use the cost equations were designed to be consistent with the financial and accounting conventions at the Times. In these terms, the model can be viewed as a rather large and sophisticated income statement or operations model. Consequently, capital stock appears only indirectly through the depreciation equations. Thus, while the fixed-variable cost specification is retained capital stock costs do not explicitly appear.

Variable costs are defined as those costs which vary directly with paging and copies produced; newsprint and production labor costs. The number of men working and the number of hours worked vary each day depend on paging size and copies produced. Although capital is held fixed, capital utilization is not. Theoretically, depreciation is a function of capital utilization and should be treated as a variable cost. Because accounting conventions do not handle depreciation in this fashion, in the model, depreciation does not vary with output.

Newsprint costs are computed by calculating the newsprint tonnage required to produce a given number of pages and copies and then multiplying by the appropriate cost per ton. There are equations for different grades of newsprint as well as equations for the many different sections of the paper.

Econometric cost equations are used to determine variable labor costs for each of the variable labor departments. While the equations differ, in general they look as follows:

$$LC_{it} = F_2(C_t, P_t, TC_t, W_{it}) \qquad (2)$$

where LC_{it} is unit labor cost for the i'th group in period t

C_t, P_t, TC_t are as defined above

W_{it} is the appropriate wage rate

Changes in production technology are rapidly making obsolete historical production and cost relationships. Eventually, these econometric cost equations, as mentioned earlier, will be replaced by an optimization analysis. For example, a given volume forecast of pages and copies can be produced in a variety of press configurations, each press configuration completely determines labor requirements. Programming techniques will be used to determine the least-cost press configuration. The labor associated with this press configuration will then be monetized at the appropriate wage rates.

To complete the cost side, non-capital fixed costs are divided into labor and non-labor fixed costs. Because the accounting system does not track these costs historically econometric methodology is not used. Fixed labor costs are determined by equations which compute manning requirements and wage rates in 32 different categories.

In the current version of the model, manning requirements are determined exogonously. They are policy variables controlled by management. Research is currently underway to determine the linkages between fixed labor, output, and profitability; and formalize the decision rules management intuitively follows in determining annual fixed manning levels. The objective is to fully endogenize these linkages.

Non-labor fixed costs (excluding capital) are treated in a very simplistic way. Dollar cost for each category is increased each period by a growth factor. These costs account for a very small percentage of fixed costs.

The demand model is a simultaneous block of non-linear econometric equations for advertising lines and copies sold (circulation). Both outputs are sold in oligopolistic markets with varying product differentiation and intense non-price competition. Because the Sunday and daily papers are distinct products, Sunday and weekday advertising and circulation are treated separately. Conceptually, the demand for each output is treated as a function of its price, the prices of its competitors, indicators of market activity, and quality. Symbolically, the demand for circulation looks like the following:

$$CIRC_{ti} = F_3(p_i, p_{ji}, Prom, M_1 \ldots, M_n, Q) \qquad (3)$$

where $CIRC_{ti}$ is average circulation in period t for the i^{th} market

p_i is the unit price per copy in the i^{th} market.

p_{ji} is the price of the competitive product in the i^{th} market

Prom is promotion expenditures

M_i are indicators of market activity

Q is quality

The variable quality is a measure of the appeal of the paper to its readers. Potential readership depends on attitudes which in turn depend on education, household income, occupation, orientation to NYC, politics.

Similarly, the demand for advertising lines;

$$L_{ti} = F_4(p_i, p_{ji}, Prom, Q, M_1, \ldots, M_n) \qquad (4)$$

where all variables are defined as above and L_{ti} is advertising lines for the i^{th} category in period t. The variable Q represents the quality of a unit of linage to the advertiser as measured by the purchasing power of its readers, prestige of the newspaper, and readership Generally, circulation can be used as a proxy variable for advertising quality. The conceptual framework for the model is summarized in fig. 1. This flow diagram depicts the dollar flows between the various components of the system. The underlying physical flows are

