Bureau of Business Research
Graduate School of Business Administration
University of Michigan

October 1971

THE IMPLICATIONS AND IMPLEMENTATION OF EXPLANATORY MODELING IN CORPORATE PLANNING

Working Paper No. 50

by

William K. Hall

Assistant Professor of Statistics
and Management Science
The University of Michigan

FOR DISCUSSION PURPOSES ONLY

ABSTRACT

A major problem in the development of corporate planning models is the determination of structural relationships which quantitatively relate planning decisions to measures of future organizational performance. In a large number of contemporary planning models, these relationships are estimated by conducting statistical analyses on data obtained from the firm's accounting information system. The intent in this paper is to illustrate some problems which can arise in using such an approach and to propose some alternative model-building methodologies. Methodologies are developed for situations where the underlying causal associations between decisions and performance measures are known, in which case time series analysis techniques can be utilized to estimate the desired structural relationships. In addition, methodologies are developed for situations in which the underlying causal associations are unknown. For these cases structural decomposition techniques and analyses of data obtained from special-purpose information systems can be utilized to make inferences on those relationships which are of interest. The techniques are applied to operations cost analysis and revenue forecasting to illustrate the underlying concepts.

BACKGROUND FOR THIS PAPER

This paper is based upon remarks delivered at the 12th American Meeting of the Institute of Management Science, Detroit, Michigan, October 1, 1971. The research was undertaken in conjunction with an informal research program on organizational planning within the Graduate School of Business Administration at the University of Michigan.

The utilization of models in the corporate planning process is the subject of much current discussion and activity. From a recent survey, Gershevski (5) concluded that one hundred corporate models were in various stages of development in late 1969. In the past two years, this effort has been significantly expanded--planning models are currently being developed by a large number of banking, utility, and manufacturing firms of varying sizes and characteristics.

A simplified conceptualization illustrating the basic components of a typical corporate planning model is presented in Figure 1. Associated with this concept are several model design questions,* notably:

1. Should the model develop an "optimal" set of planning decisions or instead evaluate alternate planning decisions?
2. Should the model be deterministic or probabilistic?
3. What level of detail (in terms of organizational and functional activity) should be incorporated into the model?
4. How should the model relate the planning inputs to various measures of future corporate performance (planning outputs)?

This paper will consider the fourth question in some detail. The major premise advanced here is that the statistical techniques for resolving this issue which have been embedded in a large number of contemporary models can result in inadequate and misleading relationships. Consequently the intent in this paper is to illustrate some problems arising from the use of these contemporary approaches and to propose some alternative model-building methodologies.

---

*The reader is referred to Gershevski (4) or Barkdoll (3) for a more complete discussion of these questions.

