80-555 #5

STANDARD FORM 162 (REV 5 - 78)
Prescribed by Library of Congress Interlibrary Loan Code

INTERLIBRARY LOAN REQUEST

Date of request: Not needed after: Requester's order no.

Request for ☐ LOAN or ☐ PHOTOCOPY
According to the A.L.A. Interlibrary Loan Code and the Federal Library Committee's I.L. Code

Call No.

U. S. NUCLEAR REGULATORY COMMISSION
LIBRARY - INTERLIBRARY LOAN
WASHINGTON, D.C. 20555
STOP 565

D

NOTICE OF RETURN
RETURN SEPARATELY

REPORTS: Checked by
SENT BY: ☐ Library rate ☐
Charges $ Insured for $
Date sent
DUE

RESTRICTIONS: ☐ For use in library only.
☐ Copying not permitted ☐

FOR YOUR RETENTION

For use of Status Dept.

Book author: OR: periodical title, vol. and date

NOT SENT BECAUSE: ☐ In use
☐ Not Owned
☐ Non Circulating
☐ Request of

Book title, edition, place, year, series: OR: periodical article author, title, pages. ☐ This edition only

Estimated Cost of: ☐ Microfilm
☐ Hard copy

BORROWING LIBRARY RECORD:
Date received
Date returned
By ☐ Library rate ☐
Postage enclosed $ Insured for $

Verified in: OR: item cited in
ISBN, or ISSN, of LC card, or OCLC, or other number if known
If non-circulating, & cost does not exceed $, please supply ☐ Microfilm ☐ Hard copy

RENEWALS: ☐ No renewals
Requested on
Renewed to
(or period of renewal)

Note: the receiving library assumes responsibility for notification of non-receipt.

Request complies with
☐ *108(g)(2) Guidelines (CCG)*
☐ *other provisions of copyright law (CCL)*

AUTHORIZED BY: (FULL NAME)
Title

162-102

N.D.

VERIFICATION OF COMPUTER SIMULATION MODELS*†

THOMAS H. NAYLOR AND J. M. FINGER

Duke University

The problem of validating computer simulation models of industrial systems has received only limited attention in the management science literature. The purpose of this paper is to consider the problem of validating computer models in the light of contemporary thought in the fields of philosophy of science, economic theory, and statistics. In order to achieve this goal we have attempted to gather together and present some of the ideas of scientific philosophers, economists, statisticians, and practitioners in the field of simulation which are relevant to the problem of verifying simulation models. We have paid particular attention to the writings of economists who have been concerned with testing the validity of economic models. Among the questions which we shall consider are included: What does it mean to verify a computer model of an industrial system? Are there any differences between the verification of computer models and the verification of other types of models? If so, what are some of these differences? Also considered are a number of measures and techniques for testing the "goodness of fit" of time series generated by computer models to observed historical time series.

The Problem of Verification

In discussing procedures and techniques used in designing computer simulation experiments with industrial systems, management scientists have had very little to say about how one goes about "verifying" a simulation model or the data generated by such a model on a digital computer. In part, the reason for avoiding the subject of verification stems from the fact that the problem of verifying or validating computer models remains today perhaps the most elusive of all the unresolved methodological problems associated with computer simulation techniques. Yet we know very well that "verifiability is a necessary constituent of the theory of meaning. A sentence the truth of which cannot be determined from possible observations is meaningless" [23, pp. 256–257].

Likewise, simulation models based on purely hypothetical functional relationships and contrived data which have not been subjected to empirical verification are void of meaning. Regarding computer models of management control systems which have not been validated, Clay Sprowls has said that "I am prepared to look at each of them as an interesting isolated case which can be described to me but from which I shall draw no conclusions" [26, p. 148]. Although the construction and analysis of a simulation model, the validity of which has not been ascertained by empirical observation, may prove to be of interest for expository

* Received April 196[illegible]; revised May 19[illegible].

† This research was supported by National Science Foundation Grant GS-110[illegible] and a grant from the Duke University Research Council and is part of a collection of studies entitled "Design of Computer Simulation Experiments for Economic Systems." We are indebted to W. Earl Sasser of the Econometric System Simulation Program at Duke University for a number of helpful comments on the original draft.

or pedagogical purposes (e.g., to illustrate particular simulation techniques), such a model contributes nothing to the understanding of the system being simulated.

