WORKING PAPER
ALFRED P. SLOAN SCHOOL OF MANAGEMENT

ALTERNATIVE TESTS FOR THE SELECTION
OF MODEL VARIABLES

Nathaniel J. Mass*
Peter M. Senge**

WP 828-76 January 1976

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
50 MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02139
ALTERNATIVE TESTS FOR THE SELECTION
OF MODEL VARIABLES

Nathaniel J. Mass*
Peter M. Senge**

WP 828-76 January 1976

*Assistant Professor of Management and Director, System
Dynamics National Modeling Project, System Dynamics
Group, Sloan School of Management.

**Doctoral Candidate and Member, System Dynamics Group,
Sloan School of Management.

In System Dynamics Group this paper is also listed as
D-2225-1.

copyright © 1976

Nathaniel J. Mass
I. INTRODUCTION

II. SINGLE-EQUATION STATISTICAL TESTS
   II.A. PRINCIPLE OF THE t-TEST AND THE PARTIAL
         CORRELATION COEFFICIENT
         II.A.1. Principle of the t-Test
         II.A.2. Principle of the Partial
                   Correlation Coefficient
   II.B. EXPERIMENTS WITH THE t-TEST AND THE PARTIAL
         CORRELATION COEFFICIENT

III. SINGLE-EQUATION AND WHOLE-MODEL TESTS OF THE IMPORTANCE OF
     A VARIABLE FOR MODEL BEHAVIOR
   III.A. SINGLE-EQUATION TESTS
   III.B. PRINCIPLE OF WHOLE-MODEL TESTS
   III.C. AN EXAMPLE OF WHOLE-MODEL TESTING--THE INFLUENCE
          OF DELIVERY DELAY ON SALES BEHAVIOR
   III.D. APPLICATIONS OF WHOLE-MODEL TESTING TO TEST
          ALTERNATIVE THEORIES OF SOCIAL BEHAVIOR

IV. CONCLUSIONS

BIBLIOGRAPHY
I. INTRODUCTION

One of the most difficult and subtle tasks confronting the mathematical model-builder is the selection of appropriate variables and functional relationships for his model. In specifying each equation, the modeler faces two distinct problems. First, he must posit a priori hypotheses about the variables or factors which he believes are important causes of change in the dependent variable of that equation. Second, once such hypotheses have been advanced, the modeler must have some means of testing whether or not, given the available empirical information, those variables which a priori reasoning has led him to consider should be included in the model. Based on the results of such testing, the modeler may be led to reject certain variables as being relatively unimportant, and may thereby begin to refine his initial causal hypotheses.

This paper is addressed to the second aspect of the problem of variable selection—testing the importance of initially-hypothesized variables. The paper examines a variety of tests used by model-builders to aid in the selection of model variables and demonstrates that some are far better suited to the task than others. In particular, the paper suggests that the only tests which provide a sound basis for including or excluding an individual variable are tests which examine the variable's impact on overall model behavior. Conversely, the single-equation statistical tests employed extensively in the social sciences are shown to be inadequate for rejecting or accepting a variable in a causal model.
The overall theme of the paper is developed in two steps. First, Section II discusses the proper role of statistical tests. The discussion focuses on the two most widely-used statistical significance tests: the t-test for statistical significance and the partial correlation coefficient. Both tests attempt to measure the impact of an individual variable in an equation containing multiple explanatory variables. Section II includes a summary of the principles underlying each test and a detailed example of how the tests might be used in a model of market growth and capacity expansion in a typical firm. The discussion shows that both the t-test and the partial correlation coefficient test the extent to which an individual variable's impact is measurable, given the available data, and therefore should be viewed as tests of data usefulness, not as tests of model specification. For example, failure to pass the t-test signifies only that the available data do not permit accurate estimation of the coefficient associated with a particular variable. Failure to pass the t-test tells the modeler that the data are not useful for estimating a variable's impact, not that the variable's impact is in any way unimportant.

The limited role of single-equation statistical tests is especially interesting in light of the predominance of such tests in the social science literature. Yet, criticism of single-equation statistical tests as a basis for testing model specification dates back at least to the 1930's. For example, in his review of Tinbergen's *Statistical Testing of Business Cycle Theories: A Method and its Application to Investment Activity*, which was the first large-scale application of multiple-correlation analysis (what we are here calling
single-equation statistical testing) to macroeconomic problems, Keynes argues that

The method [multiple-correlation analysis] is one neither of discovery nor of criticism. It is a means for giving quantitative precision to what, in qualitative terms, we know already as the result of a complete theoretical analysis [italics added].... How far are these curves and equations meant to be no more than a piece of historical curve-fitting and description, and how far do they make inductive claims with reference to the future as well as the past? If the method [multiple-correlation analysis] cannot prove or disprove a theory, and if it cannot give a quantitative guide to the future, is it worthwhile? For, assuredly, it is not a very lucid way of describing the past.*

As a response to the limitations of single-equation statistical tests, the second part of the paper, Section III, discusses how one might test the importance of a variable for model behavior. Model-behavior testing requires construction of a complete system model that includes the feedback relationships whereby the "dependent variable" in one equation feeds back and influences future values of the explanatory variables in the same equation. Within the context of a system model, the importance of a particular variable on model behavior can be tested by excluding that variable and examining the consequent change in overall model behavior. Generally, such tests require computer simulation, although analytic results are possible for simple systems. Model-behavior testing can identify variables which are essential to producing specific modes of system behavior—for example, oscillations

---

*J. M. Keynes, "Professor Tinbergen's Method" [a review of J. Tinbergen, Statistical Testing of Business Cycle Theories: A Method and its Application to Investment Activity (Geneva: League of Nations, 1939)], Economic Journal (September 1939), pp. 567, 569.
of a certain periodicity—or which have no appreciable effect on behavior. Section III includes an illustration of model-behavior testing which shows how the technique can be used to determine whether or not delivery delay is an important determinant of sales behavior in the corporate model presented in Section II.

The implications of the present paper for the process of theory testing are especially important in the social sciences where there are many examples of alternative theories which match the statistical evidence equally well. Whereas most social scientists attribute the presence of alternative "equally valid" theories to the paucity of reliable social data, the present paper suggests that in fact the prevailing philosophy of how theories should be tested is conceptually unsound and inadequate for discriminating among alternative theories of social behavior. As an illustration of the promise of a basic reorientation in approach, Section III.D includes a brief discussion of how some of the more persistent theoretical debates in the social sciences might be clarified through model-behavior testing.
II. SINGLE-EQUATION STATISTICAL TESTS

Section II shows that single-equation statistical tests should be interpreted as tests of data usefulness. The discussion focuses on two particular statistical measures—the t-statistic and the partial correlation coefficient—and shows, both theoretically and practically, that these widely-used statistical tests should not be interpreted as tests of model specification.

Demonstration of the proper interpretation of single-equation statistical tests is especially appropriate in light of the wide-spread misunderstanding, manifest by model-users and model-builders alike, of these tests. Frequently, journal articles report statistical significance test results as measures of one variable's significance in determining another. One might read, for example, that a statistical study has shown that years of education is "statistically insignificant in explaining changes in income" or that unemployment is "statistically significant in explaining changes in wage rates." The implication of such statements appears to be that education is an insignificant determinant of income or that unemployment is a significant determinant of wage rates. Yet, such inferences are quite unjustified given the nature of statistical significance tests. The same misunderstandings are manifest when researchers discuss the inadequacy of available social and economic data. When statistical testing fails to convincingly discriminate between alternative theories, authors often describe the poor data that are available for the testing, as if more generous or more accurate data would have rendered the same tests valid indicators of which theory
is superior. Hopefully, understanding the nature of statistical tests as measures of data usefulness, not tests of model specification, will enable model-builders and model-users to utilize the tests more soundly.

II.A. PRINCIPLE OF THE t-TEST AND THE PARTIAL CORRELATION COEFFICIENT.

II.A.1. Principle of the t-Test.*

The t-test provides the modeler with a measure of the confidence he can place in an estimated parameter value. A statistical parameter estimator is a random variable.** The estimator is a random variable because the equation being estimated, and therefore the data upon which the estimate is based, is assumed to have a random component. Because it is a random variable, the estimator has a mean and a variance. If the estimator has a large variance, little can be said with confidence about its accuracy. That is, even if the mean of the estimator equals the true value of the parameter being estimated, a large variance means that the probability that the parameter estimate is close to the true parameter value is low. The t-statistic allows the modeler to perform a formal statistical hypothesis test to determine whether or not the variance of the parameter estimator is "too large."