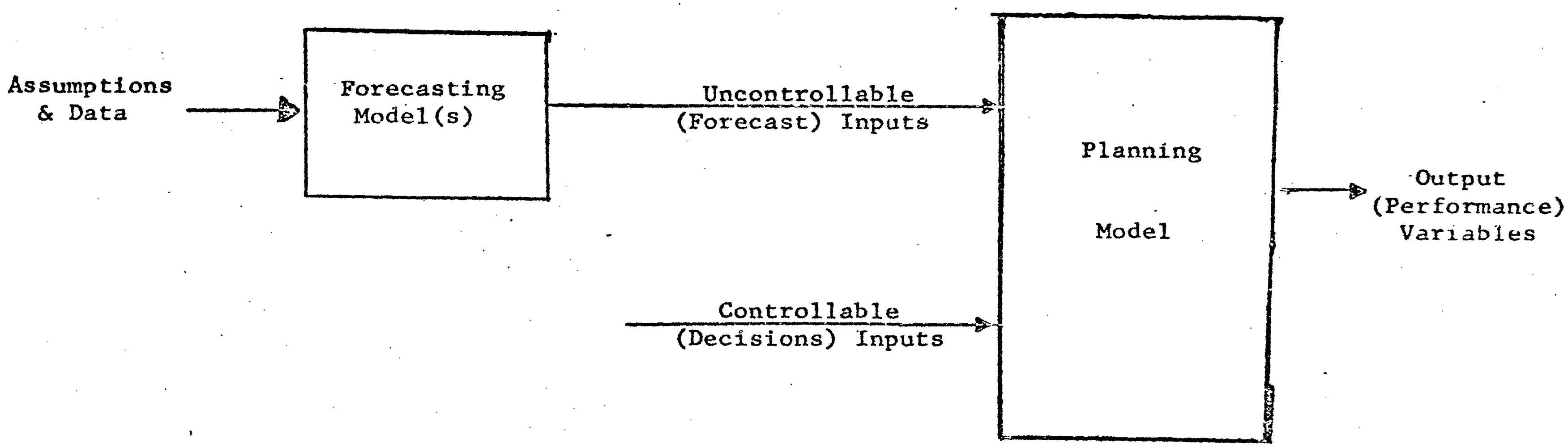

Fig. 1. Schematic conceptualization of a typical planning model.

The traditional statistical approach to this problem can best be summarized by quoting from Gershevski[1/]:

> An (input-output) diagram provided a broad conceptual framework which showed how the various functions of the company were interrelated and how key variables affected performance with each area. In this manner we devised a complete set of simple algebraic equations to project and consolidate all costs and revenues... In our case, multiple regression analysis proved to be very useful in producing equations which produce accurate forecasts. It involved finding the correct independent variables and adding them to existing equations to form more complex and accurate ones.

This utilization of multivariate statistical methods (notably multiple regression) in developing input-output relations for planning models has significantly influenced the process of model development in a large number of corporations. Indeed, many planning model computer "packages" currently offered by computer software firms include statistical analysis subroutines to aid in the development of input-output relations, leading to what the promoters call "instant models."

To illustrate some of the problems with these statistical model-building techniques, typical model relationships will be developed from an actual set of corporate accounting data. These data, presented in Table 1, consist of information (slightly disguised) taken from the accounting records of a large, NYSE-listed two-product company operating in the midwestern United States.

Two separate situations will be considered:

1. Model building when the underlying cause-effect relationships between inputs and outputs are understood (well-structured situations).
2. Model building when the underlying cause-effect relationships between input and outputs are not understood or are only partly understood (ill-structured situations)

1/ Gershevski (4), pp. 65-66.

Table 1. **Basic Financial Data - Company X**

Year	Operations Expense (in millions)	Wage & Salary Expense (in millions)	Number of Employees	Sales Product 1 ($X10^{-4}$)	Sales Product 2 ($X10^{-4}$)	Average Price Product 1
1970	$32.48	$12.80	1,196	18.81	33.40	178
1969	28.04	10.88	1,164	18.48	30.22	166
1968	24.67	10.08	1,122	17.61	27.33	162
1967	23.01	9.34	1,084	16.75	25.74	161
1966	21.62	8.65	1,050	15.87	22.33	162
1965	19.73	7.99	998	14.30	20.53	168
1964	17.71	7.47	975	12.46	18.52	178
1963	16.75	7.14	973	11.60	16.78	180
1962	15.70	6.91	974	10.85	15.68	184
1961	14.67	6.75	991	9.81	13.92	190
1960	14.40	6.51	1,004	9.55	13.54	189

(Continued)

Table 1 (cont.). Basic Financial Data - Company X

	Average Price Product 2	Sales Revenue Product 1 (In millions)	Promotional Expense Product 1 (In millions)	Disposable Personal Income (In Billions)
1970	$82	$33.5	$1.54	$ 684.8
1969	79	30.7	1.43	631.6
1968	80	28.6	1.35	590.0
1967	80	27.0	1.29	546.5
1966	85	25.8	1.20	511.9
1965	84	24.0	1.11	473.2
1964	83	22.2	1.04	438.1
1963	84	20.9	1.00	404.6
1962	83	20.0	.93	385.3
1961	82	18.6	.91	364.4
1960	75	18.1	.85	350.0

MODEL BUILDING IN WELL-STRUCTURED SITUATIONS

The first situation prevails in a large number of input-output relationships, notably those involving cost analysis. For instance, it is generally true that operati costs increase because of increased volume of production. The objective in modeling this relationship then becomes one of determining the extent of these increases.