SIMULATION MODEL: CONCEPTUAL FRAMEWORK

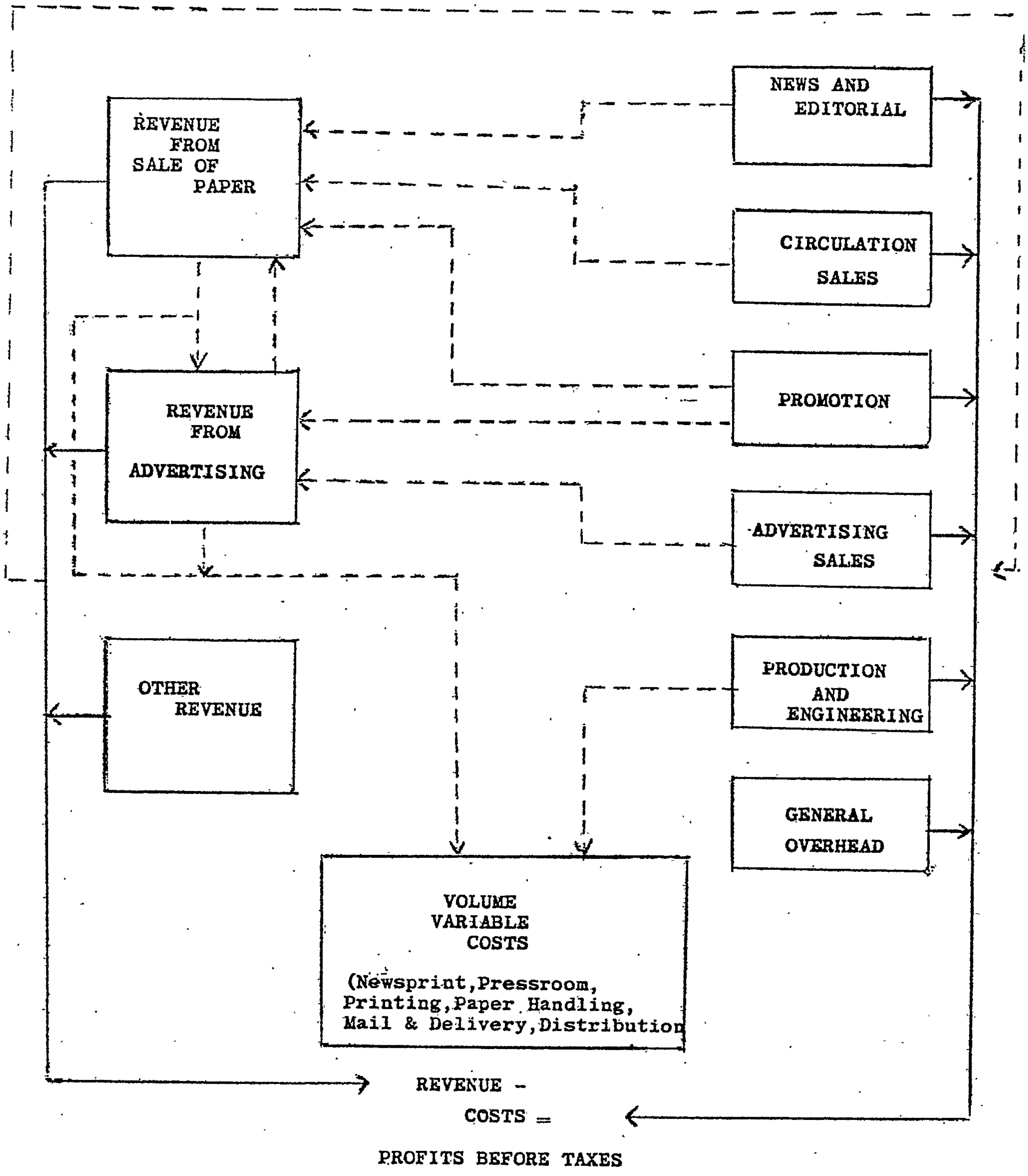


FIG. 1

NEW YORK TIMES MODEL STRUCTURE

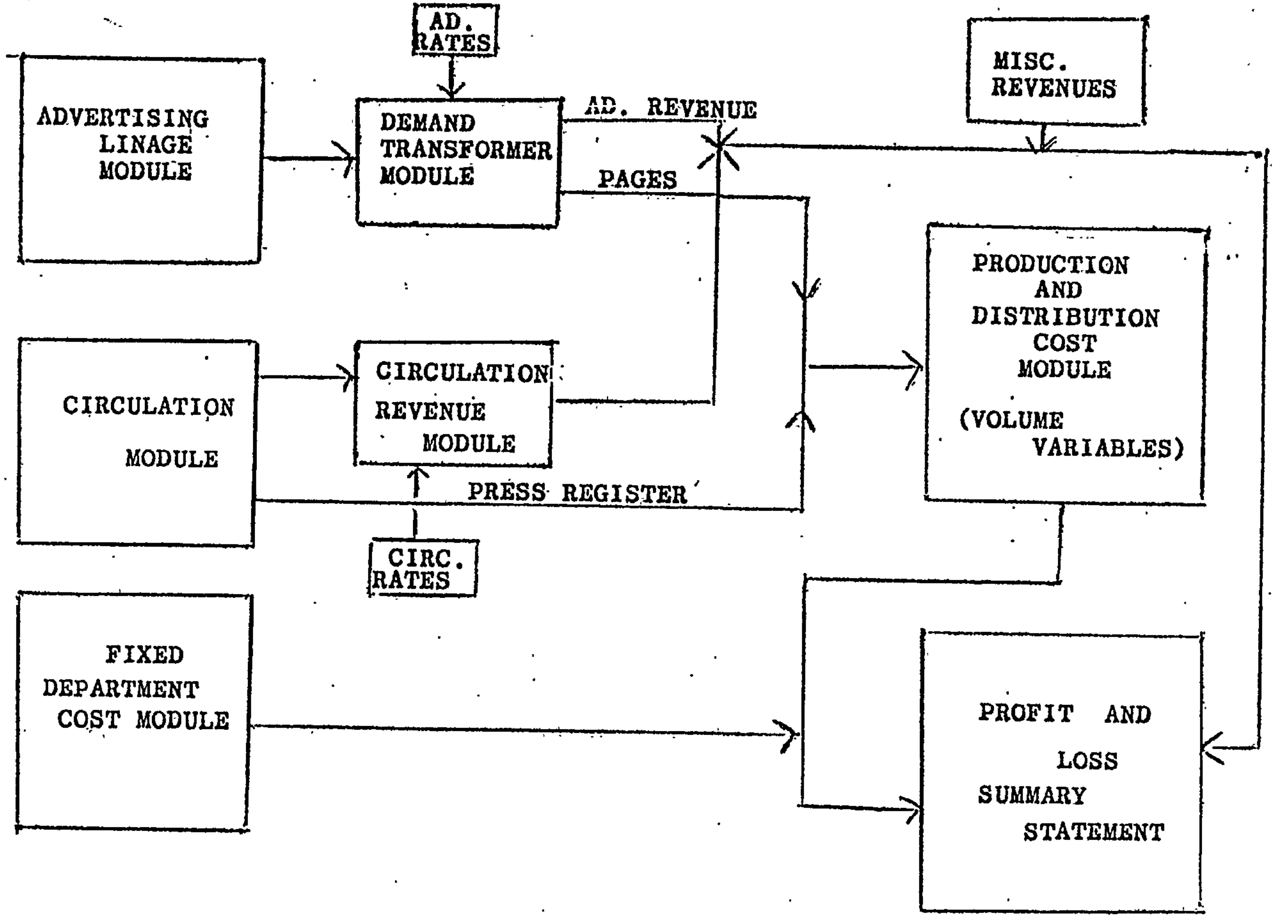
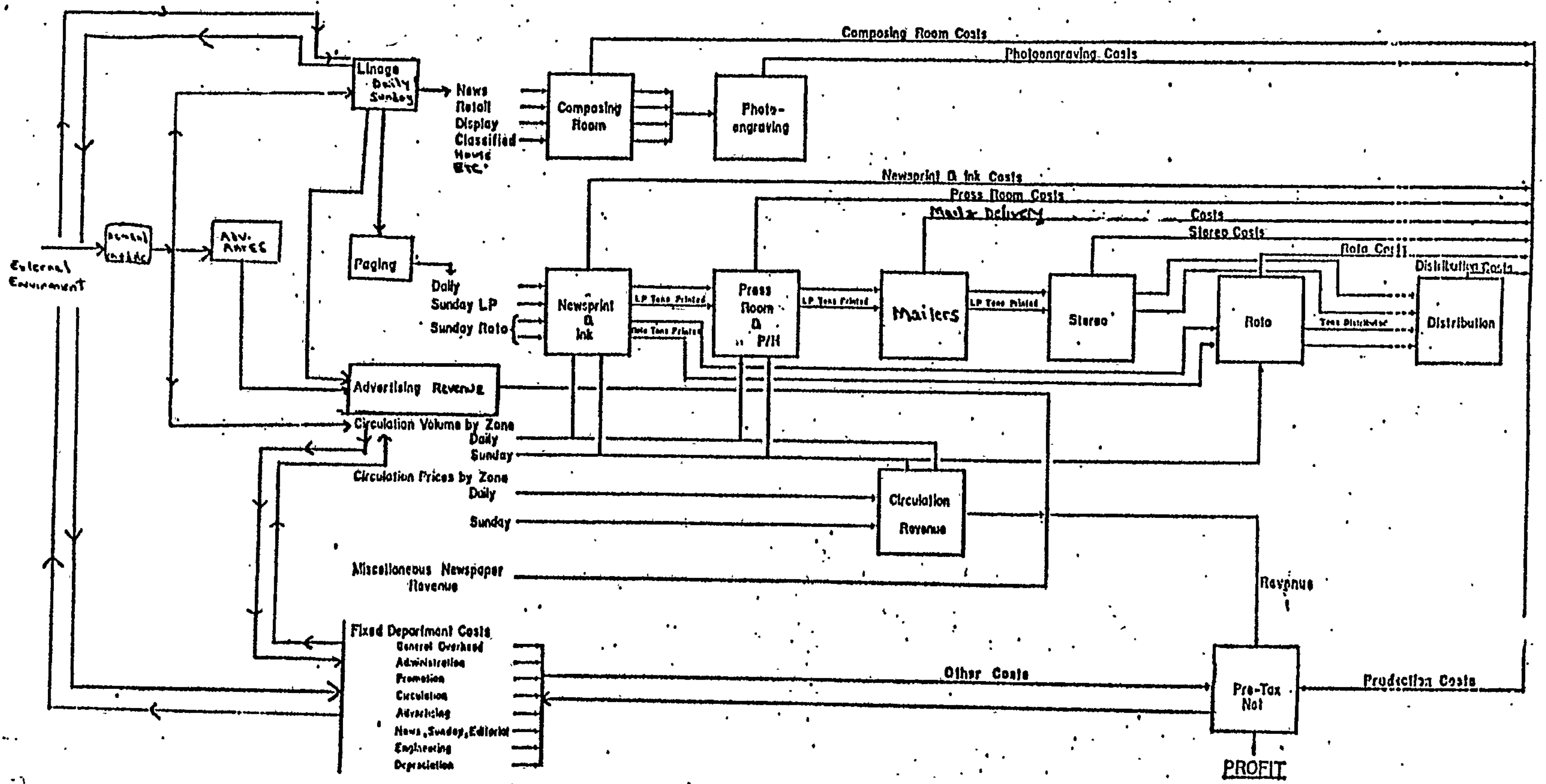


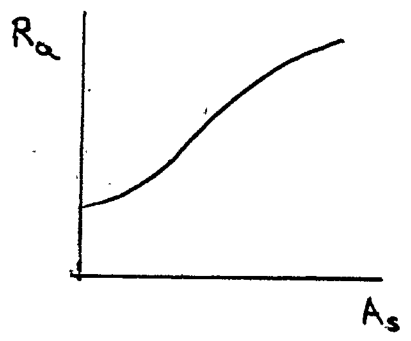
FIG. 2

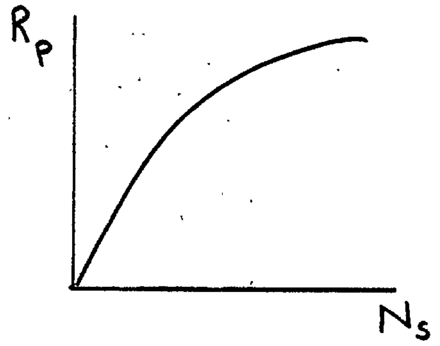
FIG. 3

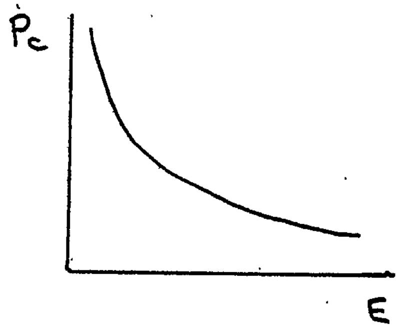
PRODUCTION - DISTRIBUTION COST MODULE

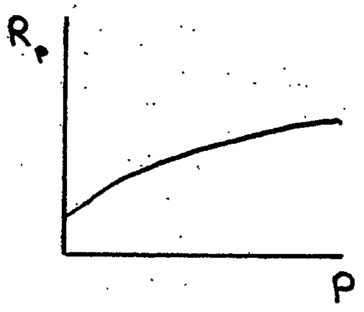


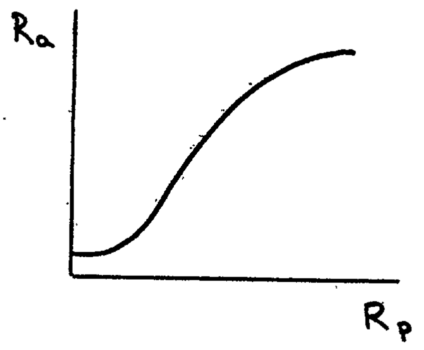
shown in figs. 2 and 3. The solid lines in fig. 1 represent dollar flows, revenues on the left and costs on the right. The dotted lines portray the interaction between these dollar flows with the arrows indicating the presumed direction of the interaction. For example, expenditures for promotion are assumed to influence advertising and circulation revenues. Conversely, advertising and circulation revenues impact the dollars available for promotion expenditures. The likely nature of some of these non-linear interactions are shown in the accompanying graphs.

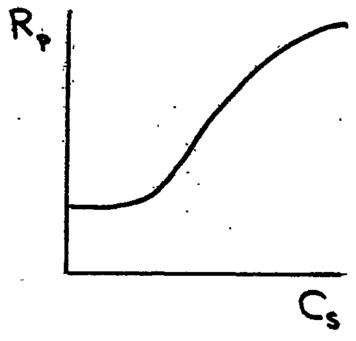


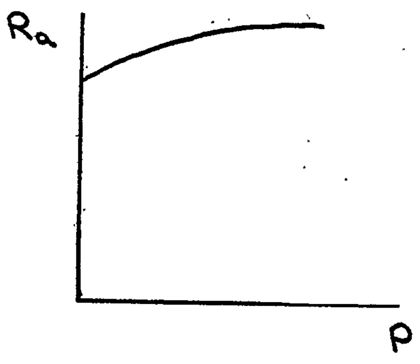












where R_p is circulation revenue
R_a is advertising revenue
P_c is production costs
N_s is news and editorial costs
C_s is circulation sales costs
P is promotion costs
A_s is advertising sales costs
E is engineering costs

Thus, real increases in news and editorial expenditures are likely to have a positive effect on the sale of newspapers although diminishing returns are probable. Similarly, if circulation sales expenditures were reduced to zero revenues from paper sales would probably drop but not to zero. The revenue loss would occur gradually not instantaneously. The flow diagrams and graphs are conceptual simplifications of complex time-dependent interrelationships.

None of the fixed department relationships are as yet built into the model. I suggested earlier that management has a rational belief concerning the nature of these interactions which they act on in determining manning and expenditure levels. As the graphs above show, these relationships can be defined mathematically and built into the model. Simulation experiments could then be performed to test the effects of various fixed department hypothesises and determine those relationships which cause simulated performance to agree with actual performance.

Shown in fig. 2 is modular structure which is a convenient basis for thinking about the model although this structure does not correspond exactly to the way the system is modeled. For purposes of discussion, each of the boxes can be defined as a module. The modules are:

1. Demand module. This represents forecasts of the demand for advertising by category. Either manual forecasts by the advertising sales force or projections from the model can be used.
2. Circulation module. This represents the manual circulation department forecasts or the projections generated by the model.
3. Overhead and fixed department module. The current exogonous manning forecasts are used to compute fixed labor costs. Non-labor fixed costs are computed in this module also.
4. Demand Transformer. Given advertising lines and rates by category, this module generates news lines, paging and advertising revenue.
5. Circulation revenue module. Forecasts of circulation by region are used to generate revenue and copies produced.
6. Miscellaneous revenues. Exogenous projections of other revenues.
7. Production and distribution module. Given total pages and copies produced, this module builds up the costs of printing and distributing the paper.

A more detailed flow diagram, which corresponds more closely to the mathematical structure of the model is shown in fig. 3. This diagram details the major physical flows of the model and the financial output which results. This financial information is summarized in an income statement for management.

III THE CIRCULATION AND ADVERTISING MODEL: SPECIFICATION AND ESTIMATION

The generalized specification of the advertising and circulation equations shown earlier briefly outline the theoretical structure of the demand model . The major departures from that structure are the absence of relative prices in the advertising equations and demographic variables in the circulation equations.

Preliminary analysis indicated that within the normal range of price change there is a high degree of price inelasticity. This finding is consistent with the economic theory of product differentiated markets. Competition between the Times and other newspapers and electronic media exists but it is based on factors other than price. I do not mean to suggest that price changes do not affect advertising sales. They do. However, price competition is minimal.

Unfortunately, data problems prevented us from including the competitive factor in the model. Ideally, market share type equations would be preferable. However, because of data problems their reliability would be suspect. In any event, changing economic conditions proved to be the major cause of cyclical swings in advertising lines.

In the circulation equations, the lack of meaningful quarterly time series demographic data precluded their inclusion as explanatory variables. There are plans to remedy this defect by linking cross-section demographic data with the available time series data

While the extent of specification bias caused by the exclusion of relevant variables cannot be precisely determined, the nature of the N.Y. market suggests that it is probably small. Over the last 10 years, population shifts have resulted in a gradual change in the spatial distribution of our circulation sales with suburban sales increasing relative to city sales. Since these changes have not been abrupt proxy variables, which mirror these changes, have been constructed and used in the circulation equations.

Another long-term problem, well documented and related indirectly to demographics, is the nation-wide decline in newspaper readership. The N.Y times is not unaffected by this trend. A number of theories purport to explain this phenomenon. It is however, not a cyclical problem. Consequently, time trends were used to capture this process of erosion.