To verify or validate any kind of model (e.g., management science models) means to prove the model to be true. But to prove that a model is "true" implies (1) that we have established a set of criteria for differentiating between those models which are "true" and those which are "not true," and (2) that we have the ability to apply these criteria to any given model. In view of the difficulty which arises in attempting to agree upon a set of criteria for establishing when a model is verified, Karl R. Popper [22] has suggested that we concentrate on the degree of confirmation of a model rather than whether or not the model has been verified. "If in a series of empirical tests of a model no negative results are found but the number of positive instances increases then our confidence in the model will grow step by step. Thus, instead of verification, we may speak of gradually increasing confirmation of the law" [3, [illegible]].

The rules for validating computer simulation models and the data generated by these models are sampling rules resting entirely on the theory of probability. Both the simulation models which have been programmed into a computer and the data which have been generated from these models represent the essence of inductive reasoning, for they are the joint conclusions of a set of inductive inferences (behavioral assumptions or operating characteristics) about the behavior of a given system. The validity of a model is made probable, not certain, by the assumptions underlying the model; the inductive inference must be conceived as an operation belonging in the calculus of probability [22, p. 233].

In the following section we explore three major methodological positions concerning the problem of verification in economics which are relevant to the problem of verifying computer models of industrial systems.

Three Positions on Verification

Rationalism

Rationalism holds that a model or theory is simply a system of logical deductions from a series of synthetic premises of unquestionable truth not themselves open to empirical verification or general appeal to objective experience [illegible] p. 612]. Immanuel Kant (1724–1804), who believed that such premises exist, coined the term synthetic a priori to describe premises of this type. The [illegible] arguments in support of rationalism in economics have been outlined as follows:

> [illegible]

Unfortunately, any attempt to spell out literally and in detail all of the basic assumptions underlying a particular system soon reveals limitations to their obviousness [16, p. 136]. Reichenbach goes so far as to deny the very existence of a synthetic a priori:

> Scientific philosophy . . . refuses to accept any knowledge of the physical world as absolutely certain. Neither the individual occurrences, nor the laws controlling them, can be stated with certainty. The principles of logic and mathematics represent the only domain in which certainty is attainable; but these principles are analytic and empty. Certainty is inseparable from emptiness; there is no synthetic a priori [23, p. 304].

Empiricism

At the other end of the methodological spectrum in complete opposition to rationalism is *empiricism*. Empiricists regard empirical science, and not mathematics, as the ideal form of knowledge. "They insist that sense observation is the primary source and the ultimate judge of knowledge, and that it is self-deception to believe the human mind to have direct access to any kind of truth other than that of empty logical relations" [23, pp. 73–74]. Empiricism refuses to admit any postulates or assumptions that cannot be independently verified. This extreme form of logical positivism asks that we begin with facts, not assumptions [1, pp. 612–613].

T. W. Hutchison, a leading proponent of empiricism as a means of verification in economics, has said that "'propositions of pure theory' is a name for those propositions not conceivably falsifiable empirically and which do not exclude or 'forbid' any conceivable occurrence, and which are therefore devoid of empirical content, being concerned with language" [14, p. 161]. Continuing, Hutchison added that "Propositions of pure theory, by themselves, have no prognostic value or causal significance" [14, p. 162].

However, Blaug suggests that throughout the history of economic thought some economists have been willing to compromise on these two extreme points of view—synthetic a priorism and empiricism. The controversy is over matters of emphasis, and economists have always occupied the middle ground between extreme a priorism and empiricism [1, pp. 612–613].

Positive Economics

Milton Friedman argues that critics of economic theory have missed the point by their preoccupation with the validity of the assumptions of models. According to Friedman the validity of a model depends not on the validity of the assumptions on which the model rests (as Hutchison would have one believe), but rather on the ability of the model to predict the behavior of the dependent variables which are treated by the model.

> [illegible]

test of the validity of the hypothesis different from or additional to the test by implications. This widely held view is fundamentally wrong and productive of much mischief. Far from providing an easier means for sifting valid from invalid hypotheses, it only confuses the issue, promotes misunderstanding about the significance of empirical evidence for economic theory, produces a misdirection of much intellectual effort devoted to the development of [illegible] or tentative hypotheses in positive economics [illegible, p. 14].