One way of developing these relationships for our sample company is to postulate a general model:

Operations = f (Production Volume Product 1, Production Volume Product 2)
Cost

Although the form of the cost-volume functional relationships faced by a firm is the subject of much debate in economics, 2/ managerial economists typically assume that the relationship is approximately linear with fixed and variable cost components. When the parameters of this linear relationship are estimated using least-squares regression on the data of Table 1, the following relation is obtained.*

$$\underset{}{OC_t} = \underset{(1.192)}{4.489} - \underset{(.296)}{.686}\, V1_t + \underset{(.153)}{1.203}\, V2_t$$

where

OC_t is the total operating cost for the firm in period t (millions of dollars

$V1_t$ is the production volume for product 1 in period t (in 10,000)

$V2_t$ is the production volume for product 2 in period t (in 10,000)

All coefficients in the estimated relation are statistically significant at the .05 level. Moreover, the coefficient of determination (R^2) is .989, indicating that almost 99% of the variation in historical costs can be explained by the estimated linear relationship.

However, even though this relationship demonstrates a high degree of statistic precision, its utilization in a plannning model is extremely questionable (even if the modeler is willing to assume that trends in the historical data are likely to prevail in future periods). This is because of the negative variable cost coefficient for

2/ See, for instance, Johnston (6).

* Standard errors of the estimated parameters are shown in parenthesis.

Product 1. If the modeled relationship is used to project the effect of changes in volume on costs, the conclusion one draws is that increases in the production of Product 1 reduce the firm's operating costs (at least over the regions of historical output). In reality, such a relationship is absurd for the product under consideration. Instead the misinterpretation arises when regression coefficients based on observational data are utilized to assess the magnitude of a causal relationship. The dangers of this common misinterpretation of a statistical relationship are nicely summarized by Ackoff and Sasieni:3/

> Usually we are interested in what happens to one variable if we succeed in controlling the other, and for this we require a causal relationship. Unfortunately the existence of a linear (regression) relationship tells us nothing about causality. If we can rule out the possibility of a non-linear relationship and obtain a correlation coefficient near zero, we can be fairly sure that no causal relationship exists, but the converse is not true.

These difficulties in determining the magnitude of causation from aggregate cost data suggest that decomposing operating costs into additive components and deriving relationships for these components individually should improve the modeling effort. Since wage and salary costs form one component of operating costs, we shall now examine these costs in greater detail.

One attempt to model public utility wage and salary costs from historical accounting data is provided by Wrobleski & Achtenberg (8). Here the model is developed by regressing wage and salary costs on total electric power generation, plant investment, and the number of employees. Even though this model has high predictive power (as measured by the statistical goodness of fit), it is still impossible to use this model to assess the effects on wage and salary expenses of changing operating and investment policies.

3/ Ackoff & Sasieni (1), pp. 386-87.

To accomplish this assessment, an exceptionally simple "explanatory" model can be created which exactly incorporates the underlying causal relationships between wage and salary expenditures and the number of employees in the firm: $WS_t = F_t + W_tN_t$*

where: WS_t = wage/salary expenses in period t (millions of dollars)

F_t = Fixed employee expenses in period t

W_t = Average variable wage rate per employee in period t

N_t = Number of employees in period t

Since this model is linear in the number of employees, it is natural to use linear regression to estimate F_t and W_t from the historical data of Table 1. Such a procedure yields $F_t = -16.206$ and $W_t = .0236$, with a correlation coefficient of .963.

In this situation the statistical model provides reasonably good predictive power within the historical wage rate and employment levels. Still, it is unlikely that future wage rates and employment levels will fall within these ranges , and since the regression model estimates coefficients on the basis of these historical averages it is likely to underforecast future wage and salary expenditures.

An alternative which avoids this problem is to forecast average wages per employee directly, taking into account the nonlinearities in this series. For example, suppose that fixed wage/salary expenses are assumed to be negligible and that the following scheme is used to project average annual wage and salary expenditures:

1. Eliminate any potential (linear) trend by taking first differences in the historical series of average wages per employee.
2. Forecast the first differences one year ahead by using a simple three-year moving average.
(3) Reconstruct the time series to develop a one-year wage and salary expenditure forecast.

---

*A more sophisticated model could further disaggregate wage and salary expenditures into costs for each employee class.