The quarterly fluctuations in circulation, apart from seasonal and random variation, seem to be explaind by price dynamics and economic changes. Circulation in each category is explained by economic indicators, advertising lines, prices and circulation from other categories. There is a good deal of substitution and complementarity between the various categories of circulation. For example, people who buy the daily paper tend to also purchase the Sunday paper. A weaker reverse relationship also holds. There is also some substitution between home delivery sales and newsstand sales. The equations were designed to take these relationships into account.

Advertising in each of the equations is related to various economic indicators, circulation, and the substitute-complement relationships which hold for advertising.

The parameters of the demand model were estimated according to the following methodology:

1. Preliminary analysis indicated those variables which explained quarterly fluctuations in volume. Data problems and multicollinearity eliminated some of the variables which belong in the equations.
2. Next, the appropriate functional form for each of the equations was determined. The usual summary statistics were used in exploring various specifications. In latter versions, multiple time series techniques were used to determine dynamic lag structures.
3. Each equation was then estimated by ordinary least squares. In many cases, serial correlation was present and the standard procedures were used to correct it. Unfortunately, in dynamic simulations this correction magnifies specification biases producing explosive out-of-sample forecasts. The usual approach, variable adjustments to the constant term with these adjustments diminishing over time was applied. I am currently trying another technique which allows for more complex stochastic specification then the typical first-order auto-correlation adjustment. It involves applying time series techniques to the residuals from each equation. The forecast for each variable then consists of a deterministic part and a stochastic part. The use of time series techniques to determine stochastic <u>and</u> dynamic specification is a generalization of the transfer-model approach outlined by Box and Jenkins and studied by Zellner and Palm in the context of simultaneous econometric equation models.
4. Finally, simultaneous estimation tecniques were applied and the estimated model was subjected to a number of validation tests.

IV THE INTEGRATION OF THE MODEL INTO THE PLANNING AND BUDGETING CYCLE: SOME APPLICATIONS

The model outlined in the preceeding pages is fully integrated within the budgeting and planning cycle. Several examples will perhaps illustrate this.

A. THE ANNUAL BUDGET

In the fourth quarter, the newspaper prepares its annual budget for the coming year. The most important components of the budget are the volume forecasts of advertising and circulation. The advertising and circulation sales departments spend several weeks preparing preliminary estimates Independently, the model is used to produce several forecasts corresponding to a most likely scenario, a more optimistic scenario, and a pessimistic scenario.

Management then reviews each set of forecasts assesing their strenghts, weaknesses, and reliablility. An analysis of the differences between the model's forecasts and the sales forecasts begins the process of producing a final set of volume forecasts in which the major differences have been reconciled. These forecasts are then distributed to the managers of the operating departments and the budgeting process continues. Between the begining and end of the budget cycle, a period of about four months, the demand model is periodically run to determine whether or not there are any departures from the original forecasts.

The procedure just outlined was followed in bothe the 1975 and 1976 budget cycle. model's 1975 forecasts, produced in December 1974, were quite accurate. The forecast error for total linage was under one percent and the circulation error slightly over one percent. Individual equation errors were larger but cancelled out when aggregated.

At the end of the cycle, the full model is run and the budget is reproduced. A set of sensitivity experiments is then performed to determine the impact of of volume changes and policy options in response to those changes. A set of income statements summarizes these simulations. As a result of these exercises, contingency planning is greatly facilitated.

B. THE FIVE YEAR PLAN

The planning and budget cycles at the Times overlap: the annual budget is the first year of the five year plan. The planning cycle begins in March and ends in the early part of the fourth quarter when the budget cycle begins.

In March, a complete run of the model is produced. The volume forecasts and income statement summary are compared to the previous year's plan and the current operating goals set by management. The results of this analysis are distributed to the operating managers for comment and analysis. The advertising and circulation departments, using the model's volume forecasts as a guide, then prepare their own forecasts. This process eventually results in a set of consensus volume projections.

With this new set of volume numbers, the production-cost model is re-run and a preliminary five year income statement is produced. If at this time the financial results still fall short of the plan's objectives a set of action plans, strategies designed to improve performance, are developed and simulation experiments are performed, to determine their impact on profits. This process continues until the five year plan is finalized.

Note the pivotal role that the model has in the planning process. Indeed, in the absence of a model the planning process, in the true sense of the word, does not really exist. The manual procedures followed by the various departments prior to the model's development limited the process to essentially one pass. The model has given management the capability to explore alternative plans quickly and react promply to changes in external conditions which result in a departure from the established plan.

C. NEW PRODUCTS

The cost and revenue implications of current and planned product changes have been analyzed using the simulation model. For example, responding to shifting population, special regional editions are being produced. In the 1976 plan the model was used to determine profit and loss statements for each of the new products. Following a base run, each new product change was introduced separately. In this manner the incremental costs and revenues for each product were generated. From this information a break-even analysis could be prepared and an appropriate pricing strategy developed. This analysis provided the information management needed to determine the economic feasibility of each product, information which would not have been available had manual procedures been followed.

V CONCLUSIONS

The preceding examples document the integration of newspaper model into the planning process at the New York Times. The model is used to set goals, design strategies to achieve these objectives, forecast, and monitor changes which might prevent those goals from being reached. Management uses the model and believes its output.

The model, however, is far from finished. Some of the changes, which will eventually improve the model, were suggested in the preceding pages. New information and changing technology require constant modification of the model's structure. It is a living organism, constantly growing and changing. In this way we hope to insure its continued usefulness to management.

BIBLIOGRAPHY

1. Box,G . and Jenkins,G.,Time Series Analysis: Forecasting and Control,Holden-Day,1970

2. Zellner,A. and Palm, F., "Time Series Analysis and Simultaneous Equation Models", Journal Of Econometrics.Vol.2, May 1974, p 17-54

WILLIAM A. SHERDEN is a member of Dynamics Associates, Inc., a firm that provides corporate planning consulting and computer software for a wide range of industrial, financial, and public sector clients. While at Dynamics, he has assisted a number of major corporations in the design and implementation of financial models for long-range planning, budgeting, and investment analysis. Mr. Sherden holds a Masters degree from the Graduate School of Business at Stanford University and a Bachelors degree in Industrial Engineering from Georgia Institute of Technology.

ORIGIN OF SIMULTANEITY IN CORPORATE MODELS

William A. Sherden
Dynamics Associates, Inc.
Cambridge, Massachusetts 02138

Corporate planning models are firmly established in the formal planning process of many firms. Of particular popularity is the deterministic corporate model, which represents the entire firm or some subdivision of the firm. This paper defines the nature of simultaneous relationships that are found in these corporate models. Three classes of simultaneity are identified that do not correspond to the actual causal behavior of the firm. Two of these classes are useful simulation devices, while the third is a class of problems common to corporate models.

INTRODUCTION

Recent advances in simulation software and modeling expertise have contributed to the success of the corporate model as a planning tool [4]. Unlike the earlier mission of the corporate model [3] to represent every facet of the firm in great detail, the concept of a corporate model described here is a comprehensive, yet concise, mathematical representation of the firm or some subdivision of the firm. As demonstrated by Warren and Sheldon [7], a simultaneous equation approach is essential if a corporate model is to be both concise and capable of expressing the interdependent characteristics of the firm.

This paper defines the nature of simultaneity found in corporate models as it relates to causal behavior. In particular, this paper identifies three classes of simultaneity that do not correspond to actual causal relationships: simultaneous decision rules, reverse simultaneity, and spurious simultaneity. Simultaneous decision rules and reverse simultaneity are useful computational devices, while spurious simultaneity is a misrepresentation of relationships in a corporate model.

CAUSALITY, SIMULTANEITY, AND CORPORATE MODELS

It is widely acknowledged that in theory dynamic models of social and economic systems are purely recursive [1][8]. That is, actions are caused by previous actions because time delays in the natural stimulus-response processes eliminate the possibility of simultaneous causality. Adhering to this principle, some schools of modeling maintain that derived mathematical models should also consist of only recursive relationships. A notable field in this purist school is Industrial Dynamics [2], which proposes that the time period of a model should be sufficiently small as to dissolve any interdependencies between variables within the same time interval, thereby eliminating any simultaneous relationships. Attempting to preserve these recursive relationships in a mathematical model poses resource problems, and an approach with larger time periods and concise relationships becomes a desirable alternative. By choosing such an alternative, one finds that simultaneous equation models are useful in expressing interdependent feedbacks that are captured within concise relationships and larger time periods.

The focal points or "bottom-lines" of social and economic models depend upon their subsystems. Individually, each of these subsystems is sufficiently complex to encumber a model with immense detail. Perpetually faced with this problem, the model designer must balance the tradeoffs of detail versus realism by summarizing the behavior of the subsystem. For instance, dividends and capital structure are customarily two focal points in corporate models. If a corporate model projects dividends per share and new equity issues, then it must also express the valuation of these new equity issues. The process of equity valuation implies the inclusion of equity market behavior, which is a subsystem of tremendous complexity.

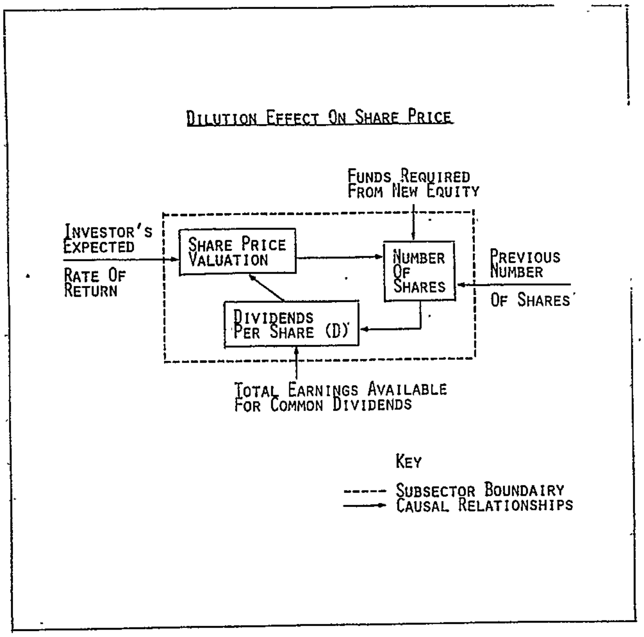
A simplification of this subsystem is shown in Illustration 1. The characteristics of this subsector are based on a classical valuation model [5] that reflects the discounted return on investment to the owner of the equity share. Upon issuing new equity shares, the following chain of causes and effects between the number of shares (S), the "per-share" return on investment (R), and the price per share (P) is anticipated.

$S\uparrow\rightarrow R\downarrow\rightarrow P\downarrow\rightarrow S\uparrow$[1] ...

---

[1] ↑ indicates an increase.
↓ indicates a decrease.
→ indicates dependency.