Although the notion that conformity to observed behavior is a desirable check on the validity of an economic model is indeed an appealing methodological position, Friedman has by no means escaped criticism for maintaining such a position. "Friedman's position is unassailable until it is realized that he is insisting on empirical testing of predictions as the sole criterion of validity. He seems to be saying that it makes no difference whatever to what extent the assumptions falsify reality" [[illegible], pp. 612–613].

Critics of Friedman's brand of positive economics as applied to [illegible] by [illegible] of predictions argue that to state a set of assumptions, and then to exempt a subclass of their implications from verification is a curiously roundabout way of specifying the content of a theory that is regarded as open to empirical refutation. "It leaves one without an understanding of the reasons for the exceptions" [16, p. 139].

Multi-Stage Verification

Computer simulation suggests yet a fourth possible approach to the problem of verification—multi-stage verification. This approach to verification is a [illegible] stage procedure incorporating the methodology of rationalism, empiricism, and positive economics. Multi-stage verification implies that each of the aforementioned methodological positions is a necessary procedure for validating simulation experiments but that neither of them is a sufficient procedure for solving the problem of verification. Although multi-stage verification may be applicable to the verification of models in general, we shall argue in this section that multi-stage verification is particularly applicable to the verification of computer simulation models of industrial systems.

The first stage of this procedure calls for the formulation of a set of postulates or hypotheses describing the behavior of the system of interest. To be sure, these are not just any postulates, for what is required in stage one is a diligent search for Kant's "synthetic a priori" using all possible information at our disposal.

[illegible] can do nothing but look for his basic postulates. But this is what he can do, and he is willing to do it with the [illegible] [illegible] [illegible] [illegible] [illegible] [illegible] [illegible]

We [illegible]

[illegible]

cannot subject all possible postulates to formal empirical testing and must therefore select, on essentially a priori grounds, a limited number of postulates for further detailed study. He is, of course, at the same time rejecting an infinity of postulates on the same grounds. The selection of postulates is taken here to include the specification of components and the selection of variables as well as the formulation of functional relationships. But having arrived at a set of basic postulates on which to build our simulation model, we are not willing to assume that these postulates are of such a nature as to require no further validation. Instead we merely submit these postulates as tentative hypotheses about the behavior of a system.

The second stage of our multi-stage verification procedure calls for an attempt on the part of the analyst to "verify" the postulates on which the model is based subject to the limitations of existing statistical tests. Although we cannot solve the philosophical problem of "what does it mean to verify a postulate?," we can apply the "best" available statistical tests to these postulates.

But in management science we often find that many of our postulates are either impossible to falsify by empirical evidence or extremely difficult to subject to empirical testing. In these cases we have two choices. We may either abandon the postulates entirely, arguing that they are scientifically meaningless since they cannot be conceivably falsified, or we may retain the postulates merely as "tentative" postulates. If we choose the first alternative we must continue searching for other postulates which can be subjected to empirical testing. However we may elect to retain these "tentative" postulates which cannot be falsified empirically on the basis that there is no reason to assume that they are invalid just because they cannot be tested.

The third stage of this verification procedure consists of testing the model's ability to predict the behavior of the system under study. C. West Churchman states flatly that the purpose of simulation is to predict, and considers the point so obvious that he offers no defense of it before he incorporates it into his discussion of the concept of simulation [4]. This point does indeed seem obvious. Unless the construction of simulation models is viewed as a game with no purpose other than the formulation of a model, it is hard to escape the conclusion that the purpose of a simulation experiment is to predict some aspect of reality.

In order to test the degree to which data generated by computer simulation models conform to observed data, two alternatives are available—historical verification and verification by forecasting. The essence of these procedures is prediction, for historical verification is concerned with retrospective predictions while forecasting is concerned with prospective predictions.

