MONTE CARLO TESTS OF CONCLUSION ROBUSTNESS

BY

W G B PHILLIPS
Systems Analysis Research Unit
Department of the Environment,
London, U.K.

ABSTRACT

Conclusions derived from world models have little value if they do no include an estimate of the uncertainty associated with the outputs. This paper describes the Systems Analysis Research Unit World Model and gives an account of the application of Monte Carlo techniques to testing the model. Samples of uncertain data encoded in probability densities are used as input for model runs. The model output is analysed statistically and the contribution to total uncertainty by the variance of the inputs is determined. The output is also analysed by a multiple linear regression technique; variances are demonstrated to be additive over a limited range. Due to the strong negative feedback loops in the model, the model usually attenuates any variation in inputs. The cost of Monte Carlo methods is justified by the quality of the results obtained.

The view expressed in this paper are those of the author and do not necessarily coincide with those of the Department of the Environment.

TABLE OF CONTENTS

I. Introduction

The Monte Carlo technique is a method of assessing confidence in model conclusions, given the inherent imprecision in real-world structure and parameter values. The techniques permit the modeller to estimate the certainty of conclusions drawn from model output, within the very wide bounds set by his ignorance of the exact values to ascribe to model input. In a model-building project, the application of Monte Carlo techniques would normally take place at a late stage when the main structure of the model had been formulated. At this stage, the model has the status of an independent system which can be studied objectively, and may be treated in the same way that the communications engineer investigates a "black box". Assuming ignorance of the internal connections in the box, the evidence then depends solely on the inputs and outputs and the relationship between them. By varying the inputs and observing the effect on the outputs, the modeller can deduce the internal circuitry.* The structure deduced may, or may not, coincide with what actually exists inside the black box, but for our purposes it does not matter. The point is that the behaviour of the black box and the model of it will be the same, at least over a restricted range. This paper will describe a technique for estimating linear approximations to non-linear simulation models which appeals to the principle of equivalent behaviour. These approximations need be linear only in the empirical parameters, not the model parameters.

In a sense, this paper touches upon a higher level of abstraction. If a simulation model can be regarded as an abstraction from the real world, then a linear approximation to that simulation model becomes, as it were, a model of a model. The value and adequacy of such second-order models must be judged by the same criteria applied to first-order models. The utility of both orders is determined by how well they satisfy the objectives selected as important by the user.

*These techniques form the subject matter of several branches of engineering, see for example Jenkins and Watts.

World modellers must deal with formidable data problems. The amount of data being gathered by international agencies from all over the world is increasing, but it varies considerably in quality. The data are usually of a simple demographic or economic nature, such as the number of people living in a country or the quantity of iron ore exported. Information essential for world modelling purposes, such as the size of the capital stock, may never be collected, or collected under definitions quite unsuitable for our purposes. Such data might be collected if we had the resources of man-power, time and money but other essential facts can only be found by indirect methods. For example, parameters relating to the dynamic properties of models, such as the average length of time required to train a man to work in a new industry. The modeller might spend a great deal of time and effort getting one parameter value exactly right, and then find that the model results were insensitive to the value of that particular variable. However, modellers are seldom completely ignorant of the values which should be ascribed to their inputs. Certainly a given variable is either always negative or always positive. The modeller may be able to make a shrewed guess about a variable's mean value, or be practically certain that its value cannot exceed some number. Consider the problem of retraining labour. Who knows how long it takes on average to train a man for a new industry? But it cannot take a negative amount of time and surely it takes no more than twenty years. Though only a small amount, this information is not trivial; it suffices to define a probability density in the positive domain, such that 2.33 standard deviations are equivalent to about twenty years. The implication is that there is only one chance in a hundred that it would take longer than twenty years to retrain a man for a job. To exclude negative time constants, the modeller might choose a distribution such as the log normal*, which is defined only in the positive domain.

Given exact values with probability densities assigned to all input variables of the model, the information is sufficient to embark on a Monte Carlo investigation

*If X is normally distributed and can take values between $-\infty$ and $+\infty$ then if

$$X = \log_e Y$$

then Y is said to be log-normally distributed and can only take values between 0 and $+\infty$ (see Aitchison and Brown).

of the model's properties. The results may indicate that the behaviour of the model is particularly sensitive to one or more of the input variables. In that case, it will be worthwhile allocating further time and resources to refining the values of these particular items in the data set. This aspect of Monte Carlo technique has a close relationship to sensitivity testing.

