Forecasting, the evaluation of effects of various strategies, is discussed. Emphasized are the quantitative techniques used in forecasting and the formulation of equations to represent functional relationships.

Also presented are two example forecasting applications—demand analysis of a consumer product and a financial forecasting model.

Forecasting techniques

by M. Aiso

Forecasting is currently becoming more important and complex in the planning activities of business and government. Computer-assisted forecasting helps to make these forecasts more accurate.

One type of forecasting required in planning activities is, in many cases, the evaluation of effects of various planned projects or strategies such as investments or marketing promotion. Another type of forecasting, which simply sketches the future without assuming particular changes in strategies, is the setting of a standard against which the performance of a specific strategy is evaluated.

Generally speaking, the forecast of the first type, an evaluation of strategy, is performed by forecasting models that express functional (also called causal) relations among relevant factors. This type of model, if adequately formulated, can indicate the turning points of future trends in response to possible changes in management action. Detection of the turning points is one of the most important concerns of forecasting in the planning process. Therefore, the need for relational forecasting models has increased substantially in industry at various organizational levels.

This paper emphasizes the use of relational models for forecasting. Presented are a general review of forecasting techniques, the formulation of equations to represent the relationships among factors, the estimation of parameters in a model, the evaluation of a model, and forecasting using the established model. An example forecasting program based on these techniques, the IBM Forecasting And Modeling System (FAMS) program product, is explained and some example forecasting applications are discussed.[1]

For the reader who desires an overview of forecasting techniques and applications, it is suggested that he read the following section and then proceed to the section entitled "A forecasting system." The remainder of this paper is a more rigorous and technical presentation of the mathematical equations used in forecasting.

Forecasting techniques and considerations

Techniques of forecasting vary depending on the context of the forecast. A number of forecasting techniques have been developed and can be classified into two major categories: qualitative techniques and quantitative techniques. Qualitative techniques are mainly based on human judgment, and future estimates may be obtained through qualitative data such as experts' opinions or information about special events. The DELPHI methods and panel consensus are typical examples. They are mainly used when data are scarce as when a product is first introduced into a market. Quantitative techniques apply various statistical techniques to historical time-series data for predicting future events. These techniques can be divided into three groups:

- Time-series analysis.
- Single-equation regression model.
- Simultaneous-equation regression model.

time-series analysis

Time-series analysis attempts to discover underlying trend and particular patterns from historical data. Based on the analysis, forecasting is performed extending the trend of the past into the future. In this type of forecasting a continuation of historical patterns is assumed, and the influence of outside factors is not taken into account.

single-equation regression model

Statistical forecasting is widely based on *regression*, a statistical procedure to determine the relationships of, for example, sales volume to various external indicators that are thought to have significant influence on it. The following is an example where a_1, . . . ,a_4 are the parameters to be estimated:

$$\text{Sales volume} = a_1 + a_2(\text{Price}) + a_3(\text{Disposable personal income}) + a_4(\text{Amount of advertising})$$

simultaneous-equation regression model

In a relationship in which sales volume is influenced from price, amount of advertising, and so forth, one can observe that sales volume is likely, in turn, to influence price and amount of advertising. For example,

$$\text{Price} = b_1 + b_2\ (\text{Sales volume}) + b_3\ (\text{Costs})$$

Table 1 Criteria for choosing forecasting techniques[2]

	Techniques		
Criteria	*Time-series analysis*	*Single-equation regression model*	*Simultaneous-equation regression model*
Consideration on external conditions	No	Yes	Yes
Turning-point detection	Yes (limited to seasonal or cyclic change)	Yes (limited)	Yes
Consistent prediction or related events	No	No	Yes
Dynamics of the flow system	No	Yes (limited)	Yes
Accuracy			
Short term (0–3 months)	Fair to Excellent	Good to Very Good	Good to Very Good
Medium term (3 months–2 years)	Poor to Good	Good to Very Good	Very Good to Excellent
Long term (2 years and up)	Very Poor	Poor	Good

can be formulated simultaneously with the previous relationship to form a two-way causation in sales volume and price. The presence of a two-way causation makes a simultaneous-equation model necessary (rather than the only approach to a single-equation model). Simultaneous-equation models have the ability to describe more complex and mutually influencing relationships by introducing as many equations as are necessary to represent the relationship.

There is no universal criteria to determine which forecasting technique is best. Some suggested criteria of choosing better forecasting techniques, summarized in Table 1, are now discussed.

Forecasting should consider and accommodate external conditions—that is, reflect existing environments, possible anticipated environmental changes, and potential policy changes. A capability to detect forecast turning-points due to changes in strategy is also a requirement. A forecasting technique should provide consistent numbers for related events such as multiple forecasts of sales, costs, and revenue. Additionally, a forecasting model

should be able to express the dynamic nature of the model variables since they may have time-dependent and interacting influences among themselves. The accuracy of a forecast is dependent on the adequacy of model formulation and treatment, but it can be associated with particular techniques involved. Although it is difficult for the author to indicate the accuracy of each technique in general, some other effort has been expended in this area and has been reported.[2]

Single-equation regression model

The single-equation regression model, which expresses a linear relationship between a dependent variable Y and k-1 independent variables $X_2, X_3, \ldots, X_k$, has a general form of:

$$Y_t = b_1 + b_2 X_{2t} + b_3 X_{3t} + \cdots + b_k X_{kt} + u_t$$

for $t = 1, 2, \ldots, n$; denoting n observations. The dependent variable Y_t can denote, for example, the sales amount of a product, and the independent variable X_2 can indicate advertising expenditure, and X_3 can represent the price of the product. The u_t is called *disturbance* (or error) that is added in the model because of the following reasons:

- Human behavior consists of many relevant factors.
- Human response has unpredictable elements of randomness.
- Observation contains error.