ILLUSTRATION 1



Apart from these simplified assumptions, building a detailed recursive model of an equity market poses difficult questions. What are the time delays and durations in this sequence of causes and effects? Is the effect of the dilution anticipated far in advance of the formal sale of new shares, or is the market less efficient? In building comprehensive corporate models that must express many complex subsystems, these questions become distracting. Alternatively, the intricate behavior of the equity market can be summarized by a simple simultaneous system as shown in Equations 1, 2, and 3.

$$P_t = \frac{D_t}{(r-g)} \tag{1}$$

$$S_t = \frac{F_t}{P_t} + S_{t-1} \tag{2}$$

$$D_t = \frac{E_t - R_t}{S_t} \tag{3}$$

Endogenous Variables

P_t = Equity share price
S_t = Number of equity shares
D_t = Dividends per share

Exogenous Variables

r = Investor's expected rate of return
g = Dividend growth rate
F_t = Funds required from new equity
E_t = Total earnings available for common dividends
R_t = Desired retention of earnings

The purpose of this example is not to propose a particular price valuation theory or dividend policy, but to demonstrate a simple scheme for expressing the dilution cost accruing to the sale of new equity issues. Building a more detailed recursive model is an unreasonable alternative. The increase in development expense of such an approach is a certainty, while the payoff in increased predictability is doubtful.

In many cases the availability of reported data forces the period size of derived models to exceed that which is necessary to accurately reflect the recursive qualities of a system. Corporate models, however, are an exception. The restrictions on period size of corporate models arise from the long-range perspective of the planning function of a firm. As a planning tool, corporate models generally project a subset of variables over a three- to five-year range of large time periods, requiring the approximation of intraperiod feedbacks as a simultaneous system. The objective of corporate models is not to foretell the future in minute detail, but rather to express the long-range implications of strategic plans over a range of hypothetical economic and competitive environments. In adhering to these objectives the larger time intervals of corporate models prohibits a purely recursive representation of a firm. The example shown in Illustration 1 demonstrates the classical use of simultaneous equations to approximate recursive feedback relationships trapped within the larger time periods of a corporate model.

This simultaneous approximation is an attempt to concisely express actual cause and effect relationships. The following sections are a departure from this attempt. These sections present three classes of simultaneous relationships that do not correspond to causal relationships.

SIMULTANEOUS DECISION RULES

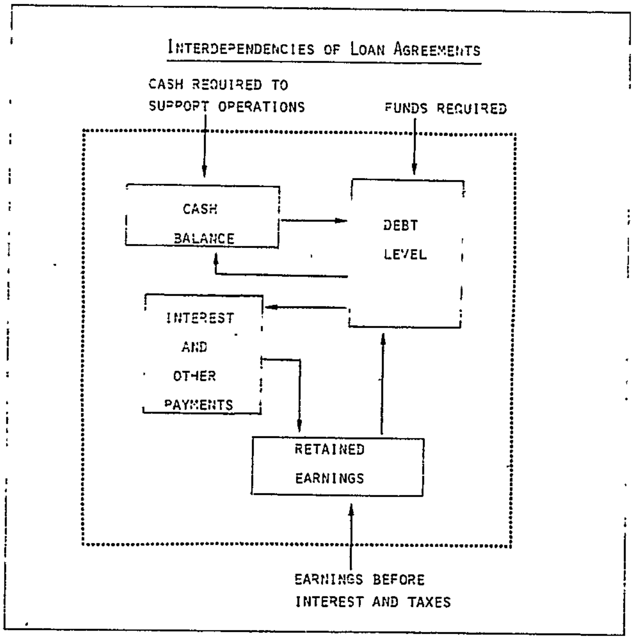
The behavior of a corporation is highly constrained by explicit and implied commitments. As an integral part of the environment of a firm, these constraints include agreements with external agents, industry conventions, stockholder expectations, loan covenants, and so forth. Realistically, an accurate forecast of the financial position of a firm cannot be prepared independently of these restrictions. Within the limitations of a corporate model, simultaneous decision rules are devices that ensure the compatibility of the simulated performance of a firm with these environmental constraints.

To realistically specify a decision process, decision rules and their effects must be jointly considered. Often these decision rules pose problems in identifying the ultimate response dictated by the decision rules of the model. The impact of each decision simultaneously alters the state of the projected firm upon which the decision was based, which consequently may evoke a different decision and so on. As a device, simultaneous decision rules provide a joint solution for the state of the firm, the actions prescribed by the decision rules, and the effects of these actions.

A common example of the use of simultaneous decision rules in corporate models is the problem of accounting for the impact of debt financing. As part of a corporate model, the decision rule that chooses the level of debt financing affects other corporate variables, which in turn determines the state of the corporation upon which the debt-financing decision is based. These feedback relationships are caused by the payment of interest and other fees that the lender requires as inducement to make the loan. Commercial loan covenants specify a package of direct and indirect charges consisting of a combination of interest charges, commitment fees, compensating balances, and inflationary hedges that together create an additional expenditure for which the firm must find more funds.

Illustration 2 demonstrates the mutual dependency between the decision rule specifying the use of debt financing and the resulting effects of a loan agreement.

ILLUSTRATION 2



The projected funds required in this example are presumed to be financed entirely with debt. This debt level is determined by the funds required (independent of the debt balance) and two components that are directly related to debt: funds invested in compensating balances and deductions from earnings.

Given a specific amount of required funds, the debt funding decision combines with the effects of the loan agreement to generate an endless chain of computations. For example, to determine the level of required financing, a firm includes in its calculations money that will be drained by "front-end" fees and cash committed to required balances, both of which are a function of the eventual debt balance. In addition, the level of financing required at the end of the period must be sufficient to cover interest expense accruing during the period. Compution of a solution is vastly simplified by using simultaneous equations to represent the interdependencies of debt financing as shown below.

$$(D_t - D_{t-1}) = XF_t + (C_t - C_{t-1}) - (RE_t - RE_{t-1}) \quad (4)$$

$$C_t = MAX(\mathcal{c}D_t, XC_t) \quad (5)$$

$$I_t = \mathcal{i}D_t + \mathcal{f}MAX(D_t - D_{t-1}, 0) \quad (6)$$

$$(RE_t - RE_{t-1}) = (EBIT_t - I_t)(1-\mathcal{t}) \quad (7)$$

Endogenous Variables

C_t = Cash balance at time t
D_t = Debt balance
I_t = Payments as per loan agreement
RE_t = Retained earnings balance

Exogenous Variables

XF_t = Funds required
XC_t = Cash balance required by operations
$EBIT_t$ = Earnings before interest and taxes
$\mathcal{c}$ = Compensating balance commitment
$\mathcal{i}$ = Interest expense
$\mathcal{f}$ = Commitment fee rate
$\mathcal{t}$ = Effective tax rate

Equation 4 determines the change in debt balance as a function of the funds required (XF_t), the change in cash balance ($C_t - C_{t-1}$), and the change in retained earnings ($RE_t - RE_{t-1}$). As specified in Equation 5, the corporate cash balance is equal to the greater of the cash requirements for operations or for compensating balance agreements. The change in retained earnings specified in Equation 7 is directly affected by the deductions from earnings summarized in Equation 6.

The simultaneous equation system 4 through 7 is a device that provides a joint solution for the four endogenous variables that satisfies both the debt financing rule and the loan agreements. It is unrealistic, however, for a firm to satisfy all of its funds requirements with long-term debt because shareholders insist on a more balanced capital structure. Therefore, a corporate model must be capable of specifying combinations of several financial sources. With the addition of new financing alternatives, the decision rules for anticipating the behavior of management become complicated. They are confounded by combinations of states, actions, and the presence of interdependencies between these states and actions. A more complete version of the previous example is presented below to illustrate the use of simultaneous decision rules to simplify this added complexity.

The following example includes a subset of decision rules found in the financial sectors of corporate models. The model first determines either a surplus or a deficit of funds and then employs the following hierarchy of decision rules.

Surplus Funds State

1. Retire short-term debt
2. Invest residual surplus funds in short-term securities

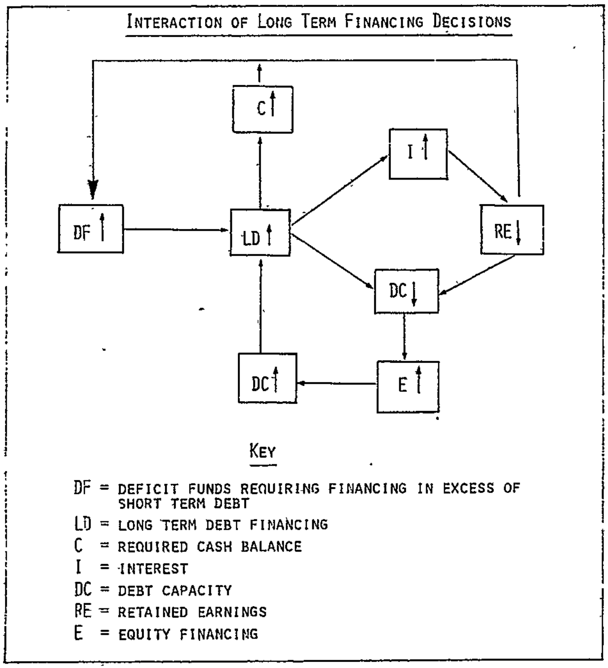
Deficit Funds State

1. Liquidate short-term securities
2. Increase short-term debt to maximum level
3. Increase long-term debt to maximum debt capacity
4. Fund residual deficit with common equity

Additionally, all assets and liabilities must be nonnegative; and long-term debt and equity levels are presumed to be nondecreasing.

Individually, each of these decision rules is simple. However, as part of an interdependent system, each decision leads to an endless sequence of effects, altered states, and, ultimately, a modification of the decision. For example, Illustration 3 demonstrates this interaction in funding a deficit in excess of allowable short-term debt. The deficit is also presumed to exceed the debt capacity of the firm, and therefore, the decision rules specify a mix of debt and equity financing.

ILLUSTRATION 3



This network of effects is an example of the many possible interactions dictated by the decision rule hierarchy. While representing the entire financing sector in a block diagram would be confusing, the total financing sector is completely represented by the simultaneous decision rules 8 through 15.

$$F_t = XF_t - (RE_t - RE_{t-1}) + (MC_t - C_{t-1}) \tag{8}$$

$$DF_t = MAX(F_t, 0) \tag{8A}$$

$$MC_t = MAX(XC_t, b(SD_t)) \tag{9}$$

$$C_t = MAX(MC_t - F_t, MC_t) \tag{10}$$

$$SD_t = MIN(DF_t, u) \tag{11}$$

$$LD_t - LD_{t-1} = MIN(DF_t - SD_t, k(E_t + RE_t + SD_t + LD_t) - LD_{t-1}) \tag{12}$$

$$E_t - E_{t-1} = MAX(DF_t - SD_t - (LD_t - LD_{t-1}), 0) \tag{13}$$

$$I_t = i_1(SD_t + LD_t) - i_2(C_t - MC_t) \tag{14}$$

$$RE_t - RE_{t-1} = (EBIT_t - I_t)(1-t) \tag{15}$$

Endogenous Variables

F_t = Effective surplus or deficit funds
DF_t = Deficit funds
MC_t = Required cash balance
C_t = Cash and short-term securities
SD_t = Short-term debt
LD_t = Long-term debt
E_t = Equity
I_t = Total income deductions
RE_t = Retained earnings

Exogenous Variables

XF_t = Funds required
XC_t = Cash required to support operations
$EBIT_t$ = Income before interest and taxes
b = Average compensating balance requirement
i_1 = Average interest rate on borrowed funds
i_2 = Interest rate on invested funds
t = Effective tax rate
u = Maximum allowable short-term debt
k = Debt to total capital constraint

Equation 8 defines the magnitude of the surplus or the deficit of funds as the difference between projecte assets and liability levels for time t. (Auxiliary Equation 8A simplifies the definition of deficit funds in subsequent relationships.) Since Equation 8 computes the funds required net of short-term debt (short-term debt is assumed to mature at year-end), any surplus funds are added to the level of cash and short-term securities. Fund deficits are financed first by obtaining short-term debt in Equation 11 and then by a combination of long-term debt and equity. Equations 12 and 13 determine the mix of long-term debt and equity by the debt to total capital constraint that is controlled by the parameter k. As expressed by Equations 9, 14, and 15, the effects of these decision rules modify the state of the fund (Equation 8) and thus complete the interdependent cycle. The alternative to these simultaneous decision rules is guessing combinations of decisions that would satisfy the financial requirements of the firm. This would be a difficult task that would involve a lengthy and undesirable sequence of trials and errors.