The results of this procedure are shown in Table 2, where these forecasts are compared with the actual expenditures and the forecast expenditures derived by using the regression model discussed earlier. These results are then plotted in Figure 2. From these two figures it is apparent that the intrinsic time-series forecasting procedure is substantially more accurate than the regression procedure over the range of the historical data.*

Another alternative in forecasting average wage per employee values is to develop projections based upon the anticipated results of labor negotiations or employee compensation policies. When major changes have occurred in these factors it is clear that historical values should be adjusted to reflect this additional information. The point in posing this alternative, of course, is that subjective techniques may be superior to more sophisticated quantitative methodologies in those cases where historical situations are unlikely to prevail in the future.

MODEL BUILDING IN ILL-STRUCTURED SITUATIONS

In an exceptionally large number of situations the underlying cause-effect relationships between corporate inputs and outputs are at best poorly understood. Consider, for instance, the relation between sales revenues and various promotional and pricing policies.

Here again one approach to the development of this relationship is to develop statistical models from the historical accounting data in Table 1.

Consider, for example, a model which projects the sales revenue from Product 1 as a function of promotional expenditures on the product. A linear regression model based upon the relevant historical data yields:

$$R1_t = -.97 + 22.1\ PR1_t$$
$$(.64) \quad (.54)$$

---

* The fact that the intrinsic procedure minimizes the forecast error for historical values does not indicate that this procedure will be superior for future values. However, the capability of the technique to rapidly adapt to changes in the series while "smoothing" short-term (high-frequency) variations should provide accurate short-range forecast values.

Table 2. Comparisons of Forecasting Procedures: Wage & Salary Expenditures (millions)

Year	Actual	Forecast 1 (Regression on number of employees)	Forecast 2 (Using moving average technique)
1961	6.51	7.55	6.61
1962	6.75	7.24	6.76
1963	6.91	6.84	6.91
1964	7.14	6.81	7.17
1965	7.47	6.86	7.44
1966	7.99	7.40	7.94
1967	8.65	8.63	8.72
1968	9.34	9.44	9.25
1969	10.08	10.34	10.01
1970	10.88	11.33	10.80
1971	12.80	12.09	11.58

Forecast 2 - Compute average annual wage per employee
Take first differences
Forecast first differences using 3 period moving average
Reconstruct series to forecast annual wage and salary expenditures