II. An outline of Sarum

The Systems Analysis Research Unit has applied Monte Carlo techniques to test a simulation model (SARUM) of the world economy. For the model's purpose, the world is divided into three strata depending upon income per capita. This division defines a rich stratum, a medium stratum, and a poor stratum. Each stratus is divided into thirteen sectors which correspond to major industrial groupings. Since we are particularly interested in food problems, several of the sectors deal with agriculture. Each sector forms a viable sub-system within the whole; that is the sector is able to pay for its raw materials and factors of production and yield a surplus. The sectors buy and sell goods and services to one another, and also supply final demand. Production and consumption are mediated by a system of stocks, and a certain coverage level is always maintained by a system of negative feed back loops.

The state of the system at any given time is determined by a set of state variables. The behaviour of the system is governed by a set of equations which relate the rate of change of the state variables to the values of the state variables at any given time. These equations may be regarded as differential equations, and as such are soluble by standard integration methods. The output of the model is a set of trajectories for the state variables, or the solutions of the differential equations given the starting conditions. For the sake of realism and stability, some model variables are smoothed, that is to say, the variables do not assume their optimum value in the current time step but approach the desired value at a rate determined by the exogenously supplied time constants of the model.

III. Using Monte Carlo Techniques

a. Generating Input Data

The first step in the application of Monte Carlo techniques is to identify the items of input data which are open to uncertainty. In our pilot study the value of the time constants, the rate of growth of productivity, the rate of growth of the labour force, the depreciation rate, and the relationship between the supply of labour and the fraction of gross national product paid as wages were chosen. In each case, a reasonable value for the parameters could be deduced and this information was encoded in a set of normal and log normal probability densities. The mean and standard deviation reflected the average value and range over which the variable might change. For each run of the model, it was necessary to draw a number at random from each of these distributions. The computer library contains a random number generator that selects a number at random between zero and one. This uniform distribution is mapped on to a normal or log normal distribution by an appropriate transformation with the desired mean and standard deviation. The inputs to the model using this method are independent random variables with known statistical properties. A sequence of 700 runs of the model was performed using sets of input data generated as described. This process can clearly be expensive both in computer time and in the effort required to organize and process the large quantities of stored model outputs. Since SARUM is compact and efficient, it was possible to contain these costs and obtain a satisfactory return on the effort.

b. Analysis of output

The next step in using Monte Carlo is to compare the statistical properties of the outputs and with the statistical properties of the inputs. For the sake of concreteness, consider the output of one variable: production in the third model sector. If the model has been run for fifty years, the tape will contain five thousand observations of this variable in sets of one hundred at yearly intervals. The hundred observations corresponding to each year themselves form a probability density, which may be plotted in the form of a histogram*. The modeller could fit a theoretical distribution to this histogram, but the simplest method of description is in terms of its decile points, that is, the abcissae of the points which divide the total number of observations into ten equal subsets. According to this convention, 80% of the observations lie between the first and the ninth decile points. The median corresponds to the fifth decile point. The modeller can carry out this procedure for each of the fifty annual time steps, then plot the 10% and 90% points against time for each year. The resulting diagram (Figure 2) demonstrates how the uncertainty increases as the model moves into the future. After fifty years, the uncertainty amounts almost to an order of magnitude. However, this preliminary report on SARUM is probably far too pessimistic about the uncertainty in the input variables. The results, in any case, are only intended to illustrate the Monte Carlo technique.

By holding a subset of the variables constant while allowing the remainder to be drawn at random, the modeller can estimate how much of the uncertainty in the output is due to variations in a subset of the input. The three pairs of decile lines shown in Figure 2 correspond to:

1. the contribution to uncertainty of the variation in the time constants;

2. the variation due to the time constants and other variables less the uncertainty in the rate of growth of productivity and, finally

3. the uncertainty due to the total variation.

* See Figure 1. The area of the bars represents the number of events that fall within a given class interval.