In a linear regression, the t-statistic is computed as the ratio of the computed parameter estimate \( \hat{\beta}_{LS} \) to the estimated standard deviation of the parameter estimate \( \hat{\beta} \).

*The following exposition of the t-test does not attempt to explain fully the mechanics of the test. For such explanation, the reader is referred to an introductory econometrics text (for example, Theil [19]).

**The term "estimator" refers to the estimation technique as an operator which converts a set of data into a set of parameter estimates. The distinction between the estimator and the estimate is analogous to the distinction between a random variable and one particular value of the random variable.
As the estimated standard deviation $\hat{\sigma}_\beta$ increases relative to the parameter estimate $\hat{\beta}$, the t-statistic diminishes. If one makes a certain set of assumptions, the t-statistic can be used to test the hypothesis that the true parameter value $\beta$ equals zero.* For example, for large samples a t-statistic whose absolute value is greater than 2.3 implies that the true parameter $\beta$ is non-zero with probability above 0.99. Therefore, when the parameter estimate exceeds its own estimated standard deviation by a factor of approximately two or greater, the t-statistic suggests that confidence can be placed in the accuracy of the estimate. A t-statistic of less than one suggests that little confidence can be placed in the estimate.

As a measure of confidence in a parameter estimate, the t-statistic guides the modeler toward judicious use of the available sample of data. It does not tell him whether or not the variable associated with a particular parameter is important. Passing the test occurs because the estimate is fairly "tight"—that is, the estimated standard deviation of the estimator is fairly small relative to the estimated parameter value. However, such an outcome only tells

\[ t\text{-stat} = \frac{\hat{\beta}_{LS}}{\hat{\sigma}_\beta} \]  

*The present paper focuses on the logical limitations of the t-test and makes no attempt to evaluate the realism of its underlying assumptions. These assumptions, which include perfect specification of the model being estimated (including zero-mean, normally-distributed noise inputs in each equation) and perfect measurement of all variables, are discussed elsewhere (Senge [18]) in the context of evaluating the accuracy of least-squares parameter estimates.
the modeler that the parameter can be estimated with some accuracy; even though accurate estimation is possible, the associated variable may still be quite unimportant in the model.

Similarly, failure to pass the t-test does not imply that a variable can be omitted from a model. Failure to pass the t-test occurs when the estimated standard deviation of the estimate \( \hat{\beta} \) becomes too large relative to the estimated parameter value (Equation [1]). The only conclusion justified by failing the t-test is that little confidence can be placed in the accuracy of the computed parameter estimate. The associated variable could have a very important impact in the model but, for a variety of reasons, that impact is not subject to accurate estimation.

Therefore, the t-test is best interpreted as a test of data usefulness, not as a test of model specification. Failure to pass the t-test says that the available data are not particularly useful for estimating the parameter \( \beta \). Conversely, passing the t-test says that the data are useful for the purpose of estimating \( \beta \)—that is, \( \beta \) can be estimated with high accuracy. Obviously, neither outcome directly addresses the importance of the variable associated with the parameter \( \beta \) itself.

II.A.2. **Principle of the Partial Correlation Coefficient**

The partial correlation coefficient provides the modeler with a measure of the incremental contribution of a single explanatory variable in accounting for variation in a dependent variable.
Denote the variable in question $x_h$. If one computed the fit of the estimated equation twice, once with the additional variable $x_h$ and once without, he would have two computations of the coefficient of multiple determination $R^2 - R^2_{\text{total}}$ and $R^2_h$, the latter computed with $x_h$ omitted from the equation. The partial correlation coefficient for the variable $x_h$, $r_h$, would then be determined:

$$r_h^2 = \frac{R^2_{\text{total}} - R^2_h}{1 - R^2_h}$$

(7)

Clearly, the partial correlation coefficient $r_h$ increases as the incremental contribution of the additional variable $x_h$--($R^2_{\text{total}} - R^2_h$)--increases. The partial correlation coefficient $r_h$ also increases as $R_h$ (the coefficient of multiple determination without $x_h$) becomes smaller. For example, if the variable $x_h$ improves the equation fit by 0.1 (i.e., $R^2_{\text{total}} - R^2_h = 0.1$), the partial correlation coefficient for $x_h$ equals 1 if the $R^2_h$ equals 0.9 and it equals $(0.167)^{1/2} = 0.410$ if $R^2_h$ equals 0.4. The latter property reflects the increasing difficulty of improving equation fit as the $R^2$ approaches unity (perfect fit).

A close relationship exists between the partial correlation coefficient $r_h$ and the t-statistic $t_h$ for the coefficient associated with the variable $x_h$ in a least-squares regression:

*Theil [19], p. 174, provides a derivation of Equation (8).
where \( r_h \) = partial correlation coefficient for \( x_h \)
\( t_h \) = t-statistic for \( x_h \)
\( n \) = periods of data available
\( k \) = number of coefficients to be estimated in equation

Taking the partial derivative of \( r_h \) with respect to \( t_h \),

\[
\frac{\partial r_h}{\partial t_h} = \frac{(n - k)}{(t_h + n - k)\sqrt{t_h^2 + n - k}}
\]

Equation (9) shows that the partial correlation coefficient decreases with decreases in the t-statistic \( \frac{\partial r_h}{\partial t_h} > 0 \) provided there are more data points than parameters to be estimated \( (n - k > 0) \). Equation (9) implies that the same factors which cause an individual t-statistic to fall cause the partial correlation coefficient to fall. Equation (8) implies that, if the t-statistic approaches zero, the partial correlation coefficient likewise approaches zero. In short, the partial correlation coefficient indicates a negligible contribution of \( x_h \) under the very same conditions which cause the parameter estimates associated with \( x_h \) to be statistically insignificant, and vice versa. Therefore, the arguments advanced in Section II.A.1 apply to the partial correlation coefficient as well as to the t-test: \( r_h \) should not be viewed as a test of model specification.

One interesting implication of the above discussion of the partial correlation coefficient is that the incremental contribution of an individual variable to equation fit is an inadequate measure of the importance
of a variable. For example, a finding that variable \( x_1 \) "explains" 60% of the variation in the dependent variable \( y \) \((R^2_{\text{total}} - R^2_1 = 0.6)\) and variable \( x_2 \) "explains" only 10% \((R^2_{\text{total}} - R^2 = 0.1)\) does not imply that variable \( x_1 \) is a more important determinant of changes in \( y \) than variable \( x_2 \). Equations (8) and (9) suggest that such a discrepancy between the measured contributions of \( x_1 \) and \( x_2 \) to \( y \) corresponds to a similar discrepancy between \( t_1 \) and \( t_2 \) (the t-statistics corresponding to \( x_1 \) and \( x_2 \)). However, Section II.A.1 showed that \( t_1 \) greater than \( t_2 \) means only that the parameter associated with \( x_1 \) can be more accurately estimated than the parameter associated with \( t_2 \). Section III.A will explain further why single-equation measures like the incremental contribution to equation fit are fundamentally inadequate for determining the importance of a variable in a system model.

II.B. EXPERIMENTS WITH THE t-TEST AND THE PARTIAL CORRELATION COEFFICIENT

The following series of experiments illustrates the theoretical limitations of the t-test and partial correlation coefficient described above. In each experiment, a feedback model of marketing and production capacity acquisition in a typical firm (see Forrester [7]) is defined as the real world and used to generate time-series data. The data are then used to estimate the model equation for order backlog assuming that the modeler has perfect a priori knowledge of the specification of the equation. Order backlogs in the model represent the accumulated difference between incoming orders (sales) and deliveries. One of the determinants of sales in the model is delivery delay (that is, the delay in filling new orders). In each experiment, a t-test is performed for the
estimated impact of delivery delay on sales and a partial correlation coefficient for delivery delay is computed. The experiments examine whether or not these tests provide reliable indicators of the importance of delivery delay as a determinant of sales.