The previous example demonstrates the use of simultaneous decision rules as a device to ensure the compatibility of the simulated performance of the firm with environmental constraints. This class of simultaneous relationships serves to constrain simulation results and does not correspond to actual cause and effect relationships. For instance, the interaction between long-term debt, debt capacity, and equity as shown in Illustration 3 is merely a device that maintains the simulated capital structure of the firm within a realistic tolerance. This notion of causal representation is completely absent from the classes of simultaneity presented below.

REVERSE SIMULTANEITY

Typically, the bulk of a corporate model is a straight sequence of nonsimultaneous relationships. Largely due to accounting convenience, this "top-down" flow of relationships starts with key input assumptions. For example, a sales forecast is a key input assumption that starts a sequence of relationships which cascades downward throughout the relationships in an income statement. In addition to simulating the results of these key inputs, corporate models are also used to determine the key inputs that are necessary to achieve given target results. For instance, a corporate model could be used to determine the level of sales that is necessary to achieve an earnings-per-share target. These "backward" simulations are generated by reoriented corporate models that contain reverse simultaneous relationships, that is, relationships that solve for the cause which will produce a given effect.

In order to determine the key inputs necessary to achieve a target result, a corporate model must be reoriented in a "bottom-up" fashion. This process of reversing a model is accomplished by reversing the original identity relationships. These reversed identities now define the variable that was originally the key input in terms of a variable which was originally dependent upon the key input. For example, the following "top-down" sequence of equations defines gross profit (G_t) in terms of revenue (R_t) and cost of goods sold (C_t). Cost of goods sold is assumed to vary as some proportion k of revenue as shown in Equation 16, which combined with revenue determines gross profit as shown in identity 17.

$$C_t = kR_t \tag{16}$$

$$G_t = R_t - C_t \tag{17}$$

This straight sequence of relationships generates the results (gross profits) resulting from the key input (revenue). Alternatively, a reversed form of Equations 16 and 17 calculates the revenue necessary to achieve a specified target gross profit. This reversed form is created by reversing identity 17 as shown in Equation 18.

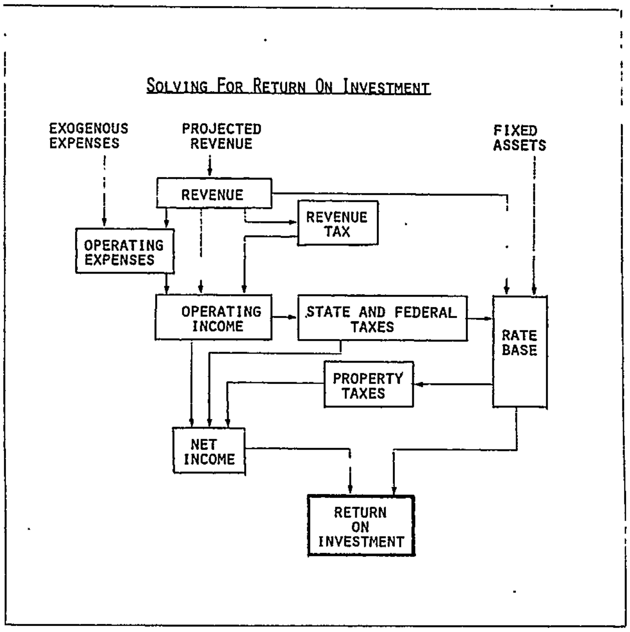
$$R_t = G_t + C_t \tag{18}$$

Equations 16 and 18 now solve for revenue in terms of a gross profit target. Note that Equations 16 and 18 are simultaneously related; revenue both determines cost of goods sold (Equation 16) and is defined by cost of goods sold (Equation 18). The simultaneous relationship is produced by the reversed identities that redefine a variable in terms of components which are dependent upon the variable. This simultaneity, however, does not describe a causal relationship, but rather a reverse-causal relationship; contrary to the actual cause and effect relationship, revenue is now dependent upon cost of goods sold and gross profit. This reverse simultaneity provides a means of solving for key inputs necessary to achieve a target result, while preserving the original assumptions of the model.

A more complete example of reverse simultaneity is shown below. The example illustrates the use of a reversed corporate model of a utility that solves for the level of revenue (a key input) which is implied by a return on investment fixed by a regulatory commission (a given target result). The original model consists of a "top-down" sequence of accounting identities, regulatory rules, and behavioral relationships. In solving for revenue implied by return on investment, however, the objective of the model flows "bottom-up". Reverse simultaneity results from this flow of objects contrary to the flow of relationships in the original model.

The downward flowing dependencies in the original model are demonstrated in Illustration 4 and in Equations 19 through 26.

ILLUSTRATION 4



Corporate Models (continued)

The sequence of dependent relationships begins with projected gross revenue for a given year (referred to as the "test year")[2], which determines revenue taxes in Equation 19 and operating expenses in Equation 20[3] and which defines operating income in identity 21. Net income is defined in Equation 24 as the excess of operating income over income and property taxes. Referred to as the rate base, the value of the net investment of the firm as expressed in Equation 25 is dependent upon Federal taxes[4] and gross revenue.[5] Finally, net income and the value of the rate base combine to yield return on investment in Equation 26.

$$RT = t_1 R \tag{19}$$

$$OE = XOE + eR \tag{20}$$

$$OI = R - OE - RT \tag{21}$$

$$PT = t_2 XRB \tag{22}$$

$$FT = t_3 OI \tag{23}$$

$$NI = OI - PT - FT \tag{24}$$

$$RB = XRB - r_1 FT + r_2 R \tag{25}$$

$$ROI = \frac{NI}{RB} \tag{26}$$

Endogenous Variables

RT = Revenue tax

OE = Operating expenses

OI = Operating income

PT = Property tax

FT = Federal and State tax

NI = Net income

RB = Rate base

ROI = Return on rate base

Exogenous Variables

R = Gross revenue

XOE = Operating expenses unrelated to revenue

e = Relationship of operating expenses to revenue

t_1 = Revenue tax rate

t_2 = Property tax rate

t_3 = Effective Federal and State tax rate

XRB = Exogenous rate base components (e.g., fixed plant)

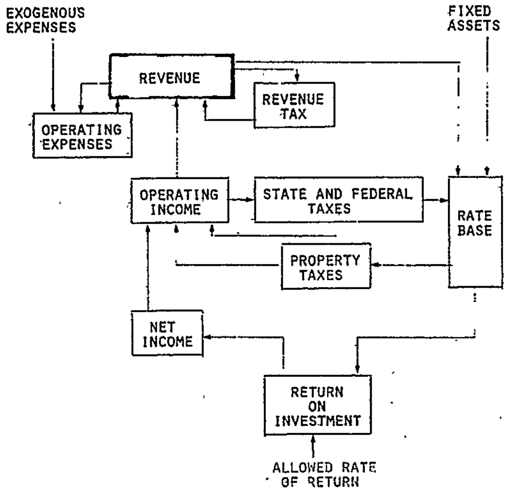
r_1 = Adjustment for accrued taxes

r_2 = Relationship of working capital to revenue

The corporate model above is incomplete. Though the flow of dependencies is clearly a "top-down" sequence, the objective of the analysis is to solve for revenue as opposed to determining a return on investment resulting from a revenue estimate. Illustration 5 is a block diagram of the reversed version of the original corporate model, which now solves for revenue.

ILLUSTRATION 5

SOLVING FOR REVENUE



The presence of reverse simultaneity is clear in Illustration 5. Observe the circular flow of relationships resulting from the "top-down" flow of relationships from the original model (Illustration 4) and the "bottom-up" flow of the objective in the reversed model.

The following equations represent the reversed version of the original Equations 19 through 26.

$$RT = t_1 R \tag{19}$$

$$OE = XOE + eR \tag{20}$$

---

[2] For convenience, the time subscripts have been omitted.

[3] Operating expenses such as collections are dependent upon gross revenue.

[4] The rate base is adjusted for the Federal loan implied by tax accruals.

[5] Working capital such as receivables are dependent upon gross revenue.

$$R = OI + RT + OE \tag{27}$$

$$PT = t_2 XRB \tag{22}$$

$$FT = t_3 OI \tag{23}$$

$$OI = NI + PT + FT \tag{28}$$

$$RB = XRB - r_1 FT + r_2 R \tag{25}$$

$$NI = RB * ROI \tag{29}$$

New Endogenous Variable

R

New Exogenous Variable

ROI

Note that the only changes to the original system of equations are the reversed identities 27, 28, and 29; the equations 19, 20, 22, 23, and 25 remain unchanged.

The reversed identities 27, 28, and 29 specify simultaneous relationships by defining a variable in terms of components that are dependent upon the variable. For example, Equation 27 defines revenue (R) in terms of revenue tax (RT) and operating expense (OE), both of which are dependent upon revenue itself. In addition, these simultaneous relationships are reverse-causal. For instance, Equation 27 does not correspond to the actual causal relationships of the utility, as revenue is not determined by expense items. These simultaneous relationships result from relationships that are exactly opposite to the actual cause and effect dependency of the utility.

Despite this lack of correspondence to the actual relationships of a firm, reverse simultaneity is useful for solving for the key inputs necessary to achieve target results.

SPURIOUS SIMULTANEITY

Corporate models often contain spurious simultaneous relationships that do not correspond to causal behavior; they, in fact, result from a distortion of the true cause and effect relationships of a system. These false interdependencies occur when a flow, determined as a function of a level, also determines this level. For instance, Equation 30 determines a flow (F_t) during period t as a function of a level (L_t). The level (L_t) is then determined by the flow (F_t) in Equation 31.

$$F_t = f(L_t, \ldots) \tag{30}$$

$$L_t = f(F_t, \ldots) \tag{31}$$

The intended process underlying spurious simultaneity is one of assignment; the flow during period t is computed and then assigned in the equation describing the level at the end of period t. To imply that a flow during an interval of time is dependent upon the level it determines at the end of the time period is a misrepresentation of time in a model, which causes spurious simultaneity.

Spurious simultaneity is eliminated by removing the dependency of the flow on the level it determines. This dependency is removed either by changing the time notation of the level to represent the beginning of the period (as shown in Equation 32) or by excluding the level altogether from the relationship describing the flow.

$$F_t = f(L_{t-1}, \ldots) \tag{32}$$

The interpretation of the flow determines the correct alternative. If the level is determined by the flow and the previous value of the level, then the flow should be dependent upon the previous level. Alternatively, if the level is dependent upon the flow alone, and not the flow plus the previous value of the level, then the level should be excluded from the relationship describing the flow.