DECILE POINTS FOR NON FOOD CONSUMER GOODS PRODUCTION: STRATUM 2, YEAR 50 EVERYTHING VARIED

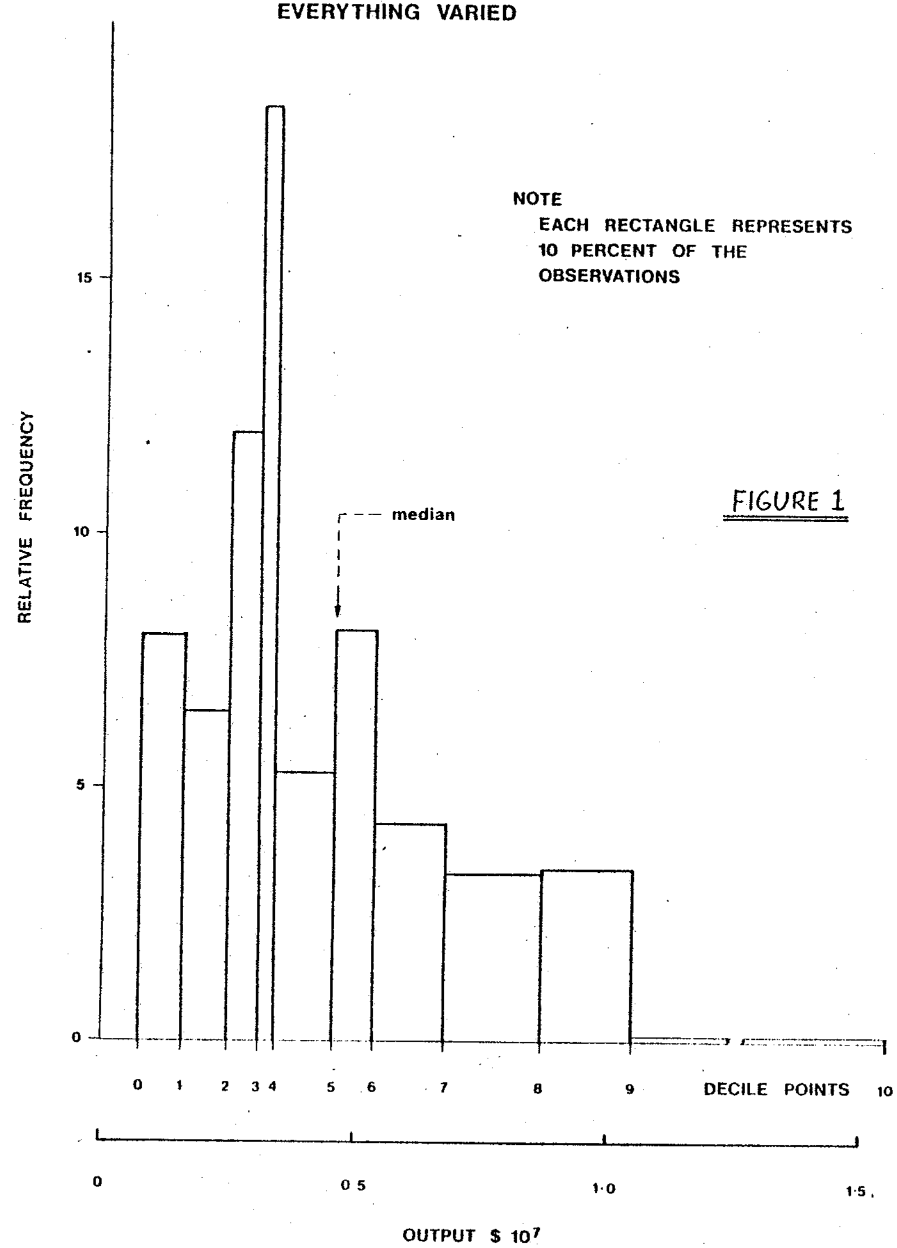

FIGURE 1

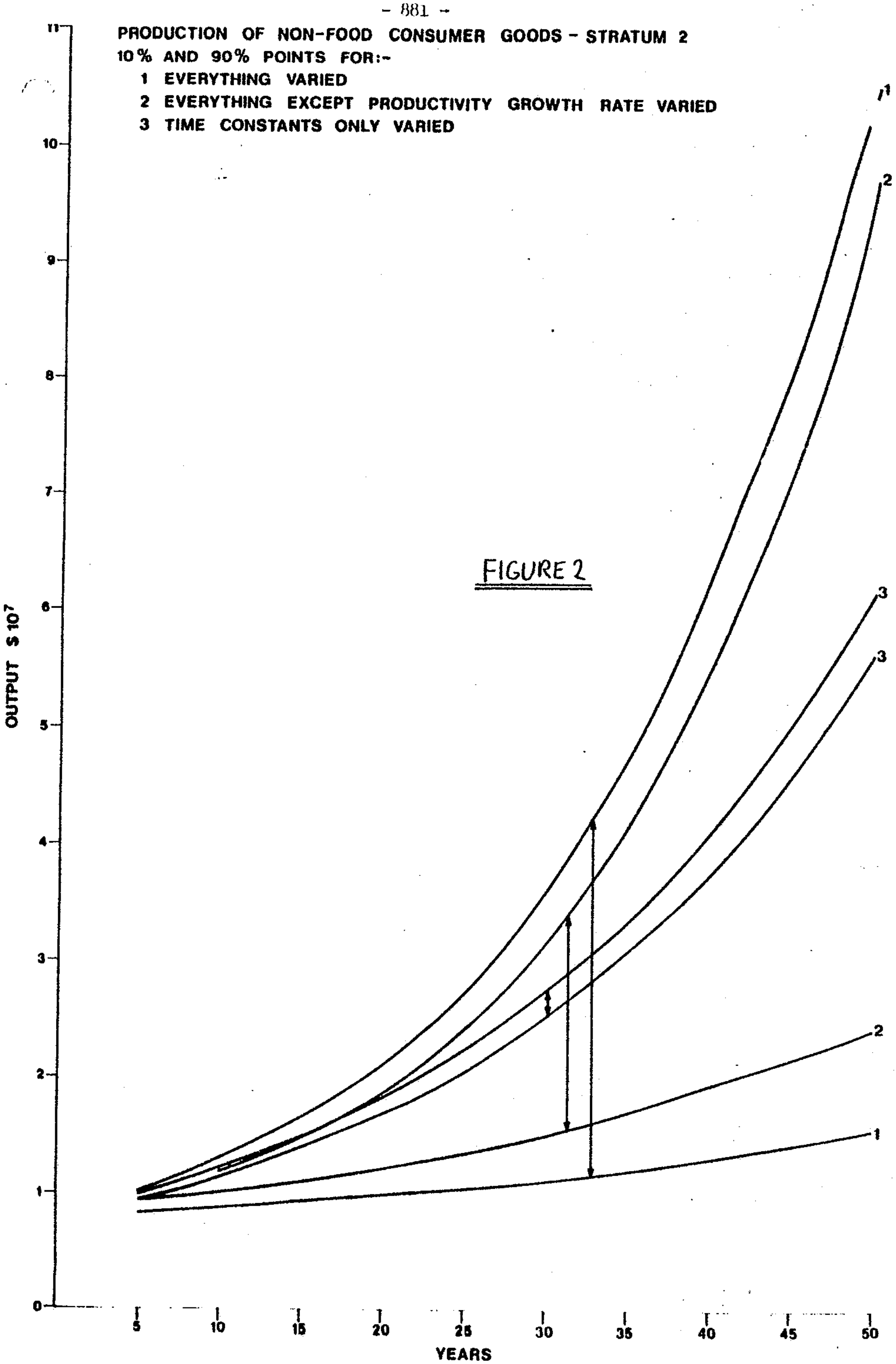

FIGURE 2

The uncertainty in the time constants has little effect on the uncertainty in the outputs. However, lacking exact information about the rate of growth of productivity or the rate of growth of population, little reliance can be placed on the model's predictions of future states of the world economy, as demonstrated by the wide divergence between the 10% and 90% decile lines labelled (3) in Figure 2.