The experimental format has been used elsewhere to examine the accuracy of statistical parameter estimates (see Senge [18]). In the present context, the estimation experiments serve two purposes. First, the experiments provide a concrete demonstration of the type of information supplied by the single-equation statistical tests. Second, in looking at a specific case, the experiments establish a background for the discussion of model-behavior testing in Section III.B. In that section, the market growth model will serve to demonstrate how model-behavior testing can be used to show that delivery delay is in fact an important determinant of sales.

The equation to be estimated in the following experiments relates changes in order backlogs (that is, the net of incoming orders and outflowing deliveries) to four explanatory variables: the number of salesmen employed by the firm S, the delivery delay in filling orders as recognized by the market DDRM, the firm's production capacity PC (measured in maximum orders per month which can be filled), and the current order backlog BL:

\[ \Delta BL = OB_t - DR_t \]  
\[ OB_t = S_t \cdot SE_t + \epsilon_{1t} \]  
\[ SE_t = g_1(DDRM_t) \]  
\[ DR_t = PC_t \cdot g_2(BL/PC)_t + t_{2t} \]  
\[ \Delta BL = BL_{t+1} - BL_t \]
where \( BL = \) order backlog
\( DR = \) delivery rate (orders shipped per month)
\( SE = \) sales effectiveness (sales per salesman per month)
\( PC = \) production capacity (maximum possible orders shipped per month)
\( DDRM = \) delivery delay recognized by market (months)
\( \varepsilon_1, \varepsilon_2 = \) random error processes
\( g_1(\cdot) = \) nonlinear function
\( g_2(\cdot) = \) nonlinear function

Because the ensuing discussion focuses on the estimated impact of delivery delay on incoming orders in the model firm, a brief description of the correct equation for orders booked (Equation [3]) is appropriate. Orders increase as the number of salesmen increase, all other factors remaining unchanged. Delivery delay, however, modulates orders. As the firm's delay in filling orders increases, more customers turn to competitors, thereby depressing sales. Delivery delay has a nonlinear impact on sales in the market-growth model, as shown in Figure 1, causing orders to fall from a maximum of 400 orders per salesman per month to practically zero when extremely long delivery delays prevail.

\[ OB_t = \varepsilon_t \cdot g_1(DDRM_t) \]
A nonlinear relation is necessary to describe the impact of delivery delay on sales because sales cannot fall below zero for large delivery delays and because, at extremely low (or high) delivery delays, customers become insensitive to further reductions (or increases) in delivery delay.

The order backlog equation (Equation [1]) is part of a complete system of equations which relate order backlog to capacity-acquisition and marketing policies in the model firm.* Figure 2 gives an overview of the structure of the model used to generate data in the following experiments. The model firm expands or contracts its sales force depending upon the number of salesmen that could be supported by the marketing budget compared with the existing sales force. Delivery delay is taken as a measure of excess demand or supply for the firm's product. The firm orders additional production capacity when delivery delay becomes longer than desired (the firm's desired delivery delay is taken as two months) and reduces capacity when delivery delay falls below the desired value. Two distinct delivery delays appear in Figure 2—delivery delay recognized by company DDRC and delivery delay recognized by market DDRM—because the customer is responding, with a delay, delivery delays quoted by the company. (Both DDRC and DDRM are smoothed values of the current delivery delay, given by the backlog divided by the delivery rate average, BL/DRA.)

In order to estimate the order backlog equation (Equation [2]), the nonlinear functions \( g_1(\text{DDRM}) \) and \( g_2(\text{BL/PC}) \) must be parameterized. In the following experiments, the function \( g_1(*) \) in Equation (1) is approximated as linear and \( g_2(*) \) is specified as a third-order polynomial.

*See Forrester [7] for a complete description of the model structure.
Figure 2: Major Feedback Loops in Forrester's Market-Growth Model
(g2(·) is definitionally constrained to be zero when (BL/PC) equals zero.) Consequently, delivery delay recognized by market DDRM enters in the second coterm multiplied by the parameter K2:

\[ \Delta BL = S \cdot (K1 + K2 \cdot DDRM) - PC \cdot (K3(BL/PC) + K4(BL/PC)^2 + K5(BL/PC)^3) \]

\[ = K1 \cdot S + K2 \cdot S \cdot DDRM + K3(BL) + K4(BL^2/PC) + K5(BL^3/PC^2) + \varepsilon \]  

(6)

\[ \Delta BL \equiv BL_{t+1} - BL_t \]

\( \varepsilon \) is a zero-mean, normally distributed, uncorrelated random process

K1, ..., K5 = unknown parameters

The following experiments examine whether the t-test for K2 and the partial correlation coefficient for the coterm (S•DDRM) provide (1) reliable indicators of the importance of delivery delay as a determinant of orders booked (sales) or (2) reliable indicators of the usefulness of available data in estimating the impact of delivery delay on orders booked.

In the first sequence of experiments, the magnitude of the noise inputs in the model used to generate data progressively increase. Each equation in the data-generating model has an additive random term.

---

*Approximating \( g_1(\cdot) \) as linear and \( g_2(\cdot) \) as a third-order polynomial allows the modeler to draw the maximum statistically significant information from the experimental data. The variation in delivery delay recognized by the market DDRM (the input to \( g_1(\cdot) \)) does not carry that variable into significantly nonlinear regions of the relationship shown in Figure 1. The opposite holds for the backlog to production capacity ratio (BL/PC), which ranges well into the nonlinear regions of \( g_2(\cdot) \).
The random variables are zero-mean, normally distributed, homoscedastic (constant standard deviation), free of autocorrelation, and mutually independent. The standard deviation of each random term is set as a fraction of the mean value of the dependent variable in the same equation. In the three estimations shown in Table 1, the standard deviations of the random inputs is progressively increased from 1% to 5% to 10% to 15% and then to 20%. In each case, Table 1 gives the true parameter values (TRUE VALUE), the ordinary least-squares (OLS) parameter estimates (ESTIMATED VALUE), the estimated standard deviations (\(\hat{\sigma}_\beta\)) and the t-statistic (T-STAT) for each estimate, the partial correlation coefficient (\(r_h\)) for each coterm, and the coefficient of multiple determination \(R^2\) for each estimation.

Table 1 shows that, as the magnitude of the noise inputs increases, the statistical significance of all parameter estimates diminishes. The partial correlation coefficients (\(r_h\)) likewise diminish. In particular, the parameter \(K_2\), which measures the impact of delivery delay on changes in order backlog, has a t-statistic whose absolute value diminishes from 327.7 to 3.19 as the noise inputs increase from 1% to 20% of the mean value of the order backlog \(BL\) variable. Therefore, statistical measures indicate increasing statistical insignificance as the magnitude of the random errors (\(\varepsilon\) in Equation [6]) present in each equation increases, even though the structure of the equation in question remains perfectly specified.

In the next sequence of experiments, random measurement errors are included in the data taken from the data-generating model. Whereas random noise inputs in the data-generating model circulate
TABLE 1

EFFECT OF INCREASED NOISE INPUTS ON t-STATISTIC
Ordinary Least-Squares Estimation

\[ BL = BL_{-1} + K1 \cdot S_{-1} + K2 \cdot S_{-1} \cdot DDRM_{-1} + K3 \cdot BL_{-1} + K4 \cdot (BL_{-2}/PC_{-1}) + K5 \cdot (BL_{-3}/PC_{2}) \]

| COEF  | TRUE VALUE | ESTIMATED VALUE | \( \hat{c}_\beta \) | T-STAT | \( r_h \) |
|-------|------------|----------------|-------------------|--------|---------|
| 1%    | K1         | 475            | 486.6             | 2.185  | 222.8   | 0.9991  |
|       | K2         | -61.5          | -61.13            | 0.1865 | -327.7  | -0.9996 |
|       | K3         | -0.6178        | -0.6686           | 0.00406| -164.5  | -0.9983 |
|       | K4         | 0.1324         | 0.1440            | 9.029 \times 10^{-4} | 159.4  | 0.9982  |
|       | K5         | -0.00975       | -0.0107           | 8.449 \times 10^{-5} | -126.2 | 0.9971  |

R² = 1

| 5%    | K1         | 475            | 519.3             | 27.92  | 18.60   | 0.8867  |
|       | K2         | -61.5          | -63.03            | 2.296  | -27.45  | -0.9429 |
|       | K3         | -0.6178        | -0.7718           | 0.06128| -12.59  | -0.7924 |
|       | K4         | 0.1324         | 0.1733            | 0.01553| 11.16   | 0.7550  |
|       | K5         | -0.00975       | -0.01315          | 0.00158| -8.33   | -0.6517 |