An example of spurious simultaneity is shown below in Equations 33 and 34. Invested cash (C_t) is presumed to grow with interest earned during period t (I_t), assuming an annually compounded interest rate k.

$$I_t = kC_t \tag{33}$$

$$C_t = C_{t-1} + I_t \tag{34}$$

Equations 33 and 34 are simultaneously related. However, this simultaneous relationship is not equivalent to the "intended" growth in the cash balance. Equation 35 is the reduced form of Equations 33 and 34. When compared to the "intended" relationship shown in Equation 36, the spurious simultaneity problem in Equations 33 and 34 becomes clear as Equations 35 and 36 are not equivalent.

$$C_t = \frac{1}{(1-k)} C_{t-1} \tag{35}$$

$$C_t = (1 + k)C_{t-1} \tag{36}$$

The problem is found in Equation 33 where the flow of interest earnings during period t is dependent upon the level of cash at the end of period t. The period-end level of cash is determined by adding interest earnings during period t to the cash level at the beginning of period t, which modifies the flow of interest earnings in Equation 33 and so on. Since the cash level is determined by the flow of interest earnings and the previous cash level (Equation 33), the spurious simultaneity is eliminated by changing the time notation for the level of cash in Equation 33 to represent the level at the beginning of the time period t. Equation 37 correctly specifies the flow of interest earnings during period t. Note that the reduced form of Equations 37 and 34 restores the intended expression for the level of cash shown in Equation 36.

$$I_t = kC_{t-1} \tag{37}$$

The complexity of corporate decision rules is the primary cause of spurious simultaneity. When faced with this complexity, the model designer often specifies each relationship independently. When the perspective of the model as a total system is lost, spurious interdependencies are overlooked. Since the financing sector of a corporate model consists of the most complex relationships, it is the most probable candidate for spurious simultaneity. In accordance with this popularity, the spurious simultaneity problem in the financing sector is given special treatment below.

The spurious simultaneity problem in the financing sector originates with the specification of funds required. When the funds requirement for a period (a flow) is dependent on a level that is assigned all or a portion of the funds required, the model will contain spurious simultaneity. This problem is illustrated in Equations 38, 39, and 40, where the funds required during period t is computed in Equation 38 as a function of period-end asset and liability levels. This funds requirement is assumed, then, to be financed with debt as shown in Equation 40. Note, however, that the flow of funds required is dependent upon the period-end level of debt because of the definition of the liability level in Equation 39.

$$F_t = A_t - L_t \tag{38}$$

$$L_t = XL_t + D_t \tag{39}$$

$$D_t = F_t + D_{t-1} \tag{40}$$

Endogenous Variables

F_t = Required funds
L_t = Projected liabilities
D_t = Projected debt balance

Exogenous Variables

A_t = Asset levels
XL_t = Other liability levels

The equation system 38, 39, and 40 contains a spurious simultaneous relationship between the flow of funds required (F_t) and the period-end level of debt (D_t). Solving for the period-end level of debt in the reduced form of this equation system is shown in Equation 41.

$$D_t = \frac{A_t - XL_t + D_{t-1}}{2} \tag{41}$$

When contrasted with the intended reduced form shown in Equation 42, the spurious simultaneous relationship between Equations 38, 39, and 40 becomes clear.[5]

$$D_t = A_t - XL_t \tag{42}$$

The intended specification of this financing sector is restored by replacing the period-end debt level in Equation 39 with the beginning period debt level as shown in Equation 43.

$$L_t = XL_t + D_{t-1} \tag{43}$$

Avoiding spurious simultaneity is difficult in financing sectors more complex than the previous example. Recall the expression for required funds (F_t) in Equation 8.

$$F_t = XF_t - (RE_t - RE_{t-1}) + (MC_t - C_{t-1}) \tag{8}$$

Equation 8 is properly specified because the flow F_t is not dependent upon levels that are determined by F_t.[6]

A misspecified version of Equation 8 is shown in Equation 44.

$$F_t = XF_t - (RE_t - RE_{t-1}) + (C_t - C_{t-1}) - (E_t - E_{t-1}) - (LD_t - LD_{t-1}) - (SD_t - SD_{t-1}) \tag{44}$$

Observe that in Equation 44 the flow of funds required (F_t) is dependent upon the level of cash (C_t), equity (E_t), long-term debt (LD_t), and short-term debt (SD_t). Since each of these levels is dependent upon F_t, a spurious simultaneous relationship exists

---

[5] The factor "2" in the denominator of Equation 41 is found in the reduced form of a spurious simultaneous equation system when the forms of the flow and level equations are:

$$F_t = f(cL_t, \ldots)$$

$$L_t = f(-cF_t, \ldots)$$

where c equals either 1 or -1.

[6] There is a subtle difference between the spurious simultaneity described above and the dependency of the flow of interest expense on the level of debt (Equation 13). In the latter case, the debt balance is assumed to be changing during the period as a function of the changing requirement for financing. The flow of interest expense is dependent on an average level of debt during the period.

between Equation 44 and each of the Equations 10, 11, 12, and 13.

Often spurious simultaneity deceptively appears to be a logical representation of causality. Through a misrepresentation of time, spurious simultaneity specifies the dependency of a flow on a level that is determined by the flow. What is intended to be an assignment of the flow to the level becomes a simultaneous relationship that distorts the actual causal behavior of the system.

CONCLUSION

Models of social and economic systems employ simultaneous equations to express feedback relationships captured within the concise relationships and large time intervals. Corporate models also contain three classes of simultaneous relationships that do not correspond to the actual cause and effect relationships of the firm. This paper identifies these three classes as simultaneous decision rules, reverse simultaneity, and spurious simultaneity. In the first two classes, simultaneous equations are employed as a balancing device. Simultaneous decision rules provide a balance between the simulated performance of a firm and an anticipation of environmental constraints. Reverse simultaneity balances the process of solving for the key inputs necessary to achieve target results, while preserving the original assumptions of the corporate model. The third class of noncausal simultaneity, spurious simultaneity, results from a distortion of time in a level and flow relationship.

Simultaneous decision rules and reverse simultaneity are simple devices for generating answers to questions that are otherwise complex. Conversely, spurious simultaneity is an elusive pitfall that destroys the intended relationships in corporate models.

BIBLIOGRAPHY

[1] Fisher, F.M. A Correspondence Principle for Simultaneous Equation Models, Econometrica: 38, January, 1970.

[2] Forrester, J.W. Industrial Dynamics. Cambridge: MIT Press, 1965.

[3] Gershevski, G.W. Building a Corporate Financial Model, Harvard Business Review 47, July, 1969.

[4] Naylor, Thomas H. The Future of Corporate Planning Models, Managerial Planning, March/April, 1976.

[5] Robichek, A.A. and Meyers, S.C. Optimal Financing Decisions. Engelwood Cliffs: Prentice Hall, Inc., 1965.

[6] Suelflow, J.E. Public Utility Accounting: Theory and Applications, Michigan State University, Public Utilities Studies, 1973.

[7] Warren, J.M. and Shelton, J.P. Simultaneous Equation Approach to Financial Planning, Journal of Finance 26, December, 1971.

(Reply with rejoinder: Carter, E.E. 28, September, 1973.)

[8] Wold, H. Econometric Model Building. Amsterdam: North Holland Publishing Co., 1964.

76
Bicentennial

WINTER
SIMULATION
CONFERENCE

DATA BASE MANAGEMENT SYSTEM SIMULATION METHODOLOGIES

Chairman:

Thomas G. DeLutis
Ohio State University

GORDON D. EDGECOMB

Gordon D. Edgecomb is a Senior System Engineer in the Systems Engineering Department of Chemical Abstracts Service. Mr. Edgecomb joined CAS in 1967 as a computer specialist and assumed the position of Manager of the Data Processing Operations Department in June 1970. He assumed his present position in November 1971. He has been primarily concerned with the development of techniques for computer simulation at CAS.

Mr. Edgecomb currently is pursuing a bachelor's degree in business administration at Franklin University.

RICHARD J. THOMPSON

Richard J. Thompson is Manager of the System Engineering Department of Chemical Abstracts Service. His experience includes engineering, analysis, and design of computer systems and management of associated technical functions. In his present position he is responsible for the management of the Data Base Administration and computer performance evaluation functions within the R&D Division.

Mr. Thompson has received the A.E.E. in Electronic Engineering, the B.B.A. in Engineering Management from Northeastern University, and the M.B.A. from Xavier University. He is a member of IEEE's Professional Group on Instrumentation and the American Society for Information Science.

VALIDATION -- THE BOTTLENECK IN SYSTEM SIMULATION

Gordon D. Edgecomb
Richard J. Thompson
Chemical Abstracts Service

ABSTRACT

In simulation exercises some degree of validation is necessary in order to create confidence in the results obtained. Experienced users of generalized simulation software and commercially available simulation tools often find that the majority of their simulation efforts are localized in the validation cycle. Discussed in this paper are some of the factors which have caused the validation process to become more complicated and more time consuming. This paper also summarizes the major problems with simulating complex computer systems and suggests some basic requirements and design philosophies for new simulation packages.

INTRODUCTION

The major use of simulation at Chemical Abstracts Service (CAS) is as an aid to computer configuration planning. During the expansion of the computer system over the past five years, CAS's simulation goals have been to predict the impact of changes in workload, hardware, and control system software. The spectrum of project analyses has varied widely in magnitude and complexity. It has included analyzing major model changes in the computer system, forecasting the operational impact of new applications software, identifying system components which have low utilization or negligible growth factors, and answering some very practical computer loading questions.

During the time CAS has been active in simulation work, significant changes have occurred in the company's environment, in the capabilities of the simulation tools which are available commercially, and in the methods employed in the simulation process. Some of these changes have made the simulation analysts' job easier, some have made the job much more difficult.

BACKGROUND

Early simulations at CAS dealt with relatively simple environments. For example, in our first use of computer simulation we evaluated the operation of a fixed batch load on several candidate computer configurations. We used the results of these simulations to aid in the selection of a new computer to replace our IBM 360/65 system. The environment at that time was limited to batch-type processing and the primary metrics under investigation were CPU time, channel load, and memory capacity (the number of jobs that could run concurrently).

The current environment is more complex. CAS is no longer restricted to batch work in an MVT operating system. We now have a mixture of batch, specialized on-line, and TSO applications in a Virtual Storage (VS) operating system. In this environment, the metrics of CPU time, channel load, and memory capacity now become secondary in importance to paging rate, working set size, and response time. Potential changes to the environment include MVS, mass storage, and distributed on-line processing. This will further complicate the simulation task and will introduce newer metrics for consideration such as the utilization of individual resources on an "on-demand" basis as well as staging and destaging time and device considerations.

Changes in computer simulation packages include new attempts to more closely approximate the internal operation of today's computers as well as improved mechanisms for building simulation profiles of the desired workload. The latter change has a significant impact on the batch program model building task by automating this time consuming process (1). However, part of the time that is saved through the use of an automatic model builder is now lost through the increased complexity associated with defining today's computer hardware and software characteristics to the simulator.

Within our range of experience, the basic approach to simulation has not changed. However, the traditional data input and logic building steps are now more appropriately named "Developing Workload Profiles" and "Defining Computer Hardware/Software Characteristics." There are two major reasons for making this distinction. First, the new environment requires a different set of skills for each step and second, the latter step becomes an integral part of the traditional validation loop. The validation loop compares simulated results with empirical data. If the results do not compare favorably, then the simulated computer hardware/software characteristics must be altered and the validation cycle must be repeated. This process continues until the simulated results approximate actual results within acceptable tolerances. Then the model can be considered "calibrated" and subsequent simulations can be made with some level of confidence in the results.