c. Linear Regression Analysis

The Monte Carlo runs can be regarded as experiments the result of which is the value of the output variable. This result can be "explained" by the values assumed by the input variables, provided there are no internal stochastic processes. In the experimental situation, the results might be analysed by means of a multiple regression model. Moreover, since the modeller would probably be justified in assuming linearity over a narrow range, he might select a multiple linear regression model. This procedure, which can be justified over a narrow range by appeal to Taylor's theorem, has been implemented by means of the BMDO2R, STEPWISE multiple linear regression program. This program selects the explanatory variables yielding the highest correlation coefficient and calculates the coefficients and residuals for each regression. For example, expenditure per capita (EPC) has been regressed on the six time constants. The results showed that only three of the time constants were significant; that is, 99 per cent of the variation was explained by the labour mobility constant (TLM), the retraining time constant (TPR) and the consumption smoothing constant (TSM). This procedure yields several interesting results. The Taylor series expansion reveals that the regression coefficients correspond to the partial differential coefficients (for example, the rate of change of expenditure per capita with respect to the different time constants). Since this relationship is equivalent to the definition of a sensitivity, the procedure provides a completely general method of detecting the variables whose values must be exactly quantified. Conversely, variables with little effect on the model results can also be identified. However, the linearization is valid only at a given time step; the contribution of the variables might be different at the beginning of a run in a transient situation, or at the end where behaviour is affected by depletion. The results exhibited below refer to a situation of equilibrium growth. Moreover, if the sectors were changing rapidly in relative size, no doubt all the time constants would have proven significant.

Examples of Regression Analysis

Time constants varied:

The expenditure per capita has been regressed for stratum 1, year 50, on 6 time constants for a subset of 99 runs in which only these constraints were varied. The sample means and standard deviations are:

Variable name	Symbol	Mean	Standard deviation
Debt Erosion Constant	DEC	1.376	1.273
Labour Mobility time constant	TLM	2.627	2.240
Retraining time constant	TPR	17.833	17.128
Consumption smoothing constant	TSM	1.312	0.983
Trade fractions smoothing constant	TID	1.504	1.200
Production smoothing constant	TCF	6.590	6.616
Expenditure per Capita	EPC	18.395	0.744

The time constants are measured in years, while expenditure per capita (EPC) is measured in thousands of dollars per annum. A straightforward linear model gives an exceelent fit:

Multiple correlation	0.9936	
Regression SS	53.56	
Residual SS	0.69	
F ratio	1189.00	(6 on 92)

The constant term (since the absolute values of variables were used rather than their departures from the mean) was 19.279 and the coefficients (partial derivatives) were:

Variable	Coefficient	Standard Error	F to remove
DEC	0.00309	0.00697	0.1971
TLM	- 0.32369	0.00395	6726.3
TPR	- 0.00537	0.00052	106.07
TSM	0.05802	0.00989	34.41
TID	- 0.00432	0.00791	0.2982
TCF	- 0.00162	0.00146	1.243

Therefore, for all practical purposes, the effects of DEC, TID and TCF may be ignored.

Twelve input parameters varied

A regression was set up using the results from 737 model runs. Each of the 12 input variables was randomly sampled in at least some of the runs. A relationship between the output of the non-food consumer goods sector (year 50, stratum 1) and the inputs was sought. Previous regressions had shown that the output fluctuations were dominated by variations in productivity and labour growth rates (PI and GL). Since the model run simulated sustained growth (negative growth for some extreme input samples) and the two dominant variables appeared as exponents, it seemed likely that a linear relationship might hold between the logarithm of the sector output and the dominant inputs. This premise is supported by the residual plots for a straightforward linear regression. The sample means and standard deviations are:

Variable	Mean	Standard Deviation
DEC	1.413	1.332
TLM	2.907	2.671
TPR	16.105	13.866
TSM	1.364	1.212
TID	1.387	1.265
TCF	7.206	7.082
OMEGA	- 2.563	2.059
Y1	0.00244	0.0638
Y2	0.62557	0.0894
PI	0.01960	0.00874
GL	0.02065	0.01045
DEP	0.05673	0.03357
$\log_{10}(Q(3))$	6.417	0.2828

where Q(3) denotes the output of the non-food consumer goods sector.

The regression algorithm stopped with 11 of the 12 variables in the equation*. The 12th variable, OMEGA, had a partial correlation of only - 0.0007. The results were as follows:

Multiple correlation	0.9905	
Regression SS	57.75	
Residual SS	1.11	
F - ratio	3436.1	(11 on 725)
Constant	5.1143	

Variable	Coefficient	Standard Error	F to remove
DEC	- 0.00058	0.00109	0.287
TLM	- 0.00872	0.00054	259.9
TPR	- 0.00016	0.00011	2.37
TSM	- 0.00183	0.00120	2.31
TID	0.00018	0.00115	0.0253
TCF	- 0.00007	0.00021	0.1016
Y1	- 0.13367	0.02309	33.52
Y2	0.82618	0.01640	2536.8
PI	24.688	0.1665	21973.3
GL	15.733	0.1386	12881.2
DEP	0.15801	0.0435	13.2

*BMDO2R STEPWISE multiple linear regression program stops when the F-ratio of those variables not included in the regression falls below a value specified by the user.