R² = 0.993

| 10%   | K1         | 475            | 457.7             | 37.05  | 12.36   | 0.7851  |
|       | K2         | -61.5          | -54.62            | 5.638  | -9.69   | -0.7049 |
|       | K3         | -0.6178        | -0.6484           | 0.06398| -10.13  | -0.7207 |
|       | K4         | 0.1324         | 0.136             | 0.02018| 6.74    | 0.5689  |
|       | K5         | -0.00975       | -0.00975          | 0.00299| -3.26   | -0.3171 |

R² = 0.9934

| 15%   | K1         | 475            | 423.1             | 42.84  | 9.88    | 0.7118  |
|       | K2         | -61.5          | -44.5             | 7.732  | -6.11   | -0.5340 |
|       | K3         | -0.6178        | -0.6014           | 0.07414| -8.11   | -0.6397 |
|       | K4         | 0.1324         | 0.1211            | 0.03850| 3.15    | 0.3072  |
|       | K5         | -0.00975       | -0.00868          | 0.00722| -1.20   | -0.1223 |

R² = 0.9731

| 20%   | K1         | 475            | 380.3             | 50.95  | 7.46    | 0.6080  |
|       | K2         | -61.5          | -34.57            | 10.85  | -3.19   | -0.3107 |
|       | K3         | -0.6178        | -0.5186           | 0.0862 | -6.02   | -0.5252 |
|       | K4         | 0.1324         | 0.0637            | 0.0625 | 1.02    | 0.1041  |
|       | K5         | -0.00975       | -0.00003          | 0.01504| -0.002  | -0.0003 |

R² = 0.9232
through the data-generating model and thereby affect model behavior, measurement errors do not affect model behavior. Measurement errors only limit the precision with which system variables can be measured. In Table 2, random measurement errors of 1%, 5%, 10%, 15%, and 20% are successively added to data generated with noise inputs of 10%. Measurement errors of 5% to 10% are typical of much socio-economic data, and larger errors occur in some data (see Morgenstern [15]).

Table 2 shows that moderate measurement errors lead to unacceptable t-statistics and partial correlation coefficients for the estimated impact of delivery delay on sales. For example, when 10% measurement error is included in the data, the t-statistic for the parameter K2 equals -1.24 which is well below the 0.99 confidence level and the partial correlation coefficient equals -0.127. When measurement errors of 15% enter the data, K2 has a t-statistic of -0.257 and the coterm (S*DDRM) has a partial correlation coefficient of -0.02638.

How might a model-builder interpret the estimated delivery delay impact in Table 2? Consider the case of 10% measurement error in particular. If he interpreted the statistical results as tests of the specification of the order backlog BL equation, he would conclude that delivery delay recognized by market DDRM is not an important determinant of the sales component in that equation. When viewed in this way, the low t-statistic would signify a "statistically insignificant" relationship between delivery delay and sales. Likewise, the low partial correlation coefficient would be seen as evidence that "delivery delay explains only a small fraction of the variation in backlog." When viewed as tests of
TABLE 2
EFFECT OF INCREASED MEASUREMENT ERROR ON t-STATISTIC
(all experiments have 10% noise inputs)
Ordinary Least-Squares Estimation

\[
BL = BL_{-1} + K_1 S_{-1} + K_2 S_{-1} \cdot DDRM_{-1} + K_3 BL_{-1} + K_4 (BL_{-1}^2/PC_{-1}) + K_5 (BL_{-1}^3/PC_{-1})
\]

| COEF | TRUE VALUE | ESTIMATED VALUE | \( \hat{\beta} \) | T-STAT | \( t_h \) | \( R^2 \) |
|------|-------------|----------------|--------------|--------|---------|---------|
| 1%   | K1 475      | 455.8          | 39.43        | 11.56  | 0.7645  | 0.992   |
|      | K2 -61.5    | -55.25         | 6.097        | -9.062 | -0.6809 |         |
|      | K3 -0.6178  | -0.6474        | 0.0703       | -9.209 | -0.6868 |         |
|      | K4 0.1324   | 0.1362         | 0.0222       | 6.126  | 0.5321  |         |
|      | K5 -0.00975 | -0.00972       | 0.00322      | -3.016 | -0.2956 |         |
| 5%   | K1 475      | 444.7          | 66.93        | 6.644  | 0.5633  | 0.9580  |
|      | K2 -61.5    | -49.16         | 12.79        | -3.842 | -0.3667 |         |
|      | K3 -0.6178  | -0.6182        | 0.1168       | -5.294 | -0.4773 |         |
|      | K4 0.1324   | 0.1253         | 0.0408       | 3.068  | 0.3002  |         |
|      | K5 -0.00975 | -0.00908       | 0.00546      | -1.661 | -0.1680 |         |
| 10%  | K1 475      | 408.2          | 89.48        | 4.562  | 0.4249  | 0.8486  |
|      | K2 -61.5    | -26.85         | 21.51        | -1.248 | -0.1270 |         |
|      | K3 -0.6178  | -0.5563        | 0.1530       | -3.637 | -0.3496 |         |
|      | K4 0.1324   | 0.07719        | 0.05362      | 1.440  | 0.1461  |         |
|      | K5 -0.00975 | -0.00411       | 0.00606      | -0.6792 | -0.0695 |         |
| 15%  | K1 475      | 371.4          | 100.7        | 3.687  | 0.3538  | 0.7053  |
|      | K2 -61.5    | -7.000         | 27.21        | -0.2573 | -0.0264 |         |
|      | K3 -0.6178  | -0.4501        | 0.1406       | -3.202 | -0.3121 |         |
|      | K4 0.1324   | 0.01443        | 0.03600      | 0.4009 | 0.0411  |         |
|      | K5 -0.00975 | -0.00102       | 0.00259      | 0.3930 | 0.0403  |         |
| 20%  | K1 475      | 361.0          | 108.9        | 3.314  | 0.3220  | 0.5606  |
|      | K2 -61.5    | 3.968          | 30.81        | 0.1288 | 0.0132  |         |
|      | K3 -0.6178  | -0.4545        | 0.1112       | -4.087 | -0.3867 |         |
|      | K4 0.1324   | 0.00929        | 0.01539      | 0.6034 | 0.0618  |         |
|      | K5 -0.00975 | -2.111 \times 10^{-5} | 4.591 \times 10^{-5} | -0.4599 | -0.0471 |         |
model specification, the t-statistic and partial correlation coefficient would suggest that delivery delay could be omitted from the order backlog equation with little loss in the model's explanatory power.

On the other hand, if properly interpreted, the t-statistic and partial correlation coefficient for the 10% case in Table 2 tell the modeler that he can learn little about the impact of delivery delay on sales from the available statistical evidence. The low t-statistic implies that the data have "little power" for the purpose of estimating the impact of delivery delay—that is, it is simply not possible to determine the value of the parameter K2 with much precision using OLS estimation. Likewise, if properly interpreted, the low partial correlation coefficient says that, in the available sample of data, the correlation between delivery delay and movements in order backlog is low. The low correlation clearly does not support the hypothesis that delivery delay is an important determinant of sales; however, it also does not imply the converse—that is, the low partial correlation does not imply that delivery delay is an unimportant influence on sales. Therefore, proper interpretation of the single-equation statistical tests leads the model-builder to conclude that he can learn little about the impact of delivery delay from the statistical estimation based on available data: the data permit neither precise estimation of the delivery delay parameter nor reveal statistical support for the a priori hypothesis that delivery delay should be included in the model.

In conclusion, the preceding sections have shown that, if they are used as tests of specification, the statistical tests lead to rejection
of a priori relationships when they cannot be estimated reliably. For example, the experiments in Section II.B showed that, if data contain measurement errors of 10% or greater, the estimated delivery delay impact is statistically insignificant and the partial correlation of delivery delay with changes in order backlog is low. If the model-builder rejects the a priori hypothesis that delivery delay is important on the basis of low statistical significance or low partial correlation, he implicitly places estimability as his criterion for specification. The above experiments show how such a criterion can lead to rejection of a correct model specification and suggest that estimability is not appropriate as a guide to constructing valid social theories.