While computer and simulator technology have improved over the years, the simulation validation process has had little attention and support. This "tuning" process continues to be the activity that requires the most effort in terms of human resources. Validation is at the core of many simulation failures and frustrations. It also emphasizes the frailties of the simulation education process and the complexity of today's simulation packages.

SIMULATION INTEGRITY

Armed with formal knowledge and perhaps experience in systems analysis, modeling processes, queuing theory, and the mysteries of "black box" software, the simulation analyst feels he is prepared to answer the real world's question, "What if...?" Instead, the analyst may be faced with a downfall as he is confronted with the practical realities and intricacies of the simulation process. The downfall doesn't come during the drudgery of data gathering or in working out the logical subtleties of the model. It comes when the output is presented and the user, client, or customer asks the next question, "How good is the answer?"

Over time, there are advancements in technology and analysts become more proficient in generating models and developing generalized simulation software. But over the same time, circumstances arise which make it more difficult, if not impossible, to ascertain the reliability of simulation results with any guarantee of accuracy, confidence, or credibility. Consequently, the analyst now finds that more time and effort must be spent in the validation cycle to establish the integrity of the simulation results. Indeed, there are occasions where the inability to complete the validation cycle forces the abandonment of the whole simulation exercise.

The causes of the validation dilemma cannot be attributed to any single factor or chain of events. Our experience indicates that the difficulties contributing to the validation problem exist in three critical areas: simulator complexity, data availability, and personnel skill requirements.

SIMULATOR COMPLEXITY

It's a simple truth that today's simulation software is more complex because the computer systems themselves are more complicated. Commercially available or home-grown software has matured to the point of adequately simulating hardware components in a relatively simple batch system. With this software, the impact of model changes or workload variations can be predicted -- and validated within a reasonable degree of certainty. Today's more sophisticated computer systems not only feature advanced, integrated circuits with faster components, but also involve highly complex control software. A combined environment of batch, time sharing, and other real time applications operating under the control of a VS operating system is not uncommon.

This shift in emphasis from hardware to control software leaves batch-type simulators with a severe handicap. It is not enough for them to treat the operating system as a simplified black box when 50% of the CPU cycles are caused by operating system functions in a VS configuration. This amounts to simulating half a computer model. One half of the model is difficult to validate, and the other half is impossible to validate.

The emphasis on control software also changes the focus on the types of questions asked of a simulation. The concern of DP management now is to determine how to make the best use of this operating system overhead and thus defer the time when more hardware acquisitions need to be made. The objective of the simulations is now optimized throughput, and the questions are in terms of the desired number of active initiators, their priorities, and changes caused by varying input/output (I/O) rates. Internal queuing, paging activity, and their causes and effects are also of major concern. Since the operating system black box approach does not allow access to these specific metrics -- either for manipulation of the model, or for validation of the results -- the simulator algorithm must be made more detailed and more explicit. This of course leads to more complicated simulator software and thus completes the circle of system complexity.

There is an independent development in technology which has not made systems more complex, but has magnified the problem of validating simulation results. This independent development includes the increased and widespread use of hardware monitors, machine accounting, and resource utilization systems. Through the use of this technology, many installation personnel are now more familiar with the operation of their systems than ever before. This increased awareness of computer performance characteristics focuses more attention on the validation process. More metrics are visible and more opportunities exist to validate the specific situations being modeled. Users no longer accept general answers based on loosely-validated general models. They expect the validation process to compare a selected subset of models against the empirical data which is available to them. Unless this comparison is within some acceptable tolerance at this level of detail, the users will place no confidence in the results of a larger-scale simulation.

DATA AVAILABILITY

It's well known that the availability of empirical data is essential to the simulation process. But, the simulation analyst soon finds that the classical sources for this data, namely, system documentation and the analyst who designed the system, are not adequate for his purposes. Systems documentation was never designed to support simulation projects except at a very general level, and thus it rarely contains the detail that is necessary to construct a model. It is often out-of-date and does not coincide with the current operation of the system. Most systems analysts are deeply involved in designing a system under severe time constraints, or they have been reassigned to another project by the time the simulation interview is ready to take place. In either case, the analyst has neither the time nor the recollection that is necessary in order to define an accurate model of the system.

The most reliable and available sources of empirical performance data that can be used for model generation are found in three places:

- Data base access and file usage statistics
- Computer-oriented performance data
- Computer-produced machine accounting data

The data base statistics can provide detailed information on file accessing patterns and data element usage. Hardware and software monitors gather performance data on equipment, and internal software, such as the IBM System Management Facility (SMF) feature of the operating system, supplies the machine and job accounting data. In most cases, all these sources of data must be used for building models of complex computer systems. These sources of data are also vital in the validation cycle.

Organizing this supply of data into terms and structures meaningful to the simulator software has long been a time-consuming manual and intellectual process. Advances have been made in this area, however. Improved mechanisms exist for accepting this data in machine form and automatically building workload profiles for input to the simulation software. But, much of the intellectual effort still remains. Due to either the lack of specificity in the model building algorithms, or to a batch orientation in the simulator itself, a considerable effort in hand-coding models is still required for some applications. For example, if the automatic model builder does not or cannot distribute on-line or TSO transactions over the "connect" time of the terminal, the simulator will receive transaction models that look like batch processes. As a result, the simulator will process a model of typical file records and report that eight hours of on-line transactions can be processed in one minute. Needless to say, the validation process will have little success in correlating these results with actual performance. To eliminate this problem, the model of each type of on-line transaction must be coded separately. For those installations which perform the majority of their work in an on-line mode, the volume of manual model preparation required to handle the message traffic can become almost overwhelming.

PERSONNEL SKILLS

Simulation is not a programmer's tool; it is a specialist's tool. It requires the combined skills of an applications programmer, systems analyst, statistician, systems programmer, and hardware specialist to fully utilize its capabilities. Application of these skills begins with the model building process, continues through the simulation exercise and validation cycles, and ends only when the simulation results are accepted.

In the model building process, analyst skills are required to define the problem, to identify essential metrics and the sources of data, and to structure the model. If simulation software is already available, the model must be expressed in terms acceptable to that software. If model building software is unavailable or inadequate, then some programming may be done to extract, translate, and compare the raw data into suitable input for the simulator.

These same skills are applied in the validation process. Since most empirical data is in the form of machine accounting data, the output from the simulator must be interpreted or converted in order to place the measurements on a common base. Further, this base must be in a report form that is familiar to management or other users of simulation. To obtain this common base, it is often necessary to write special conversion programs that will translate the results of the simulation into report formats which can be compared with system utilization reports that are already in use at the installation.

For many simulation applications, defining the computer hardware and control software characteristics requires the knowledge of skilled systems programmers and/or hardware experts who are generally familiar with the interaction of hardware and software on the computer. It is especially important to have an understanding of the various system interactions that are required in order to execute a hardware I/O operation. Also required is a knowledge of the operation of the major queues in the operating system, the overall mix of instructions on their system, and the potential simultaneous processing capability of the files in each system to be simulated.

Simulation does not come easy, even for someone who possesses all of the above characteristics. The analyst must thoroughly understand the simulator itself and the manner in which it processes model inputs. In general, a lack of knowledge about the internal logic of the simulator is the major obstacle in the validation process. During validation, the analyst modifies simulator inputs to bring the results closer to actual experience. Since the workload profile input is based on empirical data, the majority of the tuning involves modifications to the hardware and software definitions. This tuning can be a time-consuming trial-and-error process since some hardware/software definition changes have little effect while other changes may have gross effects on the results of the simulation. Moreover, tuning one aspect to match empirical data may seriously detune another aspect if the analyst doesn't understand

Validation (Continued)

how the simulation program is constructed or is not alert for these possible side effects.

Assume, for the moment, that the validation cycle is successful, and the subset of the model problem tracks reality within acceptable tolerances. The analyst must now introduce additional workload profiles, extrapolate new volumes, and/or change hardware/software definitions to simulate the full scope of the postulated problem. Even with full knowledge of how the simulator handles these inputs, there can be no guarantee that the final simulated results are as accurate as the validated case. It is more probable that the results are less accurate since some simulator functions may not have been exercised or tested during validation. This is particularly true when the object of the simulation is an activity which does not yet exist and for which calibration data is unavailable. And yet, this is exactly the case where a good simulation can be most helpful to a designer or decision-maker.

The problem of educating simulation personnel can be separated into two levels: one, learning to accurately define and code workload profiles and configuration definitions and two, learning to manipulate the simulator to achieve the desired results. While the first level is concerned with the basic techniques required to adequately represent a computer application in simulator terms, the second is concerned with manipulating the logic of the simulator itself in order to insure that it is representing a life-like situation. Training for the first level is inadequate in that formal education does not really address the topic except at a theoretical level. Training in the second level is non-existent in vendor-supplied packages and is a lengthy on-the-job learning process for proprietory software.

SOME POSITIVE STEPS

There are a number of general problems facing the simulation analyst today. Most experienced users anticipate difficulties in system complexity, data gathering, or skilled resources and take steps to reduce the impact of these problems. However, it is interesting to note that they do not always attack the real issues. Contrary to the conventional wisdom, it is not the complexity of the application which causes difficulty; it is the complexity of the simulator software itself. The effort required to gather data for model building is far less than the effort required to validate the results. And finally, a host of skills is required, not just a sound knowledge in simulation techniques.

Our experience has been that the impact of these problems is felt more severely in the validation process than in many other steps of the simulation process. Lengthy, and sometimes, unsuccessful validation cycles have the effect of amplifying the already high cost of running simulations, and of increasing the skepticism of those who receive the simulation results.