There was some evidence of an increase in the size of the residuals at extreme GL values, indicating a possible small synergistic effect. The results are clearly dominated by uncertainty in the growth rates (PI and GL), the depreciation rate (DEP), and the ordinates of the labour supply curve (Y1 and Y2). However, the effects of the ordinates may reflect inconsistencies in the data set.

Local Linearisation

Suppose that, at a tiven time, the output variables are dependent on the input variables of the model according to a set of functions:

$$y_i = f_i(x_1, x_2, \ldots, x_n) \qquad ; i = 1,n$$

Then, in the neighbourhood of $x_1, x_2, \ldots, x_n$, a local linear approximation to y_i can be made by expanding $f_i(x_1 + \ x_1, x_2\Delta + \Delta x_2, \ldots, x_n + \Delta x_n)$ as a Taylor series about $f_i(x_1, x_2, \ldots, x_n)$, and then truncating all terms except the first. The Taylor series expansion of the ith variable is:

$$y_i = f_i(x_1+\Delta x_1, x_2+\Delta x_2, \ldots, x_n+\Delta x_n)$$

$$= f_i(x_1, x_2, \ldots, x_n) + \sum_j \Delta x_j \frac{\partial y_i}{\partial x_j} +$$

$$\frac{1}{2!}\left(\sum_j \Delta x_j^2 \frac{\partial^2 y_i}{\partial x_j^2} + 2 \sum_{\substack{j \\ j \neq k}} \sum_k \Delta x_j \Delta x_k \frac{\partial^2 y_i}{\partial x_j \partial x_k} \right) + \ldots$$

$$\therefore \Delta y_i \simeq \sum_j \frac{\partial y_i}{\partial x_j} \Delta x_j \qquad ; i = 1, \ldots, n$$

In matrix notation

$$\Delta \underline{y} = A \Delta \underline{x}$$

where A is the Jacobian for the y's;

that is $a_{ij} = \partial y_i / \partial x_j$

If S is the variance - covariance matrix for the x's, then V --- the variance - covariance matrix for the y's -- is given by

$$V = ASA^T \quad .$$

d. Additivity of variances

The linear regression procedure also sheds light on the so-called synergistic problem. A model or any other system for that matter, can often be shown to be stable with respect to variations in any of its input variables taken one at a time. However, the change often goes, if more than one variable were allowed to change simultaneously, the variance of the sum would be greater than the sum of the variances. In that case, the modeller could not ignore the interaction or covariance term. However, if the regression procedure establishes the validity of the linear hypothesis, then the variance of the output must be equal to the sum of the variances of the inputs weighted by the squares of the coefficients. Here again, the linear hypothesis probably only holds over a limited range, but some idea of the width of that range can be obtained by examining the plot of the residuals.

In algebraic terms, the regression procedure has established that the linear hypothesis holds over a certain range,

that is, $y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n$

and if $x_1, \dots, x_n$ are random variables

$$\text{var}\,(a_0 + a_1x_1 + a_2x_2 \dots + a_nx_n)$$

$$= \sum_i a_i^2 \,\text{var}\,(x_i) + \sum_i \sum_j a_ia_j \,\text{cov}(x_i, x_j) \quad .$$

We have generated $x_1, \dots, x_n$ so that they are known to be independent, which implies that the covariance vanishes and therefore

$$\text{var}(y) = \sum_i x_i^2 \,\text{var}(x_i) \quad .$$

The modeller can therefore use the regression model to predict the variance of a dependent variable in light of any assumption taken about the variances of the independent variables.

The regression analysis showed that:

$$EPC = a_0 + a_1 \Delta\, TLM + a_2 \Delta\, TPR + a_3 \Delta\, TSM$$

therefore:

$$\text{Var}\,(EPC) = a_1^2 \,\text{var}\,(TLM) + a_2^2 \,\text{var}\,(TPR) + a_3^2 \,\text{var}\,(TSM) \quad .$$

Inserting the known values of the coefficients and the variables on the right-hand side gives:

a_1^2 var (TLM)	=	.52436
a_2^2 var (TPR)	=	.00858
a_3^2 var (TSM)	=	.00276
		.53567
Mean square residual error	=	.00740
		.54307 .