How, then, should we test whether or not a variable should be included in a model? How should we determine whether or not delivery delay is an important determinant of sales? Section III suggests one way of conducting such testing which is based on analyzing the effect of delivery delay on overall model behavior. Section III shows that, in contrast to single-equation statistical testing, analysis of model behavior does seem to provide an adequate basis for testing whether or not a variable should be included in a model.
III. SINGLE-EQUATION AND WHOLE-MODEL TESTS OF THE IMPORTANCE OF A VARIABLE FOR MODEL BEHAVIOR

Section II showed that single-equation statistical tests, while frequently construed as measures of the "importance" of a variable, in fact do not convey any information directly about whether a particular model relationship is important or whether a particular model specification is correct. Instead, the tests provide a measure of the confidence that can be placed in parameter estimates derived from use of the technique given the available data, and assuming that the theoretical assumptions underlying the technique are in fact met; this function was summarized in section II by saying that single-equation statistical tests are primarily "tests of data usefulness." Thus, for example, a high t-statistic indicates that a particular parameter being estimated is significantly different from zero; the outcome does not indicate that the variable associated with the estimated parameter is in any sense important, but only that its influence is probably non-zero. Analogously, a low t-statistic does not imply that the corresponding variable is unimportant, but only that the data do not permit any inference as to whether or not the impact of the variable is significantly different from zero.

Given that single-equation statistical tests do not directly address the question of the relative importance of different variables, what alternative techniques might provide insight into this critical problem? Section III attempts to provide some guidelines and direction for addressing this question. In particular, two viewpoints for evaluating the importance of a variable—the single-equation approach and the whole-model approach—are here contrasted. Section III.A describes single-equation tests and discusses
their potential use, while sections III.B through III.D outline the whole-model testing approach. The resulting analysis shows that successful evaluation of the importance of a variable cannot be performed reliably in a single-equation framework, but instead requires whole-model testing which involves analyzing the full set of feedback loops connecting that variable to other system elements. Moreover, Sections III.C and III.D give examples of whole-model testing to show how such an approach can yield insight into the validity of alternative theories of social behavior.

III.A SINGLE-EQUATION TESTS

Single-equation tests, of which the t-test and partial correlation coefficient are specific examples, attempt to measure the magnitude of influence of one variable—the independent variable—or some other—dependent—variable. To illustrate the nature of such single-equation tests, consider the example introduced in Section II concerning the impact of delivery delay on incoming orders in a firm. To assess the importance of delivery delay as an influence on sales, a statistician would typically construct a model that relates sales (incoming orders) to such factors as price, delivery delay, and sales effort. This would yield a single equation of the general form shown below:

\[ \text{Sales} = a + b \text{delivery delay} + [\text{other factors}] + \epsilon \]

Given the above equation, the statistician might attempt to assess the importance of delivery delay as an influence on sales. First, he might investigate whether or not the parameter corresponding to the delivery delay term is significantly different from zero. However, as described earlier, such a significance test does not imply directly whether or not the associated variable is important, only whether or not we can ascertain that the variable
has a non-zero influence given the available data. An alternative, and frequently-used, approach for analyzing the "importance" of delivery delay is to consider the size of the parameter associated with the delivery delay term. According to this viewpoint, a large absolute value of the parameter would imply that long delivery delays exert a strong depressive effect on sales; conversely, if the parameter were small in absolute value, one might infer that the impact of delivery delay is unimportant.

However, the size of the delivery delay coefficient does not constitute a valid test of the importance of delivery delays. For example, in the market-growth described in Section II, delivery delay is the primary channel through which demand and supply are equilibrated. That is, a high delivery delay, which is an indication of excess demand, lowers sales while inducing expansion of production capacity. In this situation, a low delivery delay coefficient in the sales equation would indicate that market imbalances would be sustained, while a large coefficient would indicate rapid movement towards equilibrium. Thus, a low delivery delay coefficient would mean that delivery delay interactions with sales rate are a significant cause of disequilibrium behavior in the firm's order backlog; in this situation, the contribution of delivery delay to transient behavior would be inversely related to the size of the delivery delay coefficient. This point is illustrated mathematically below to emphasize that the magnitude of an estimated parameter is not in itself a good indication of the importance of the corresponding model variable.

The impact of delivery delay on behavior of backlog can be shown by analysis of the order backlog equation in conjunction with the equation for delivery delay in the market-growth model. Delivery delay, which is defined as the average time it takes to fill an order, depends on order backlog according
to the following relationship:

\[
\text{Current Delivery Delay} = \frac{\text{BL}}{\text{DRA}} \quad (\text{weeks}) = \frac{(\text{units})}{(\text{units/week})}
\]

DRA \equiv \text{delivery rate average}

Assuming the delivery rate average DRA (which depends on a variable production capacity in the full model) remains constant, and also assuming that the number of salesmen \( S \) employed by the firm does not vary, Equation (6) can be rewritten as follows:

\[
\Delta \text{BL} = a + b' \cdot \text{DDRM} + c \cdot \text{BL} + d \cdot \text{BL}^2 + e \cdot \text{BL}^3
\]

where

\[
a = K_1 \cdot S \\
b = K_2 \cdot S \\
c = K_3 \\
d = K_4 / \text{PC} \\
e = K_5 / \text{PC}^2
\]

(Salesmen \( S \), and Production Capacity PC constant)

Ignoring the information delay inherent in the delivery delay recognized by the market DDRM, we can equate the perceived delivery delay DDRM and the current delivery delay:

\[
\text{DDRM} = \frac{\text{BL}}{\text{DRA}}
\]

Equation (11) then becomes

\[
\Delta \text{BL} = a + b' \cdot \text{BL} + \ldots [\text{higher-order BL terms}]
\]

\[
b' = K_2 \cdot S / \text{DRA} + K_3 \\
a = K_1 \cdot S
\]

Rewriting Equation (13) as a continuous differential equation,

\[
\frac{d(\text{BL})}{dt} = a + b' \cdot \text{BL} + \ldots [\text{higher-order BL terms}]
\]

Because increases in delivery delay cause decreases in sales, the coefficient \( K_2 \) in Equation (6) and therefore the parameter \( b' \) in Equation (13) are negative. (That is, delivery delay is part of a negative feedback loop.)
Ignoring the higher-order backlog terms, Equation (14) can be solved for the behavior of order backlog:

\[
BL(t) = -(a/b') + (BL(0) + (a/b'))e^{b't} \quad (15)
\]

Because the parameter \( b' \) is negative, the second term in Equation (15)--\((BL(0) + (a/b'))e^{b't}\)--will eventually decay to zero, leaving the behavior of order backlog \( BL \) equal to its "steady-state" solution--\( BL(t) = -a/b' \). However, the smaller \( b' \) is, the more slowly order backlog \( BL \) decays to its steady-state behavior. Conversely, a large value for the parameter \( b' \) causes order backlog to adjust to its steady-state value more quickly. A well-established principle in the design of feedback control systems in engineering states that the feedback loops which have long time constants tend to dominate the pattern of system behavior. A small value for \( b' \) corresponds to a long time constant for the feedback loop linking order backlog, delivery delay, and sales; and vice versa. Therefore, in this simplified version of the market-growth model, the importance of delivery delay on behavior of order backlog \( BL \) is inversely related to the size of the delivery delay coefficient \( K2 \).

Therefore, one should not reject delivery delay as an important variable simply because the delivery delay coefficient is small.

Similarly, there are numerous examples of how a large coefficient value can be misleading if one concludes that the associated variable is particularly important for system behavior. Borrowing from an example discussed in Section III.D, consider the case of a large coefficient for changes in fixed capital investment in an equation for changes
in aggregate production as a function of changes in consumption, fixed
capital investment, and inventory investment. A large capital coeffi-
cient indicates that changes in fixed capital investment account for
a large fraction of the variation in production. This might suggest
that controlling fluctuations in fixed capital investment is an effec-
tive means of controlling fluctuations in production. However, if
fixed capital investment were held constant, output inventories would
tend to vary more and producers would be more inclined to vary labor
and/or capacity utilization to adjust production rates. Thus, fluctua-
tions in aggregate production would shift from variations in fixed cap-
ital investment to fluctuations in net inventory investment. The work
described in Section III.D shows that employment and inventory policies
alone can account for much of the observed short-term business-cycle
behavior in the national economy, although many writers and theorists
have attributed the short-term cycle to fixed capital investment poli-
cies due to the relatively large variations in investment characteristic
of the business cycle.*

The above examples show the inadequacy of single-equation anal-
yses for determining whether or not a variable should be included in a
model. The delivery delay example demonstrated the need to consider the
feedback of backlog on delivery delay in evaluating the significance of
a small delivery delay coefficient in the order backlog equation. Like-
wise, the importance of fixed capital investment in determining production

*See Evans [6] for description of the magnitude of investment fluctuations
over a typical business cycle. Burns [4] notes that most business-cycle
theories center about the determinants of fixed capital investment.
fluctuations cannot be determined by analysis of the production equation alone. The impact, for example, of reducing fluctuations in capital stock will depend on the effects such a reduction will have on fluctuations in inventories, labor, and capacity utilization. These examples illustrate how the feedbacks which interlink explanatory variables in any one equation must be analyzed thoroughly in assessing the importance of any one variable in the equation.