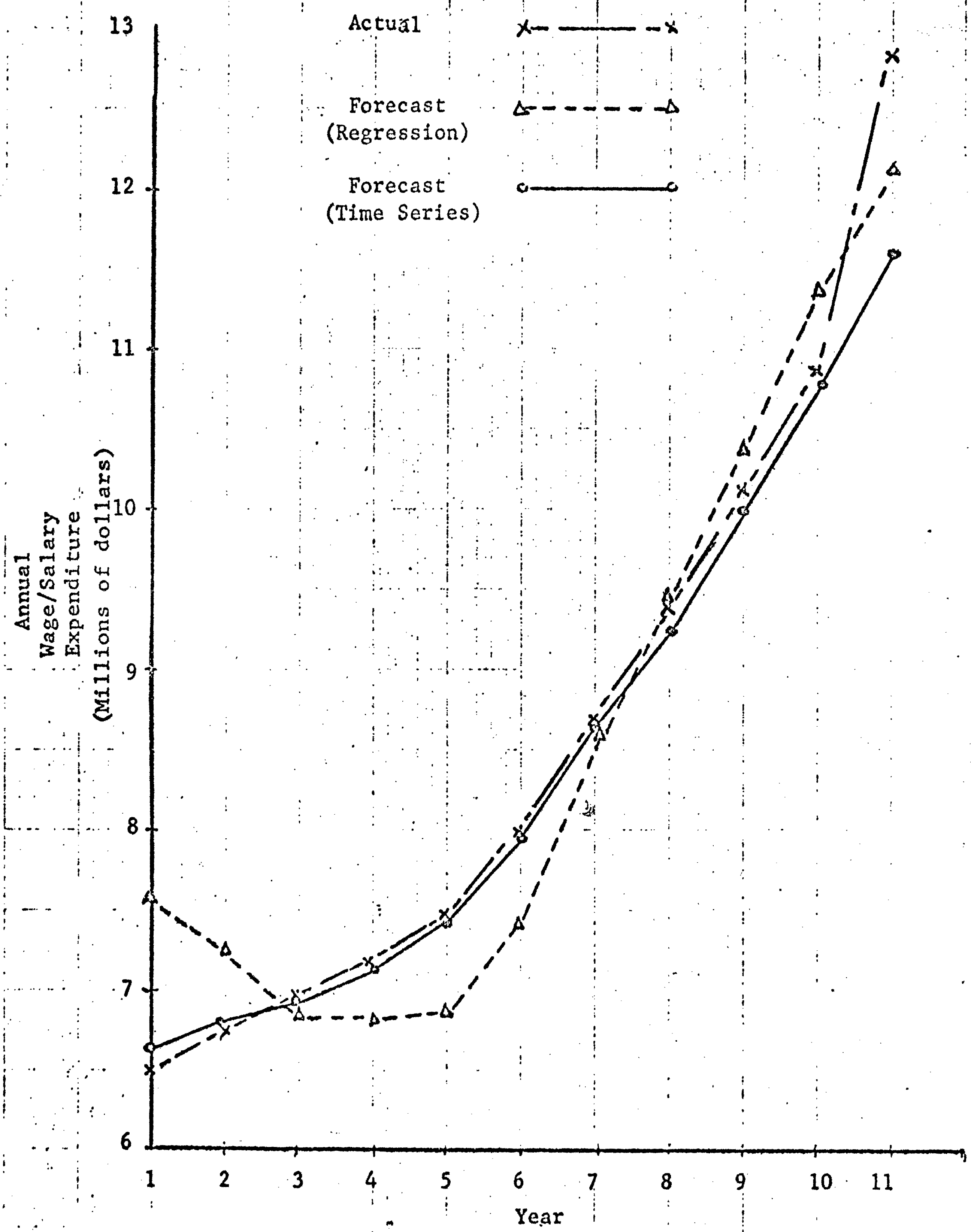

Fig. 2.
Comparison of Various Wage/Salary Forecasting Techniques

where $R1_t$ = Sales revenue for Product 1 in year t (Millions of dollars)

$PR1_t$ = Promotional Expenditures for Product 1 in year t (Millions of dollars)

An erroneous assessment of this model* would lead one to conclude that the marginal revenue generated by promotional expenditure is 22.1 over the range of historical operations. Since this is implausible, the analyst might attempt to develop a more sophisticated model which separates sales volume variations from sales price variations. A linear regression model for this situation yields:

$$V1_t = \underset{(1.6)}{13.06} - \underset{(.008)}{.076}\, P1_t + \underset{(.36)}{12.53}\, PR1_t$$

where $P1_t$ = Price of Product 1 in year t (dollars)

and the other variables are defined as in the earlier examples. The coefficient of multiple correlation in this case is again exceptionally high (.999).

One further example can be developed by adding an exogenous variable to aid in the explanation of sales volume. In this company, for instance, management might assume that disposable personal income has a positive influence on demand. The resulting linear model (with coefficients estimated by least squares regression) is given by:

$$V1_t = \underset{(.91)}{15.96} - \underset{(.004)}{.081}\, P1_t + \underset{(2.14)}{.70}\, PR1_t + \underset{(.004)}{.024}\, J_t$$

where J_t = disposible personal income in period t (billions of dollars)

Here the coefficient of multiple correlation is increased to a phenomenal .9999. Moreover, the coefficient of promotional expenditure is not significantly different from zero, indicating that there is no marginal revenue generated by increased promotional expenditures over the range of historical operations.

The problem with all three of these statistical models is that, regardless of the goodness of fit, the models cannot be used to assess the effects of changes in

*The coefficient of correlation here is exceptionally high (.997), indicating that this model fits the data with high precision.

the independent (controllable) variables. Thus, although the models may be very useful in predicting future sales volumes (based upon predictions of future values of the independent variables), they are of absolutely no use in projecting the effects of changes in promotional and pricing strategies.