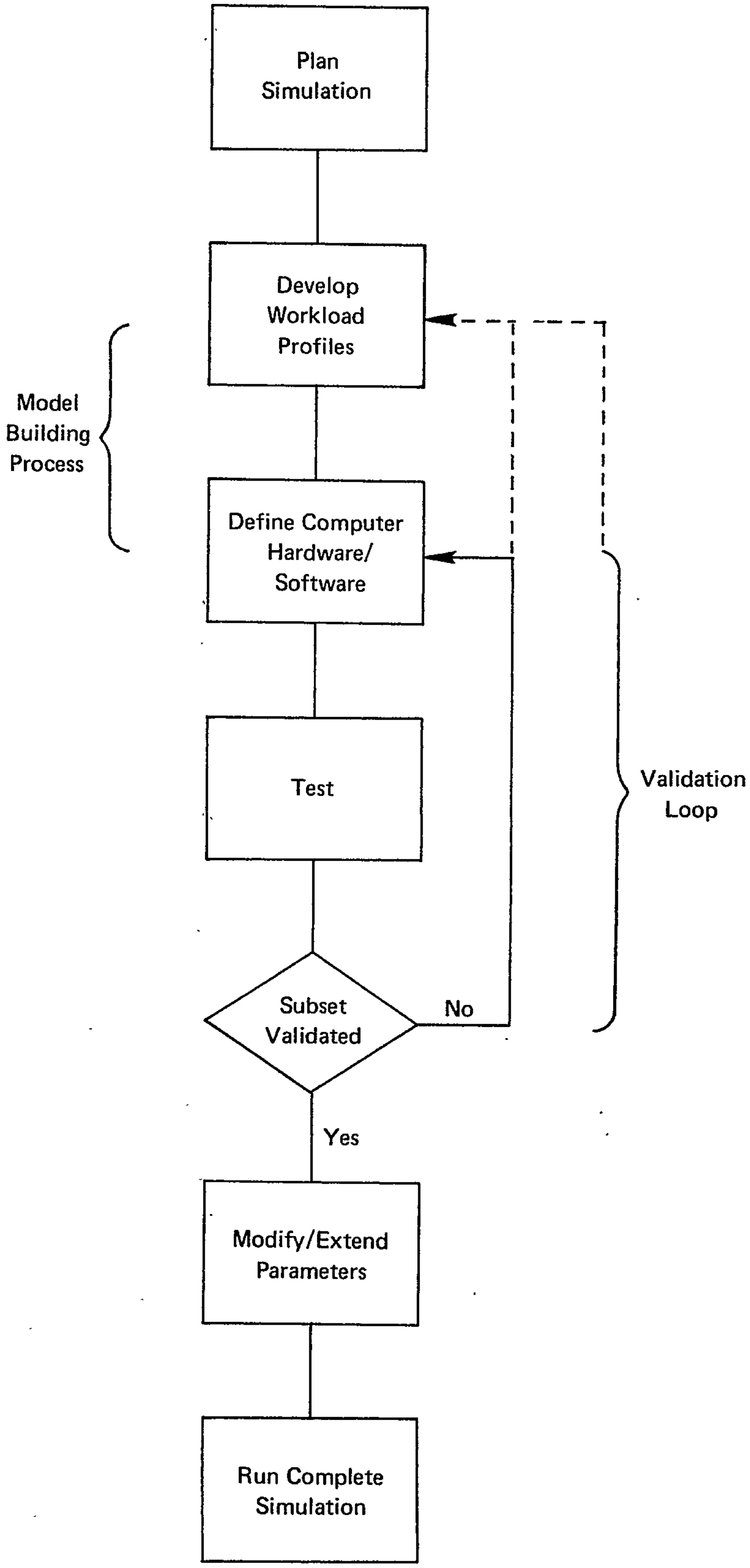
Our experience has also identified some corrective steps which can be taken by the developers of simulator software. The suggestions deal with the general principle of adding more synthetic properties to the simulation software, thus reducing the amount of manual analytical interaction that takes place in today's simulation exercises. Perhaps the ultimate limit in this direction is a simulator which adaptively tracks the solution and reports optimized results.

One step is to develop a set of algorithm-generating programs which could be executed on the simulation user's computer. These programs would act as test drivers to develop service rates of various system queue points under various conditions, thus developing their own queue curves. This approach is similar to that taken on trace-driven modeling techniques as described in earlier papers by Dr. Sherman (2) and by Noetzel and Herring (3). Only the major operating bottlenecks of a particular installation should be considered as candidates for automatic queue curve generation. We also suggest that standardized, machine-readable data sources should be used as input to the algorithm generators. Two sources come to mind -- SMF and hardware monitor data. This approach would eliminate much of the guesswork associated with setting the variable parameters in today's simulators. Of course, these queue curves would be valid only on the particular computer on which the programs were run. The challenge remains for the vendors of simulation software to adapt algorithm generators to the functions of a wide variety of computer models.

The next step is to enable the input of the simulator to be in the same terms as that of the output. Again, we suggest that these terms be in SMF, or some equivalent with which the users are familiar. To assist in the step previously mentioned, simulator pre-processors would synthesize this data into its resource queue demand components for processing by the algorithms of the simulator. This step would immediately reduce the amount of time required in the validation process to resolve differing sets of figures. Further, in the model generating process, hand-coded models would be reduced to an assemblage of SMF definitions.

Automatic model generators are becoming commonplace, but there is still little software available for automatic validation. With the combination of features just described, i.e., SMF type I/O and automatic algorithm generators, considerable progress can be made toward an automatic validation cycle. This automation could occur by comparing the simulated output to original input (both are SMF), modifying the algorithms where required, and executing the simulation again. This iterative process could continue until the simulator was validated according to some parameterized tolerance.

Some attention needs to be given to the typical need to run a sequence of simulations for a sequence of variables in certain input parameters. Simulation is an arduous enough task without having to resubmit runs with only minor alterations to the postulated problem. The ability to accept and


Plan Simulation
Develop Workload Profiles
Model Building Process
Define Computer Hardware/ Software
Test
Validation Loop
Subset Validated
No
Yes
Modify/Extend Parameters
Run Complete Simulation

process ranges of input values on a specified set of variables would add to the flexibility of the simulation process.

We also suggest that designers introduce some type of code to calculate how long a simulation will run based on the input it must process. With all the effort involved in preparation and validation, nothing is more exasperating to the analyst than to see his job abort because it took longer to run than he expected it to.

If these types of improvements were made to simulation software, then the users could devote more of their time to simulation as a tool instead of as an end in itself.

SUMMARY

Although computer and simulation technology have improved over the years, little improvement has been made in the critical simulation validation process. Throughout this paper we have attempted to focus attention on the validation process as an integral part of any simulation exercise and on the factors that influence the result of the validation process. These factors include such items as simulator complexity, contemporary computer system complexity, data availability, and the extensive personnel skills that are required in order to complete the simulation exercise. Along with the identification of problem areas we have also included some suggestions for corrective steps which the developers of simulation software may take in order to reduce the amount of manual analytical interaction required to perform a simulation exercise. Overall, we have tried to present these topics from an experienced user's point of view.

BIBLIOGRAPHY

(1) Edgecomb, G. D., "Automatic Model Building," NTIS, March 1974, PB-232 033/1WC.

(2) Sherman, S., "Trace Driven Modeling: An Update," SIMULETTER, Vol. 7, No. 4, (1976), pp. 87-91.

(3) Noetzel, A. S. and Herring, L. A., "Experience With Trace Driven Modeling," SIMULETTER, Vol. 7, No. 4, (1976), pp. 111-119.

THOMAS N. TRYGGESTAD is an Operations Research Analyst with the Federal Computer Performance Evaluation and Simulation Center (FEDSIM), Alexandria, Virginia, working primarily in the area of computer system simulation. Previous to joining FEDSIM in 1972, he spent three years working in the area of computer performance evaluation at the Air Force Data Systems Design Center, Gunter AFS, Alabama. He received a Bachelor of Science degree (Mathematics) from Augustana College (S.D.) in 1966 and a Master of Computing Science degree from Texas A & M University in 1968.

SIMULATION TECHNIQUES FOR THE EVALUATION OF DATA BASE SYSTEMS

Thomas N. Tryggestad

Federal Computer Performance Evaluation
and Simulation Center

I. INTRODUCTION

A. There are two distinct but related questions that usually generate a data base management system (DBMS) simulation study.

1. "What can a DBMS do for us?"

2. "Which DBMS shall we choose?"

B. This discussion outlines techniques that can be used in answering these questions.

C. There are some general considerations which apply to all techniques discussed.

1. The accuracy of the results depends on the accuracy of the input data and the care with which the simulation activity is conducted.

2. The use of a DBMS frees analysts and programmers from most concerns about the physical representation and storage of data. The simulation analyst is not freed from these considerations and, in fact, will likely have more difficulty in obtaining data on the physical aspects of a DBMS than on the typical file structure.

II. METHODS OF DEFINITION

A. IMS-type Definitions in SCERT or CASE

1. Description

a. Packaged simulators (SCERT, CASE) have built-in definition aids for hierarchical structures such as IMS.

b. They have no inherent network capabilities.

c. The definition structure generates multiple I/O's and the additional processor overhead encountered in DBMS accessing.

d. The overall effect is to add to the existing definition structure to account for additional processing generated by the DBMS.

2. Considerations

a. IMS has changed since the SCERT and CASE capability were developed.

b. Other DBMS using hierarchical structures can be defined using this technique as long as the accessing methods are the same as IMS or are properly redefined for the simulator.

c. The simulation analyst must always deal with the physical relationships, i.e., the actual physical activity that takes place, rather than the logical relationships.

B. Inverted DBMS Definitions

1. Description

a. Inversion is at the other end of the logical DBMS spectrum from IMS.

b. Conspicuous similarities exist between IMS and an inverted DBMS when only the physical aspects of accessing the data base are considered.

c. The differences lie in amount not in type.

d. The capability mentioned in the previous section can be used to represent the relationship between the index levels and record levels of an inverted DBMS.

2. Considerations

a. The above statements hold for retrieval and update-in-place, but not necessarily for record insertion.

b. The use of this technique requires careful examination of all library characteristics and accessing method descriptions.

C. DBMS Definition Using IA Tables (CASE)

1. Description

a. This area is unique to CASE simulator.

b. CASE uses decision-table (IA tables) logic in the calculation of device accessing time for random access devices. The decision tables can be modified or added to by the user.

c. Sequential, index-sequential, and random file organization and accessing exist within the CASE simulator.

d. A DBMS, especially one that has its own unique accessing method, can be defined using these tables.

2. Considerations

a. The level of detail required to use this technique is greater than for the previous techniques.

b. This technique is applicable to all types of DBMS, hierarchical or inverted, as well as less easily categorized data or file management systems such as TOTAL, INQUIRE, etc.

c. This technique can be used in conjunction with the previous techniques.

D. Physical Level Definition

1. Description

a. DBMS activity ultimately results in a certain number of I/O's and a certain amount of CPU activity.

b. DBMS developers have available this type of information which is referred to as physical level data.

c. With physical level data there is no need to be concerned about logical structure for definition purposes and the I/O and CPU activity can be input directly to the simulator.

2. Considerations

a. The DBMS vendors may have generalized physical level information available; however, they must know the details of particular files before their information can be specific.

b. The use of this type of information allows for ease of initial definition, but does not allow for flexibility in simulating the effect of changing workloads, changing logical structures, or adding new applications.

c. The use of physical level data is not confined to package simulation; it may be applied to discrete simulation languages and analytical modeling techniques.

E. ECSS Simulations

1. Description

a. The Extendable Computer Simulation System (ECSS) is a general purpose simulation language with no special features for the definition of DBMS-type structures.

b. When dealing with physical level data, it is not necessary to be concerned with a DBMS structure; therefore, the use of ECSS is appropriate.

c. Work is being done currently to add a structure for the definition of DBMS to ECSS.

2. Considerations

a. The same limitations on flexibility in simulating the effect of changing workloads, etc., as noted in the previous section apply here as well.

b. Using physical data with ECSS is quite similar in effect to using it with SCERT or CASE.

F. Analytical Modeling Techniques

1. Description

a. Queuing theory and other analytical techniques can be applied to physical level data to obtain results similar to simulation language or package results.

b. The type of results that can be calculated include queuing time, length of queues, throughput, wait time, and the probabilities associated with different queue lengths.

c. The various queues within the system, both software and hardware, can be defined and analyzed using queuing equations, and the composite picture then developed.

2. Considerations

a. Analytical techniques do not require the knowledge or the expense of a simulation language.

b. In a complex environment, such as a DBMS, the calculations can become difficult to handle manually.

c. There are automated queuing models available and being used, e.g., Analytic Solutions to Queues (ASQ).

III. EVALUATION CRITERIA

1. The ultimate results of a DBMS simulation study are expressed in the same terms as the results of any other simulation study, i.e., resource utilization, queuing characteristics, response times, and system throughput.

2. The presence of a DBMS structure usually adds another level of complexity to the problem.

3. Acceptable levels for each of the categories of results depend on the situation being analyzed just as in any simulation study.