Although a thorough review of all single-equation analysis techniques has not been attempted here, the results presented above tend to suggest that, logically and conceptually, the importance of a variable cannot be assessed in a single-equation context using any technique. Instead, such assessment requires detailed analysis of the feedback linkages connecting model variables. This result is especially significant because theory testing and hypothesis testing in psychology, sociology, economics, and the other social sciences is dominated by single-equation analysis techniques, especially single-equation statistical techniques. However, to reiterate, it appears that such methods are unlikely to yield reliable inferences about the importance of different variables and the validity of alternative social theories. Sections III.B through III.D describe a method of whole-model testing that is designed to overcome the logical shortcomings of single-equation analysis.

III.B. PRINCIPLE OF WHOLE-MODEL TESTS

As stated earlier, evaluation of the importance of a variable requires analyzing the full array of feedback loops connecting system
variables. The whole-model testing approach described here is intended to provide a specific framework for conducting such analysis. Section III.B, then, outlines the basic elements of whole-model testing. Section III.C provides a detailed example of the approach, and, finally, Section III.D overviews some additional areas of potential application.

The whole-model testing approach entails three principal steps. Suppose we are interested in determining whether or not variable X is an important determinant of variable Y. First, a model must be constructed that interrelates the variables X and Y. The model should contain enough endogenous structure to portray how changes in one variable, say X, affect the present and future values of both X and Y. To continue the example introduced in Section II, suppose we wish to analyze the impact of delivery delay on sales behavior. An adequate model to address this question should include the direct effect of delivery delay on sales effectiveness and sales rate; but the model should also include, for example, the impact of delivery delay on production-capacity acquisition, and the feedback of the level of production capacity on delivery rate and delivery delay. Also, the model should include the influence of sales rate on revenues, capacity expansion, and marketing effort. Analysis of such interactions is necessary to determine how production capacity, delivery rate, delivery delay, marketing effort, and sales rate change over time.

The second step in whole-model testing involves simulating the model both with and without any direct influence of X and Y. How is the behavior of variable Y altered as a result of deleting the direct
link between X and Y? In the delivery delay example, assessing the impact of delivery delay on sales behavior would involve omitting any direct link between delivery delay and sales and seeing whether model behavior is altered significantly as a result.

Finally, the third step in whole-model testing involves analysis of the causes of behavior observed in the second step. If the behavior of Y is relatively unaltered, what other variables appear to dominate the behavior? If Y's behavior is altered significantly, what direct and indirect links between X and Y account for the change in behavior? Sections III.C and III.D illustrate the application of this three-step procedure and show the kinds of insights derivable from whole-model testing.

Before proceeding to give specific examples of the whole-model testing process, it is useful to discuss several criteria for judging whether or not a given variable exerts an important impact on system behavior. The whole-model testing process consists of analyzing the change in model behavior resulting when a given factor is omitted as a direct causal influence in the model. However, in order to gauge whether or not omission of the variable is significant, some criteria are needed for judging whether or not the associated change in model behavior is significant. At least three criteria are possible:

(1) Does omission [inclusion] of the factor lead to a change in the predicted numerical values of the system?

(2) Does omission [inclusion] of the factor lead to a change in the behavior mode of the system? (For example, does it damp out or induce fluctuations in the system?)
(3) Does omission [inclusion] of the factor lead to rejection of policies that were formerly found to have had a favorable impact or to a reordering of preference between alternative policies?

In general, the results of an evaluation of the importance of a given variable will depend on which of the three criteria above are being used as a basis for the evaluation. For example, suppose we are constructing a model to explore the causes of sales fluctuations in a particular firm around a long-term growth trend in sales. Also suppose that the model is exhibiting the basic problematic sales behavior, as shown by the curve labeled A in Figure III-1. Suppose that we now omit a particular factor from the model (or, alternatively, add a particular factor), and the model behavior shifts to that described by Curve B in Figure III-1. Curve B differs from A in the exact numerical values for sales over time, but both curves clearly exhibit about the same general growth trend and about the same magnitude of fluctuations. The difference between outcomes A and B would then be judged as important by the first criterion given above, but unimportant by the second criterion. To deal with the third criterion, suppose that a number of policies were tested on both the models underlying Curves A and B and it was found that the policies that caused reduced fluctuations in one model also caused reduced fluctuations in the other model, and conversely. In this situation, the difference between the two models would be judged
insignificant or unimportant by the third criterion.* This example, although highly simplified, illustrates some of the considerations involved in assessing the importance of a given model variable or relationship. The purpose of a particular study will, in general, determine which of the above three criteria is most pertinent to the evaluation process.

![Figure III-1: Alternative Patterns of Sales Behavior](image)

III.C. AN EXAMPLE OF WHOLE-MODEL TESTING—THE INFLUENCE OF DELIVERY DELAY ON SALES BEHAVIOR

Section III.C gives a detailed example of the whole-model testing process, discussing the influence of delivery delay on sales

---

*To show still a more complex case, consider Curve C in Figure III-1. Curve C shows alternating periods of growth and leveling off rather than growth and decline as in Curve A or B. Curve C might be judged to be significantly different from, say, Curve A by both the first and second criteria (although some similarities in the overall mode of behavior are apparent—a secular rise in sales with periodic interruptions in the rate of growth). The outcomes might not be significantly different from the standpoint of the third criterion, however, if, for example, the same policies that caused reduced fluctuations in A also contracted the leveling-off periods in C, and conversely.
behavior. This example is intended, first, to illustrate how whole-model testing yields information about the contribution of variables to behavior that cannot be ascertained from single-equation tests; and second, to show that the information produced by whole-model testing can aid in discriminating which variables are and are not important or critical in generating a particular pattern of real-world behavior.

Section II overviewed the structure of a simple dynamic model of marketing and production-capacity acquisition policies. The section showed that even with perfect knowledge of the determinants of sales (orders booked) and shipments (delivery rate), statistically insignificant estimates of the influence of delivery delay on sales could be obtained under a variety of general circumstances. An econometrician, viewing this result, might conclude that delivery delay is a relatively unimportant influence on sales. However, this section shows, using whole-model testing, that delivery delays in fact exert a pronounced impact on sales behavior in the same market-growth model.

Figure III-2 shows the basic behavior of the market-growth model (in which a long delivery delay on the part of the firm is assumed to have a direct depressive effect on sales). Figure III-1 shows a behavior pattern of growth in sales (orders booked) interrupted by periods of decline. Such growth instability is caused by the simultaneous effect of delivery delay on sales and capacity expansion. At the outset of the simulation, the firm's delivery delay is low. Sales effectiveness is consequently high, thereby allowing orders booked to increase rapidly as the sales force increases. The growth of sales increases the firm's order backlog.
Delivery delay rises as a result, since backlog is growing more rapidly than production capacity. Eventually, delivery delay increases to the point where it begins to suppress orders. As the firm perceives the long delivery delay, it begins to order new production capacity. However, due to delays in perceiving delivery delay and in acquiring production capacity, delivery delay continues to rise for awhile, thereby lowering sales effectiveness and depressing orders booked. As capacity is eventually expanded, delivery delay starts to decline. The decline in orders booked and order backlogs further lowers delivery delay. Once delivery delay falls, growth in orders booked resumes and capacity expansion declines. Figure III-2 therefore shows alternating periods of growth and decline in sales and production capacity. In the following discussion, the pattern of system behavior shown in Figure III-2 will be assumed to be the real-life behavior which the modeler seeks to reproduce and understand.