This figure agrees closely with the known variance of EPC:

$$\text{var}\,(EPC) = (.744)^2 = .553 \quad .$$

Because the contribution made by the labour mobility constant to the total variance is large, it would be desirable to undertake further studies to establish more exact values for this parameter. The modeller would then be able to assign TLM a smaller variance, and could immediately deduce how much effect this variance would have on the uncertainty in EPC without having to carry out any further Monte Carlo runs. With limits, therefore, the modeller need not be too particular about the variances ascribed to the independent variables in preliminary studies, since these figures can be revised without cost at a later date.

e. Amplification

The procedure discussed here also facilitates a comparison of the variation of model inputs with the variation of outputs, and a subsequent calculation of whether the model amplifies or attenuates errors in the data. The amplification is defined as the ratio of the fractional input error to the resulting fractional output error. For the model time constants, the amplification turns out to be much less than 1.0 (that is, they are attenuated), as would be expected since the behaviour of the model is dominated by negative feedback loops. We can make an interesting analogy with amplifier design at this point. To obtain good reproduction, an audio-amplifier must have linear response, but it must be constructed of transistors which have very non-linear characteristics. To overcome this problem, modern amplifiers are constructed with strong negative feedback so that, while they may be inefficient, their overall response can be made almost exactly linear over a wide range.

The error amplification of TLM is calculated as follows:

the fractional input error for TLM is

$$\frac{\sigma\ (TLM)}{TLM} \quad .$$

The resulting fractional output error is

$$\frac{a_{TLM} \cdot \sigma\ (TLM)}{EPC} \quad .$$

The amplification is therefore

$$\frac{a_{TLM} \cdot TLM}{EPC} \quad .$$

For the first example in which only the time constants were varied we have:

Variable	Amplification
DEC	0.00023
TLM	0.0462
TPR	0.00521
TSM	0.00414
TID	0.00035
TCF	0.00058

Errors in the time constants are therefore highly attenuated in the model.

IV. Conclusions

In a scientific context, a measurement made without an estimate of its variance has very little value. Insufficient attention has been paid to this problem by world modellers. They have often offered point estimates or trajectories without indicating how much uncertainty or fuzziness surrounds them. The methods described here go some way toward providing the necessary tools to set confidence intervals* for model projections.

These techniques are expensive, but they permit the modeller to extract a great deal of information about model structure. However, if the modeller knows

*A confidence interval is constructed in the following way: if two limits, say α_1 and α_2 can be found such that if we define the logical statement B by:
B = "α lies between α_1 and α_2"
then if we can find a number c, such that $P(B) \geq 1 - c$
then c is the confidence coefficient.

that the empirical parameters exhibit linearities over some specified range, then the uncertainty of the output (its variance) can be computed from the variances of the inputs. This set of variances may differ from those used in the original Monte Carlo runs. Consequently, the modeller can obtain the equivalent of an immense number of possible Monte Carlo runs from just one set of runs (see Section d.).

There seems to be a strong argument here for parsimony in model building; for a model that is simple and cheap enough to be run many hundreds of times to explore its variational properties in full. Modellers may be tempted to build models of overwhelming comprehensiveness because the relationships seem so complex and manifold. However, unless it can be fully tested, such a model may be less helpful than a smaller transparent model whose relationships are thoroughly understood. In future, however, the decline in computing costs may make possible the use of Monte Carlo techniques on even the largest and most complex models.

V. Acknowledgement

I should like to take the opportunity to acknowledge the helpful advice that I have received from my colleagues in the Systems Analysis Research Unit and in particular the valuable assistance given by Dr K T Parker and Dr M R Goodfellow who designed the tests and made the computations.

VI. References

CHARLWOOD, F. J. and NOTON, M.
"Uncertainty in World Modelling"
Department of System Science, City University, London (Unpublished).

AITCHISON, J. and BROWN, J. A. C. (1957) "The Lognormal Distribution", Cambridge: Cambridge University Press, 1957.

JENKINS, G. M. and WATTS, D. G. "Spectral Analysis and its applications". Holden-Day Inc., San Francisco, 1968.