The following quotation from de Neufville (7) supports this point and the earlier observations of Ackoff:

> The statistical closeness of an equation to a set of observational data on a system is not a sufficient test of its validity. In fact, statistical correlation by itself may not even be a good guide to the choice of a systems model.... Indeed, an example can be reputed to fit existing data quite closely even though it is actually quite opposite to the valid causal model.

To conclude this discussion it is interesting to note that an "exact" causal relationship between sales revenues and promotional expenditures exists for these data:

$$PR1_t = .05\ R1_{t-1}$$

That is, corporate budgetary policy sets promotional expenditure at 5 per cent of previous years' sales, a completely opposite causal hypothesis from that proposed (and supported) by the earlier statistical models.

The question remaining, of course, is how the planning analyst can develop causal relationships which are useful in evaluating alternative strategies. One alternative is to postulate such relationships directly through deductive reasoning. The classical S-shaped function relating sales response to advertising expenditure provides one simplified example of this procedure. The major problem with this analytical procedure is that the form and magnitude of relationships are exceptionally difficult to assess in a direct fashion.

One exceptionally useful approach to this assessment is to decompose the relationship of interest into a sequence of interrelated causalevents. This approach is usefully illustrated by developing a "causal" chain relating advertising expenditures to future sales volumes. A sample resulting chain is conceptually illustrated in Figure 3. The advantages of this approach are twofold:

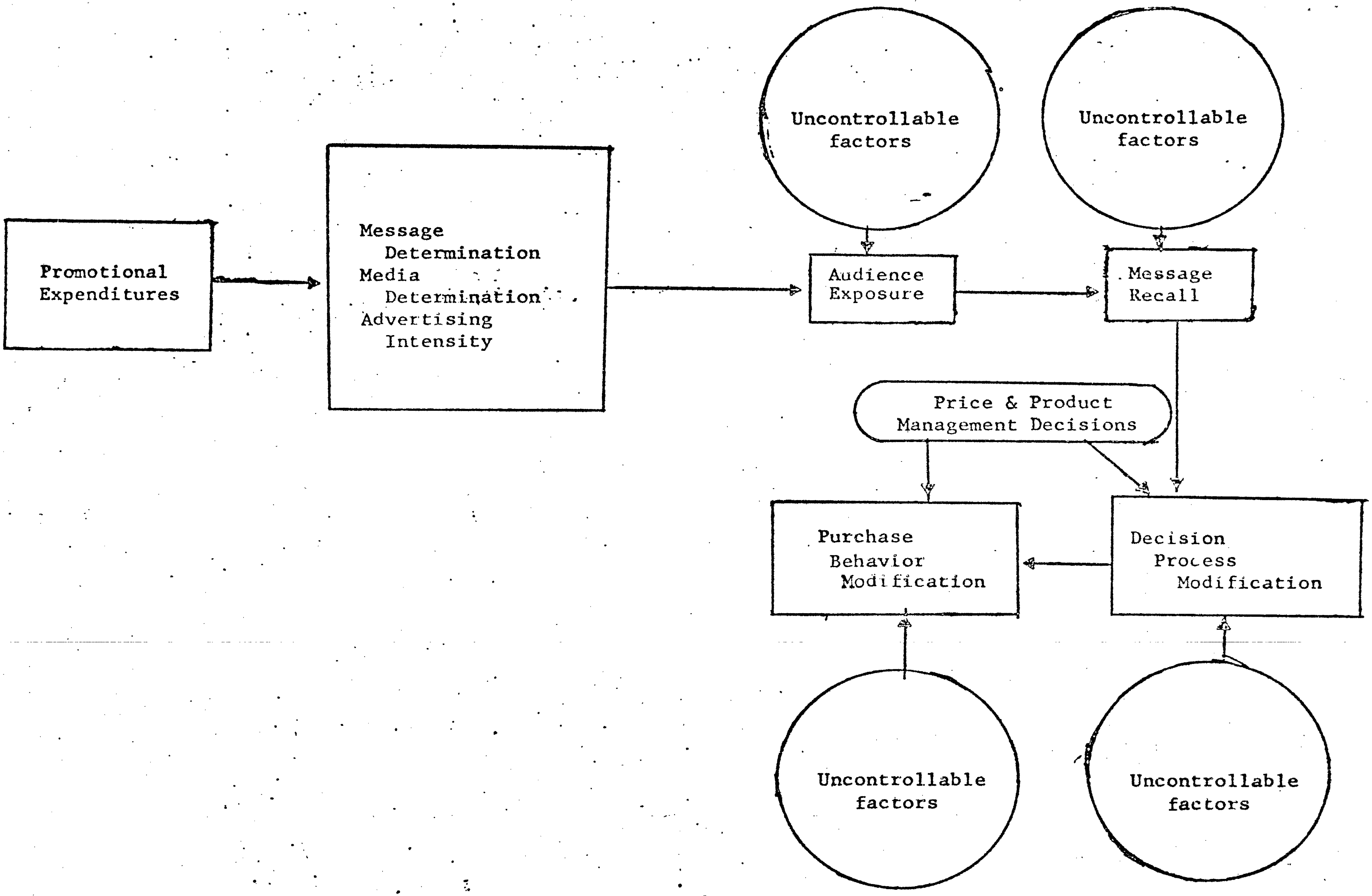

Fig. 3. "Causal chain" for promotional strategy evaluation.

1. The causal chain identifies the sequential changes which must occur before promotional expenditure can influence purchase behavior.
2. The chain identifies the influence of uncontrollable (confounding) variables on this behavioral process.

Within the overall chain, various submodels and experimental techniques can be utilized to quantify the extent of the intermediate relationships. For instance, one can conceive a well-designed experiment to determine how the selection of media mix affects audience exposure to a promotional campaign. Alternatively, special purpose market information systems can be utilized to monitor the extent of changes in customer exposure and behavior as the result of a change in promotional policy. In the event that positive relationships are detected, the results of these various efforts contribute to the development of an overall promotional effectiveness model. On the other hand, the failure to detect intermediate relationships provides an indication that subsequent effects of the promotional strategy are nonexistent and consequently that the strategy should be revaluated.

SUMMARY AND IMPLICATION

The primary intent in this paper has been to illustrate some problems which arise when statistical techniques are used to make inferences on causal relationships in corporate planning models. The point raised here is not new; for years statisticians have pointed out the problems in inferring cause and effect from observational data. Still, the use of such procedures has recently been promoted in a wide variety of corporate planning situations.

The major alternative proposed in this paper is to decompose the relationship of interest into a series of additive or sequential relationships. Various strategies are then available to develop these intermediate relationships in both well-structured and ill-structured situations.

One advantage of these strategies is that they avoid the erroneous causal inferences which can be drawn from statistical models. A more fundamental advantage is that they encourage increased understanding of the underlying input-output relationships which characterize an organization. That is, the emphasis is on the process of modeling rather than on the results of modeling.

The importance of this process-oriented view of corporate modeling cannot be overemphasized. In an age of rapidly changing organizational and structural relationships within the firm, it is likely that the insights obtained during the development of planning models may become the biggest of all the "payoffs" from the developmental effort. If this is true, those in the management science profession must learn to improve the quality and implementation of these insights, rather than concentrating on the development of more sophisticated "instant models." As Ackoff (2) has thoughtfully observed, "The principal contribution of scientists to planning may not lie in the development and use of relevant techniques, but rather in their systemization and organization of the planning process, and in the increased awareness and evaluation of this process that their presence produces."

References

(1) Ackoff, R.L., and Sasieni, M.W., Fundamentals of Operations Research, Wiley, 1968.

(2) Ackoff, R.L., A Concept of Corporate Planning, Wiley, 1970.

(3) Barkdoll, G.L.,"Using Financial Models to Improve Communications," Managerial Planning, May-June 1971, pp. 2-4.

(4) Gershewski, G.W., "Building a Corporate Financial Model," Harvard Business Review, July/Aug. 1969.

(5) Gershewski, G.W., "Corporate Models -- The State of the Art," Managerial Planning, November 1969.

(6) Johnston, J., Statistical Cost Analysis, McGraw-Hill, 1960.

(7) de Neufville, R., Systems Analysis for Engineers and Scientists, McGraw-Hill, 1970.

(8) Wrobleski, W.J., and Achtenberg, D.C., Corporate Financial Modeling: Systems Analysis in Action, Working Paper No. 21, Bureau of Business Research, The University of Michigan, 1971.