In order to test the contribution of delivery delay to sales behavior, the direct influence of delivery delay on salesman effectiveness can be eliminated from the market-growth model.* A simulation of the resulting model is shown in Figure III-3. In contrast to the previous simulation, Figure III-3 shows a continued growth in orders booked and continued increase in production capacity after month 30, although

*Note that the vertical plot scales are different in Figures III-1 and III-2 due to the expanded range of variation in salesmen, orders booked, and production capacity seen in Figure III-2.
Figure III-3: Market-growth model with no direct influence on orders booked (sales)
both orders and capacity expand at an alternating rate of growth. Thus, with the direct link between delivery delay and sales eliminated, the market-growth model no longer exhibits the recurring fall-offs in sales characteristic of the original model. With this link omitted, a high delivery delay no longer exerts a depressive effect on sales. Instead, sales are influenced only by the level of salesmen and by a constant sales effectiveness.

Analysis of the structure of the market-growth model clearly shows how omission of the influence of delivery delay on sales shifts model behavior from that seen in Figure III-2 to the outcome in Figure III-3. Figure III-4 shows two negative feedback loops and one positive feedback loop which control sales and delivery delay in the original market-growth model. In Loop 1, an increase in sales raises revenues; high revenues, in turn, increase the revenues available for marketing and sales, leading to increased salesmen and, thereby, still higher sales. Loop 1, therefore, tends to promote continued exponential growth in sales.

Loops 2 and 3 are negative feedback loops controlling delivery delay. In Loop 2, an increase in sales raises order backlog, causing delivery delay to increase. In turn, a higher delivery delay lowers salesmen effectiveness, thereby lowering sales.

In Loop 3, an increase in delivery delay encourages capacity expansion, causing a decline in delivery delay through increased shipments.

When delivery delay is omitted as a direct influence on sales rate, Loop 2 no longer serves as a restraint on sales. Sales behavior, instead, is controlled by the positive feedback loop, Loop 1. Inactivation
Figure III-4: Principal feedback loops governing behavior of market-growth model.

of Loop 2 thus accounts for the nearly steady exponential growth in sales observed in Figure III-3. In Figure III-3, production capacity grows continually after month 30 (due to steadily rising sales) but at an alternating rate of increase. These fluctuations in production-capacity
expansion are generated in Loop 3 and are due to the fluctuations in delivery delay seen in Figure III-3. The latter fluctuations, in turn, result from variations in growth in production capacity.

Analysis of Figure III-3 thus shows the critical impact of delivery delay on sales behavior.* When delivery delay is no longer assumed to influence sales, the behavior of the market-growth model is altered markedly and the model no longer exhibits alternating periods of growth and decline in sales. In other words, the overall behavior mode of the system is significantly changed when the delivery delay effect is omitted. This result concerning the important influence of delivery delay on sales behavior contrasts with the outcome of the estimation experiments reported in Section II, which suggest that delivery delay is not a statistically significant determinant of sales. Such results might be interpreted by an econometrician to suggest that delivery delay is not an "important" influence on sales. However, the outcome of the whole-model test highlights the point that single-equation significance tests do not measure the importance of a given variable, and indicates the need for whole-model testing in order to assess the contribution to behavior of a variable, or set of variables, under study.

*Note that the outputs shown in Figures III-2 and III-3 are simulations of a deterministic system containing no random noise inputs. The essential outcome of the whole-model test remains unaltered, however, if uncertainty is introduced into the system by inserting random noise generators in one or more of the model equations.
III.D. APPLICATIONS OF WHOLE-MODEL TESTING TO TEST ALTERNATIVE THEORIES OF SOCIAL BEHAVIOR

The two previous sections have outlined the principal components of the whole-model testing approach and given a detailed example of its application. This section summarizes one additional application of whole-model testing that has been conducted, and outlines a number of areas where whole-model testing may help to clarify alternative theories of social behavior.

In a recent application, Mass [13, 14] has used whole-model testing to assess the generic causes of short-term and medium-term cycles in the national economy. He develops a general model of a producing unit within the economy that can be expanded to incorporate various hypothesized causes of economic instability.

One particularly significant outcome of the above-mentioned work regards the relative importance of labor adjustments and fixed capital investment in generating short-term business cycles. As noted by Burns [4], the predominant number of business-cycle theories, including the theories of Paul Samuelson, John Hicks, Nicholas Kaldor, and James Duesenberry, emphasize fluctuations in fixed capital investment as a cause of overall fluctuations in income and output.* Such theories have been widely influential from the theoretical standpoint, and have stimulated much subsequent business-cycle research. In addition, widespread acceptance of the theories has led to formulation of economic stabilization policies designed to regulate investment opportunities.

*See Samuelson [17], Hicks [10], Kaldor [11], and Duesenberry [5] for the original statement of these theories.
However, some economists have questioned the assumed importance of fixed capital investment in generating short-term business cycles. Abramovitz has concisely presented a heuristic argument for the position that fixed capital investment is not a fundamental cause of the short cycle.

For a number of reasons, the simpler capital-stock adjustment models with their implied requirements for balanced growth rate take on heightened interest when considered in the context of long swings rather than in that of shorter business cycles. First, insofar as these models treat investment as dependent in part on current or past changes in the demand for finished goods, there has always been justifiable skepticism about their applicability to durable equipment and structures, so long as the theory was supposed to illuminate investment movements in short cycles. Since investment in durables is made for long periods of time it is doubtful whether it would respond readily to income change over short periods. This difficulty disappears, however, when we consider expansions lasting 8 to 12 years or more.*

According to Abramovitz, then, fixed capital investment is unlikely to be an essential factor in generating the short-term cycle, since the delays in capital formation have about the same magnitude as the four-year business cycle and the delays in capital depreciation run much longer. The statement that fixed capital investment is not essential in generating the business cycle has two principle dimensions: first, that business cycles can occur independently or changes in fixed capital investment; and second, that fixed capital variations cannot independently generate four-year cycles.

*Abramovitz [1].
Whole-model testing is used by Mass to evaluate the above hypotheses. He describes a sequence consisting of three basic computer simulations. In the first of these, fixed capital stock is held constant while labor is allowed to vary. The resulting simulation exhibits a four-year fluctuation resembling the short-term business cycle in terms of amplitude, phase relationships between variables, and other characteristics. This simulation indicates that a short-term business cycle can be generated independent of fluctuations in fixed capital investment.

In a second simulation, labor (employment) is held constant while capital stock is permitted to vary. This simulation exhibits a fluctuation of eighteen-year periodicity resembling the so-called Kuznets cycle. The result suggests that variations in fixed capital investment alone, without variations in employment, cannot account for the occurrence of short-term cycles in the national economy. Finally, a third simulation in which both labor and capital are permitted to vary exhibits a four-year cycle in production, employment, and inventories, superimposed on an eighteen-year cycle in fixed capital stock. This third simulation demonstrates that the periodicities of fluctuation associated with labor and capital differ sufficiently that the two periodicities retain their separate existence when labor and capital are combined in a joint production process. Overall, the results tend to indicate that fixed capital investment is not fundamentally involved in generating the short-term business cycle, but principally affects economic cycles of much longer than four years' duration. The detailed implications of this outcome for the conduct of economic
stabilization policy are currently being examined in continuing phases of the study.

A whole-model testing process similar to that described above might be employed to address some of the major controversies in economic theory and policy. For example, most attempts to date to evaluate the monetarist theories regarding economic cycles and stabilization policy have utilized single-equation statistical tests for hypothesis testing. As illustration of such testing, Andersen and Jordan [2] attempt to measure the effectiveness of fiscal and monetary action by regressing changes in income on indices of fiscal and monetary policy. However, for example, as pointed out by Blinder and Solow [3], this approach has the theoretical weaknesses that it ignores the feedback from changes in income to changes in fiscal and monetary policy. This criticism, as well as the analysis developed earlier in the paper, suggests that the single-equation statistical approach is not an adequate approach for assessing the validity of the fiscalist and monetarist viewpoints on macroeconomic management. A more fruitful approach for evaluating the monetarist theories, for example, might be to incorporate the monetarist assumptions into a whole-model context to evaluate their implications, significance, and interaction. Such a model might draw, for example, on Friedman's 1968 Presidential Address to the American Economics Association (Friedman [9]), in which he presents a descriptive summary of some of the main relationships linking money supply, interest rates, GNP, prices and price expectations, and capital investment according to the monetarist view.
Such a model should help to define and assess these assumptions in a more comprehensive framework than has been available to date.

