000	1.9667	1.9076	-1.6517	6.1172	7.7690	1.5367
Y(3)	25.500	28.945	28.926	24.865	33.746	8.8815	1.7967
Y(4)	40.600	46.199	46.234	37.286	54.164	16.878	3.4476
Y(5)	12.400	14.554	14.608	9.3038	18.379	9.0751	1.8594
Y(6)	182.60	184.76	184.70	181.14	188.91	7.7690	1.5367

..........(the same for all the years up to 1941)..........

OUTPUT FOR VARIABLE Y(1) FROM YEAR 1921 TO 1941 WITH 100 REPLICAT.

YEAR	ACTUAL VALUE	DET VALUE	MEAN.STOC VALUE	STD DEV	MIN	MAX	ACTUAL %CHANG	DET %CHANG	MEAN.STOC %CHANG
1921	41.90	45.33	45.42	2.016	39.93	49.26	.0	.0	.0
1922	45.00	48.44	48.44	2.986	37.64	55.98	7.398	6.869	6.648
1923	49.20	51.88	51.70	3.430	44.03	60.60	9.333	7.106	6.723
1924	50.60	54.84	54.58	3.705	46.04	64.76	2.845	5.702	5.576
1925	52.60	56.27	56.14	3.352	47.85	63.53	3.952	2.597	2.849
1926	55.10	54.60	54.59	3.392	45.18	61.94	4.752	-2.968	-2.757
1927	56.20	51.06	51.18	3.660	41.11	59.49	1.996	-6.482	-6.246
1928	57.30	48.17	48.35	4.140	35.27	57.40	1.957	-5.663	-5.526
1929	57.80	48.35	48.82	3.838	40.74	58.08	.8726	.3909	.9600
1930	55.00	50.29	50.80	3.883	41.53	59.98	-4.844	3.999	4.064
1931	50.90	51.66	52.48	4.018	41.40	60.51	-7.454	2.716	3.301
1932	45.60	51.95	52.80	3.619	43.82	62.89	-10.41	.5625	.6210
1933	46.50	51.20	51.87	3.088	42.97	59.13	1.973	-1.441	-1.759
1934	48.70	52.28	52.54	3.483	44.68	61.14	4.731	2.110	1.288
1935	51.30	53.73	53.68	3.876	44.00	63.09	5.338	2.786	2.157
1936	57.70	55.21	54.80	4.113	43.79	64.36	12.47	2.743	2.103
1937	58.70	54.61	54.01	3.643	45.02	62.51	1.733	-1.088	-1.455
1938	57.50	57.48	57.14	3.521	45.13	65.38	-2.044	5.253	5.792
1939	61.60	60.88	60.45	3.601	50.99	68.70	7.130	5.924	5.794
1940	65.00	63.65	63.10	3.996	51.84	71.40	5.519	4.544	4.388
1941	69.70	69.06	69.10	3.844	60.48	77.62	7.230	8.493	9.511

.......(the same for all the other endogenous variables)......

Table 4

	MEAN VALUES			DIFFERENCES	
	ACTUAL	DETERMIN	MEAN STOCH.	DETERMIN	MEAN STOCH.
Y(1)	.53995D+02	.53858D+02	.53908D+02	-.13673D+00	-.86295D-01
Y(2)	.12666D+01	.11329D+01	.11171D+01	-.13372D+00	-.14954D+00
Y(3)	.36361D+02	.36233D+02	.36268D+02	-.12841D+00	-.93426D-01
Y(4)	.58371D+02	.58100D+02	.58135D+02	-.27045D+00	-.23584D+00
Y(5)	.16890D+02	.16748D+02	.16748D+02	-.14203D+00	-.14241D+00
Y(6)	.20176D+03	.20190D+03	.20224D+03	.14171D+00	.48526D+00

REGRESSION OF ACTUAL ON SIMULATED VALUES

	DETERMINISTIC			MEAN STOCHASTIC		
	EST.COEFF	STD.ERROR	T VALUE	EST.COEFF	STD.ERROR	T VALUE
VARIABLE Y(1)						
B0	.2606D+01	.9951D+01	.2619D+00	.1320D+01	.1047D+02	.1260D+00
B1	.9541D+00	.1838D+00	-.2494D+00	.9771D+00	.1934D+00	-.1183D+00

.......(the same for all the other endogenous variables)......

REGRESSION OF ACTUAL ON SIMULATED VALUES
(FIRST RELATIVE DIFFERENCES)

	DETERMINISTIC			MEAN STOCHASTIC		
	EST.COEFF	STD.ERROR	T VALUE	EST.COEFF	STD.ERROR	T VALUE
VARIABLE Y(1)						
B1	.5512D+00	.2751D+00	-.1631D+01	.4986D+00	.2783D+00	-.1800D+01

.......(the same for all the other endogenous variables)......