If one uses a simulation model for descriptive analysis, he is interested in the behavior of the system being simulated and so would attempt to produce a model which would predict that behavior. The use of simulation models for prescriptive purposes involves predicting the behavior of the system [illegible] different combinations of policy conditions. The [illegible] would then decide on the most desirable set of policy conditions to put into effect by [illegible] predicted [illegible] When [illegible] is used for descriptive analysis the actual historical record produced by the system being simulated can be used as a check on the accuracy of the predictions, and hence on the extent to which the model fulfills its purpose. But a prescriptive analysis involves choosing one historical path along which the system will be directed. Hence only the historical record of the path actually traveled will be generated and the historical records of alternative paths corresponding to alternative policies will not be available for comparison. Though in this case the historical record cannot be used as a direct check on whether or not the model did actually point out the best policy to follow, the actual outcome of the policy chosen can be compared with the outcome predicted by the simulation model as an indirect test of the model. In either case, the predictions of the model are directly related to the purpose for which the model was formulated, while the assumptions which make up the model are only indirectly related to its purpose through their influence on the predictions. Hence the final decision concerning the validity of the model must be based on its predictions.

Goodness of Fit

Thus far, we have concerned ourselves only with the philosophical aspects of the problem of verifying computer simulation models. What are some of the practical considerations which the management scientist faces in verifying computer models? Some criteria must be devised to indicate when the time paths generated by a computer simulation model agree sufficiently with the observed or historical time paths so that agreement cannot be attributed merely to chance. Specific measures and techniques must be considered for testing the "goodness of fit" of simulation models, i.e., the degree of conformity of simulated time series to observed data. Richard M. Cyert has suggested that the following measures might be appropriate [10]:

1. number of turning points,
2. timing of turning points,
3. direction of turning points,
4. amplitude of the fluctuations for corresponding time segments,
5. average amplitude over the whole series,
6. simultaneity of turning points for different variables,
7. average values of variables,
8. exact matching of values of variables.

To this list of measures we would add the probability distribution and variation about the mean (variance, skewness, kurtosis) of variables.

Although a number of statistical techniques exist for testing the "goodness of fit" of simulation models, for some unknown reason management scientists and economists have, more often than not, restricted themselves to purely graphical (as opposed to statistical) techniques of "goodness of fit" for validating computer models [5], [19]. The following statement by Cyert and March concerning the validity of their duopoly model is indicative of the lack of emphasis placed on goodness of fit by many practitioners in this field.

[illegible]

This statement was made on the basis of a graphical comparison of the simulated time series and actual data. Not unlike most other simulation studies described in the literature, Cyert and March did not pursue the question of verification beyond the point described in the aforementioned statement.

Within the confines of this paper it is impossible to enumerate all of the statistical techniques which are available for testing the "goodness of fit" of simulation models. However, we shall list some of the more important ones and suggest a number of references which describe these tests in detail.

1. *Analysis of Variance.* The analysis of variance is a collection of techniques for data analysis which can be used to test the hypothesis that the mean (or variance) of a series generated by a computer simulation experiment is equal to the mean (or variance) of the corresponding observed series. Three important assumptions underlie the use of this technique—normality, statistical independence, and a common variance. The paper by Naylor, Wertz, and Wonnacott [20] describes the use of the F-test, multiple comparisons, and multiple ranking procedures to analyze data generated by simulation experiments.

2. *Chi-Square Test.* The Chi-square test is a classical statistical test which can be used for testing the hypothesis that the set of data generated by a simulation model has the same frequency distribution as a set of observed historical data. Although this test is relatively easy to apply, it has the problem of all tests using categorical type data, namely, the problem of selecting categories in a suitable and unbiased fashion. It has the further disadvantage that it is relatively sensitive to non-normality.

3. *Factor Analysis.* Cohen and Cyert have suggested the performance of a factor analysis on the set of time paths generated by a computer model, a second factor analysis on the set of observed time paths, and a test of whether the two groups of factor loadings are significantly different from each other [6].

4. *Kolmogorov-Smirnov Test.* The Kolmogorov-Smirnov test is a *distribution-free* (nonparametric) test concerned with the degree of agreement between the distribution of a set of sample values (simulated series) and some specified theoretical distribution (distribution of actual data). The test involves specifying the cumulative frequency distribution of the simulated and actual data. It treats individual observations separately and unlike the Chi-square test, does not lose information through the combining of categories [25].

5. *Nonparametric Tests.* The books by Siegel [25] and Walsh [29] describe a host of other nonparametric tests which can be used for testing the "goodness of fit" of simulated data to real world data.

6. *Regression Analysis.* Cohen and Cyert have also suggested the possibility of regressing actual series on the generated series and testing whether the resulting regression equations have intercepts which are not significantly different from zero and slopes which are not significantly different from unity [6].