Numerous other potential applications of whole-model testing reside in the areas of psychology, sociology, education, and the other social sciences. The whole-model approach might be used, for example, to interrelate the effects of motivation, expectations, job availability, and social conditions on the effectiveness of education in a school system or community.* Such variables are connected through complex feedback loops of cause-and-effect relationships. In complex problem areas such as economic theory or educational effectiveness, the whole-model approach affords an unprecedented opportunity, and provides a framework for, interlinking variables and testing the effects of various assumptions and theories of social behavior.

*Foster [8] and Roberts [16] present preliminary efforts along these lines.
IV. CONCLUSIONS

This paper has examined two approaches for determining whether or not a given variable should be included in a model. The first method analyzed was the single-equation statistical testing approach. The second approach entailed the analysis of model behavior. The major finding of the paper was that, of the two basic approaches, only behavior tests provide a valid basis for selecting model variables. Only by analysis of model behavior can the modeler ascertain the importance of a particular variable. He can do so by omitting the variable from the model entirely, or by constraining its movement, and examining the consequent shift in model behavior. Model-behavior testing can be used to isolate the influence of an individual variable on a particular historical behavior pattern, on a possible mode of future behavior, or on model response to alternative policies. Section III of the paper presented examples of model-behavior testing at the individual-firm and national-economy levels. In one instance, model-behavior testing was employed to analyze the impact of delivery delays on sales in a model of a growing firm. The testing revealed that delivery delay can be an important factor in generating the pattern of growth instability observed in many firms. In a second example, behavior testing was employed to demonstrate the distinct periodicities of macroeconomic fluctuations associated with labor adjustments versus capital investment. The second example drew on a recent study by Mass [16], in which model-behavior testing showed that, contrary to a considerable body of economic
theory, fixed capital investment is not an essential factor in generating the short-term business cycle.

In contrast to the demonstrated usefulness of behavior testing in assessing the importance of individual model variables, the paper shows that widely-used statistical tests do not provide a valid guide to the selection of model variables. Section II analyzed the t-test and the partial correlation coefficient to illustrate the workings of conventional statistical tests. Both statistical measures are concerned with the contribution of an individual variable X in "explaining" (in a statistical sense) movements in another variable Y. Yet, the discussion showed that neither test actually provides a reliable guide in determining the "importance" of including the variable X. For example, if the parameter estimate associated with variable X fails the t-test, this indicates that the parameter cannot be estimated reliably given the available data, not that the variable X is itself in any way unimportant. Conversely, passing the t-test indicates confidence in the estimated value of the parameter associated with X, but still leaves open the issue of whether or not the variable itself is important.

On the basis of the initial analysis presented in Section II, it appears that conventional statistical tests should be viewed as tests of data usefulness and not as tests of model specification. The data rather than the model are the true subject of statistical testing in the following sense. Failure to pass the t-test indicates that the hypothesized impact of variable X cannot be accurately estimated. Such an outcome should lead the model-builder to either view statistical estimates cautiously or reject them altogether. Either interpretation is tantamount
to rejecting the data as relatively useless for the purpose of estimating the relationship in question. Conversely, passing the t-test indicates a measurable, not necessarily an important, impact of the variable X and implies that the data are useful for the purpose of estimating the hypothesized relationship between the variables X and Y.

In order to illustrate the need for careful interpretation of statistical tests, Section II included a series of experiments with the t-test and partial correlation coefficient. In each experiment, the corporate growth model analyzed in Section III was employed to generate experimental data. The "data" then served as an input to test statistically the impact of delivery delay on sales in the model firm. The experiments showed that the impact of delivery delay on sales was statistically insignificant (low t-test) and that delivery delay showed low partial correlation with sales when the data contained realistic errors. In particular the statistical results were poor when the data contained random measurement errors of 10% or greater. Yet, as noted above, model-behavior testing in Section III revealed that the impact of delivery delay on sales was essential in generating a particular pattern of model behavior. Therefore, if the model-builder incorrectly interprets statistical tests as tests of model specification, he can be led to reject an important influence on overall system behavior.

The results reported in Section II and III of the paper suggest a basic reorientation of theory-testing in the social sciences. Although expert econometricians are well aware of the limitations of statistical testing as a guide to model specification, modeling practice continues to be dominated by the single-equation statistical framework.
The results presented here indicate that the single-equation statistical framework is inadequate, and, if used as the only guide to model specification, can misdirect the search for proper specification. Perhaps one reason for the reliance on statistical tests is the absence of an accepted alternative approach to testing. In this light, the present paper argues that model-behavior testing offers a promising alternative to conventional single-equation statistical analysis.

Of course, model-behavior testing poses new challenges as well. In particular, model-behavior testing requires that one have a system model before the importance of any individual relationship can be tested. This represents a formidable challenge for the model-builder. Yet, the principal conclusion of this paper is that analysis of the mutual interaction between system variables is necessary to ascertain the importance of each individual variable in a social model. Given the major controversies which have persisted within economics and the other social sciences over the validity of alternative theories of social behavior, we should not be too surprised that the necessary prescription entails fundamental reorientation in our testing philosophy.
BIBLIOGRAPHY

[1] Abramovitz, Moses, "The Nature and Significance of Kuznets Cycles," in Economic Development and Cultural Change, vol. 9 (April 1961).

[2] Andersen, L. C., and J. L. Jordan, "Monetary and Fiscal Actions: A Test of their Relative Importance in Economic Stabilization," Federal Reserve Bank of St. Louis Review, vol. 52 (November 1968), pp. 11-24.

[3] Blinder, A. S., and R. M. Solow, "Analytical Foundations of Fiscal Policy," in The Economics of Public Finance (Washington: The Brookings Institution, 1974), pp. 67-70.

[4] Burns, Arthur F., The Business Cycle in a Changing World (New York: National Bureau of Economic Research, 1969).

[5] Duesenberry, James S., Business Cycles and Economic Growth (New York: McGraw-Hill, 1958).

[6] Evans, Michael K., Macroeconomic Activity (New York: Harper and Row, 1969).

[7] Forrester, Jay W., "Market Growth as Influenced by Capital Investment," Industrial Management Review, vol. 9, no. 2 (Winter 1968).

[8] Foster, Richard O., "Education in the City," System Dynamics Group Memorandum D-2144, MIT, Cambridge, June 1972.

[9] Friedman, Milton, "The Role of Monetary Policy," American Economic Review, vol. 58 (January 1968), pp. 1-17.

[10] Hicks, John R., "Mr. Harrod's Dynamic Theory," Economica, vol. 16 (May 1949), pp. 10-121.

[11] Kaldor, Nicholas, "A Model of the Trade Cycle," Economic Journal, vol. 50 (March 1940), pp. 87-92.

[12] Keynes, J. M., "Professor Tinbergen's Method" [a review of J. Tinbergen, Statistical Testing of Business Cycle Theories: A Method and its Application to Investment Activity (Geneva: League of Nations, 1939)], Economic Journal (September 1939), pp. 567, 569.

[13] Mass, Nathaniel J., Economic Cycles: An Analysis of Underlying Causes (Cambridge: Wright-Allen Press, 1975).
[14] Mass, Nathaniel J., "The Dynamics of Economic Fluctuations: A Framework for Analysis and Policy Design," in *Proceedings of the 1975 Summer Computer Simulation Conference*, San Francisco, July 1975.

[15] Morgenstern, O., *On the Accuracy of Economic Observations* (Princeton: Princeton University Press, 1963).

[16] Roberts, Nancy, "A Computer System Simulation of Student Performance in the Elementary Classroom," *Simulation and Games*, vol. 5, no. 3 (September 1974), pp. 265-289.

[17] Samuelson, Paul A., "Interactions Between the Multiplier Analysis and the Principle of Acceleration," *Review of Economics and Statistics*, vol. 21, no. 2 (May 1939), pp. 75-78.

[18] Senge, Peter M., "Testing Estimation Techniques for Social Models," System Dynamics Group Working Paper D-2199-4, Alfred P. Sloan School of Management, MIT, Cambridge, 1975.

[19] Theil, H., *Principles of Econometrics* (New York: John Wiley and Sons, 1971).