	R M S E (DIMENSIONLESS)		R M S E	
	DETERMIN.	MEAN STOCH.	DETERMIN.	MEAN STOCH.
Y(1)	0.7931D-01	0.8044D-01	0.4315D+01	0.4376D+01
Y(2)	0.8123D+00	0.8220D+00	0.2997D+01	0.3033D+01
Y(3)	0.1109D+00	0.1126D+00	0.4092D+01	0.4154D+01
Y(4)	0.1212D+00	0.1229D+00	0.7188D+01	0.7291D+01
Y(5)	0.1970D+00	0.2003D+00	0.3425D+01	0.3482D+01
Y(6)	0.2477D-01	0.2429D-01	0.5003D+01	0.4905D+01

M A P E

	DETERMINISTIC	MEAN STOCHASTIC
Y(1)	0.66976741D+01	0.69844656D+01
Y(2)		
Y(3)	0.89207088D+01	0.92469067D+01
Y(4)	0.10724729D+02	0.11063192D+02
Y(5)	0.19614092D+02	0.20112123D+02
Y(6)	0.19060701D+01	0.19504500D+01

Table 5

T H E I L INEQUALITY COEFFICIENTS

	DETERMINISTIC	MEAN STOCHASTIC
Y(1)	0.97014279D+00	0.10007834D+01
Y(2)	0.34593923D+02	0.25494341D+01
Y(3)	0.93261000D+00	0.94804300D+00
Y(4)	0.89684308D+00	0.92202912D+00
Y(5)	0.90269442D+00	0.94127179D+00
Y(6)	0.77930375D+00	0.78936792D+00

6. Performances

The package is written in FORTRAN IV and ASSEMBLER 370 languages. It consists of approximately one thousand statements, in addition to the statements necessary to formalize the model, as already seen in section 3. The required storage for the program is nearly 60 kilobytes. A large work matrix is then required to hold intermediate and final results of the computation; its dimensions depend on the parameters specified in the input file (number of equations, simulation period and number of replications): one or two megabytes are its currently used dimensions. The stochastic simulation of the presented model requires about 13 seconds of CPU time for 100 replications over the 21 years of the sample period.

7. References

[1] Box,G.E.P., and Muller,M.E., "A Note on the Generation of Random Normal Deviates", Ann. Math. Stat. 29, (1958).

[2] Calzolari,G., "Interactive Management for Time Series", COMPSTAT 1974, Proceedings in Computational Statistics, Phisica Verlag, Wien, (1974).

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

[4] Goldberger,A.S., "Impact Multipliers and Dynamic Properties of the Klein-Goldberger Model", North Holland, Amsterdam, (1970).

[5] Howrey,E.P., and Kelejian,H.H., "Simulation Versus Analytical Solutions: The Case of Econometric Models", in T.H.Naylor, ed., "Computer Simulation Experiments with Models of Economic Systems", John Wiley, New York, (1971).

[6] IBM, "System/360 Scientific Subroutine Package Version III Programmer's Manual H20-0205", IBM, New York, (1968).

[7] IBM, "Virtual Machine Facility/370: Command Language Guide for General Users GC20-1804", IBM, New York, (1972).

[8] Klein,L.R., "Economic Fluctuations in the United States", John Wiley, New York, (1950).

[9] Klein,L.R., "A Textbook of Econometrics", Prentice Hall, Englewood Cliffs, (1952).

[10] Kmenta,J., "Elements of Econometrics", Macmillan, New York, (1971).

[11] Lewis,P., Goodman,A., Miller,J., "A Pseudo-random Number Generator for the System/360", IBM Syst. J. 2, (1969).

[12] Lewis,P.A.W., Learmont,G.P., "Random Number Generator Package LLRANDOM PS55LW73061", Naval Postgraduate School, Monterey, (1973).

[13] McCarthy,M.D., "Some Notes on the Generation of Pseudo-structural Errors for Use in Stochastic Simulation Studies", in B.G. Hickman, ed., "Econometric Models of Cyclical Behaviour", Studies in Income and Wealth n.36, NBER, New York, (1972).

[14] Nagar,A.L., "Stochastic Simulation of the Brookings Econometric Model", in J.S.Duesenberry, et al., eds., "The Brookings Model: Some Further Results", North Holland, Amsterdam, (1969).

[15] Neave,H., "On Using the Box-Muller Transformation with Multiplicative Congruential Pseudo-random Numbers Generators", J. of the Royal Stat. Soc. (Series c) 22, (1973).

[16] Sowey,E.R., "Stochastic Simulation of Macroeconometric Models: Methodology and Interpretation", in A.A.Powell, and R.A.Williams, eds., "Econometric Studies of Macro and Monetary Relations", North Holland, Amsterdam, (1973).

[17] Theil,H., "Applied Economic Forecasting", North Holland, Amsterdam, (1966).