7. *Spectral Analysis.* Data generated by computer simulation experiments are usually highly autocorrelated. When autocorrelation is present in sample data, the use of classical statistical estimating techniques (which assume the absence of autocorrelation) will [illegible] to underestimates of sampling variances [illegible] unduly large [illegible] Spectral analysis [illegible]

arranged in a series according to historical time. It is essentially the quantification and evaluation of autocorrelated data at which spectral analysis is aimed, after the data have been transformed into the frequency domain. For purposes of describing the behavior of a stochastic variate over time, the information content of spectral analysis is greater than that of sample means and variances. Spectral analysis provides a means of objectively comparing time series generated by a computer model with observed time series. By comparing the estimated spectra of simulated data and corresponding real-world data, one can infer how well the simulation resembles the system it was designed to emulate [12], [20], [21].

8. *Theil's Inequality Coefficient.* A technique developed by Theil has been used by a number of economists to validate simulations with econometric models [28]. Theil's inequality coefficient U provides an index which measures the degree to which a simulation model provides retrospective predictions of observed historical data. U varies between 0 and 1. If $U = 0$, we have perfect predictions. If $U = 1$, we have very bad predictions. There is no obvious reason why this technique cannot be used to validate management science models as well as econometric models.

Summary

While we *have* argued that the success or failure of a simulation experiment must be measured by how well the model developed predicts the particular phenomena in question, we *have not* argued that care exercised in selecting assumptions and statistical testing of these assumptions are purposeless or wasteful activities. Our defense of the first two stages of the three-stage process of verification we have proposed rests solidly on the law of scarcity. Any hypothesis which can be rejected on *a priori* grounds should be so rejected, because testing by this procedure is *cheaper* than formal statistical testing. Only if the experimenter had an unlimited budget could he afford to subject *all* possible hypotheses to statistical testing. Likewise, testing assumptions is cheaper than deriving and testing predictions, so any increase of validity we can obtain at an early stage is cheaper than additional validity gained at a later stage.

Having described multi-stage verification it is appropriate that we point out that this approach to verification is by no means limited to simulation models. For example, suppose that we were interested in verifying a simple economic model of consumer demand for a particular commodity, the model consisting of one or two equations. First, we might take a look at the rationale or the a priori assumptions underlying the model. These assumptions might take the form of postulates about the shape of individual marginal utility functions, the signs and magnitudes of income and substitution effects, the shape of indifference curves, etc. Are these assumptions in accordance with the body of knowledge known as economic theory? Second, if we are satisfied with the model on purely a priori grounds, then we attempt to verify ... [illegible]

[illegible]

to bypass the first two steps of the multi-stage verification procedure and concentrate on the accuracy of the model's predictions. Whether we would be willing to skip steps one and two in verifying a particular model will, in part, depend on the cost of obtaining predictions with our model. If the model is characterized by (1) a small number of variables, (2) a small number of linear equations, (3) no stochastic variables, and (4) predictions for only one or two time periods, then one may be willing to concentrate on the third step of the procedure with a minimum of risk. But if one is dealing with a complex model consisting of a large number of nonlinear difference or differential equations and a large number of variables (some variables being stochastic), and the model is to be used to generate time paths of endogenous variables over extended periods of time, then the cost of omitting steps one and two of our procedure might be quite high. Thus, it may be prudent to use steps one and two of the multi-stage procedure to detect errors in the model which otherwise might not become obvious until expensive computer runs have been made.

Thus, while most of our argument is relevant to the general problem of verifying hypotheses or theories, the nature of computer simulation experiments makes the three stage procedure particularly relevant to computer simulation models. This form of analysis is particularly useful when: (1) it is extremely costly or impossible to observe the real world processes which one is attempting to study, or (2) the observed system is so complex that it cannot be described by a set of equations for which it is possible to obtain analytical solutions which could be used for predictive purposes. Thus computer simulation is a more appropriate tool of analysis than techniques such as mathematical programming or marginal analysis when data against which predictions can be tested are not available and/or when predictions can be obtained only at great expense (in human time and/or computer time). In other words, computer simulation is most likely to be utilized when the savings derived from improving the model at earlier stages are most pronounced.

References

1. Blaug, M., *Economic Theory in Retrospect*, Richard D. Irwin, Homewood, Ill., 1962.
2. Burdick, Donald S. and Naylor, Thomas H., "Design of Computer Simulation Experiments for Industrial Systems," *Communications of the ACM*, IX (May, 1966), pp. 329-339.
3. Carnap, R., "Testability and Meaning," *Philosophy of Science*, III (1936), [illegible]
4. Churchman, C. West, "An Analysis of the Concept of Simulation," *Symposium on Simulation Models*, Austin C. Hoggatt and Frederick E. Balderston (editors), South-Western Publishing Co., Cincinnati, 1963.
5. Cohen, K. J., *Computer Models of the Shoe, Leather, Hide Sequence*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1960.
6. Cohen, Kalman J. and Cyert, Richard M., "Computer Models in Dynamic Economics," *The Quarterly Journal of Economics*, LXXV (February, 1961), pp. 112-127.
7. Conway, R. W., "Some Tactical Problems in Digital Simulation," [illegible] Vol. 10, No. 1 (October, 1963), pp. 47-[illegible]
[illegible]

10. Cyert, Richard M., "A Description and Evaluation of Some Firm Simulations," *Proceedings of the IBM Scientific Computing Symposium on Simulation Models and Gaming*, IBM, White Plains, N.Y., 1966.
11. ——, and March, James G., *A Behavioral Theory of the Firm*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1963.
12. Fishman, George S., and Kiviat, Philip J., "The Analysis of Simulation-Generated Time Series," *Management Science*, Vol. 13, No. 7 (March 1967), pp. 525-557.
13. Friedman, Milton, *Essays in Positive Economics*, Univ. of Chicago Press, 1953.
14. Hutchison, T. W., *The Significance and Basic Postulates of Economic Theory*, Macmillan & Co., London, 1938.
15. Kuhn, H. P., and Stover, R. N., "Simulation of an Industrial Environment," *Proceedings of the IBM Scientific Computing Symposium on Simulation Models and Gaming*, IBM, White Plains, N.Y., 1966.
16. Koopmans, Tjalling C., *Three Essays on the State of Economic Science*, McGraw-Hill Book Co., New York, 1957.
17. McMillan, Claude, and Gonzalez, Richard F., *Systems Analysis*, Richard D. Irwin, Inc., Homewood, Ill., 1965.
18. Naylor, Thomas H., Balintfy, Joseph L., Burdick, Donald S., and Chu, Kong, *Computer Simulation Techniques*, John Wiley & Sons, New York, 1966.
19. ——, Wallace, William H., and Sasser, W. Earl, "A Computer Simulation Model of the Textile Industry," Working Paper No. 8, Econometric System Simulation Program, Duke University, October 15, 1966.
20. ——, Wertz, Kenneth, and Wonnacott, Thomas, "Some Methods for Analyzing Data Generated by Computer Simulation Experiments," *Communications of the ACM* (1967).
21. ——, Wertz, Kenneth, and Wonnacott, Thomas, "Spectral Analysis of Data Generated by Simulation Experiments with Econometric Models," Working Paper No. 1, Econometric System Simulation Program, Duke University, September 1, 1966.
22. Popper, Karl R., *The Logic of Scientific Discovery*, Basic Books, New York, 1959.
23. Reichenbach, Hans, *The Rise of Scientific Philosophy*, University of California Press, Berkeley, 1951.
24. Robbins, Lionel, *An Essay on the Nature and Significance of Economic Science*, Macmillan, London, 1935.
25. Siegel, Sidney, *Nonparametric Statistics*, McGraw-Hill, New York, 1956.
26. Sprowls, C. R., "Simulation and Management Control," *Management Controls: New Directions in Basic Research*, C. P. Bonini et al. (editors), McGraw-Hill, New York, 1964.
27. Teichroew, Daniel, and Lubin, John F., "Computer Simulation—Discussion of the Technique and Comparison of Languages," *Communications of the ACM*, IX (October 1966), pp. 723-741.
28. Theil, H., *Economic Forecasts and Policy*, North-Holland Publishing Co., Amsterdam, 1961.
29. Walsh, John E., *Handbook of Nonparametric Statistics*, I and II, D. Van Nostrand Co., Princeton, N.J., 1962, 1965.