In matrix notation, the model can be rewritten as:

$$y = Xb + u$$

where $y = \{Y_1 Y_2 \cdots Y_n\}$

$$X = \begin{bmatrix} 1 & X_{21} & \cdots & X_{k1} \\ 1 & X_{22} & \cdots & X_{k2} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 1 & X_{2n} & \cdots & X_{kn} \end{bmatrix}$$

$b = \{b_1 b_2 \cdots b_k\}$

$u = \{u_1 u_2 \cdots u_n\}$

and $\{\}$ denotes column vector. The constant (intercept) term b_1 requires the insertion of a column of units in the X matrix.

least-square estimation of parameters

Assumptions for the least-square estimation of unknown parameters are:

1. $E(u) = 0$
   The mean of disturbances is zero, for all t.
2. $E(uu') = s^2 I_n$

The variance of disturbances is a constant (s^2), and the covariance is zero, for all t.

3. $E(xu') = 0$
   $x = \{X_{i1}X_{i2} \cdot \cdot \cdot \cdot X_{in}\}$
   Each of the independent variables (X_{it}, $i = 2, 3, \cdot \cdot \cdot, k$; $t = 1, 2, \cdot \cdot \cdot \cdot, n$) is independent from the disturbance u, and each of the independent variables is a set of fixed numbers.
4. X has rank $k < n$
   No exact linear relation exists between any of the independent variables, and the number of observations exceeds the number of coefficients to be estimated.

Let $\hat{b}$ denote the estimated values of b. The relationship then becomes

$$y = Xb + e$$

where e represents residual. One should distinguish between disturbance u and residual e. The disturbance u shows the error in y which is related to X through the unknown value of b; the *residual* e shows the error in y when it is related to X through an actually estimated value of parameter $\hat{b}$.

The coefficient b can be estimated in such a way that the squared sum of residuals:

$$e'e = (y - X\hat{b})'(y - X\hat{b})$$

has the least value. To obtain the value $\hat{b}$ which minimizes $e'e$, $e'e$ is differentiated and equated to zero:

$$\frac{\partial(e'e)}{\partial\hat{b}} = -2X'y + 2X'X\hat{b} = 0$$

If $X'X$ is nonsingular,

$$\hat{b} = (X'X)^{-1}X'y$$

and this is the estimate of unknown coefficient b.

The computation of $(X'X)^{-1}$ is possible only if the matrix $X'X$ is nonsingular. If two or more independent variables are perfectly correlated, the matrix becomes singular, and the computation of its inversion (the computation of b estimates) is impossible; and if the variables are highly correlated, the computed b values will not be reliable. These phenomena are called *multicollinearity*. Efficient ways of avoiding multicollinearity cannot always be found, but some recommendations would be to purge inappropriate variables by checking simple correlations or to weaken the relationship by transforming variables. An example is that of taking the difference from the previous period.

evaluation

After regression coefficients are estimated by the least-squares estimation, the estimated structure of the model is evaluated to

determine how well it represents reality. There are three types of information that determine the effectiveness of the estimated model:

- A priori information.
- Statistical test values.
- Comparison of actual and computed values.

A priori information. A priori information is the theoretical or practical knowledge one possesses before attempting statistical operations. Economic theory or business practice provides the main source of knowledge. Examples of a priori information are a combination of variables, and sign and magnitude of the parameters. In a combination of variables a set of independent variables must give a plausible explanation of the dependent variable in each estimated equation. Thus each variable and a whole set of independent variables must be logically related to a dependent (or explained) variable. The plus or minus sign and magnitude of the parameters are not usually known beforehand for all parameters. Some parameters, however, such as tax rates, interest rates or marginal profit ratios, are roughly known without applying estimation procedures and are checked with the results of computation.

Statistical test value. From the result of estimation computation, various measurements can be obtained to determine whether the estimation is meaningful. Presented in the sections that follow are four measurements which are most frequently used to test the validity of the equation results. They are:

- Coefficient of determination.
- The t-values for significance test of coefficients.
- Standard error of equation.
- Durbin–Watson d statistic.

The *coefficient of determination,* R^2, is defined by:

$$R^2 = 1 - \frac{\sum_{t=1}^{n} e_t^2}{\sum_{t=1}^{n} y_t^2}$$

using small letters to denote deviations from arithmetic means ($y_t = Y_t - \overline{Y}$). The second term in the right-hand side is the ratio of the variation of residuals or "unexplained" to the total variation of the Y about their sample mean; that is, the second term denotes the proportion of the variation that is not explained by the least squares. Thus the value R^2 indicates the proportion of the variation in Y "explained" by the least-squares regression equation. For instance $R^2 = 0.78$ means that 78 percent of the

variations of the dependent variable Y about its mean can be explained by the independent variables X_i $(i = 2,3,\cdots,k)$. $R^2 = 1.00$ is perfect correlation.

The degrees of freedom of $\sum_{t=1}^{n} e_t^2$ are $n - k$, and those of $\sum_{t=1}^{n} y_t^2$ are $n - 1$. Generally speaking, the smaller the degrees of freedom, the larger the coefficient of determination. This fact is inconvenient for comparison of several regressions with different degrees of freedom. To facilitate comparison, the R^2 value can be adjusted for the degrees of freedom. This is usually denoted by $\bar{R}^2$ and defined by:

$$\bar{R}^2 = 1 - \frac{\sum_{t=1}^{n} e_t^2 \Big/ (n-k)}{\sum_{t=1}^{n} y_t^2 \Big/ (n-1)}$$

$$= 1 - (1 - R^2)\frac{n-1}{n-k}$$

$$= R^2 - (1 - R^2)\frac{k-1}{n-k}$$

To derive *t-values for significance tests of coefficients* for the $\hat{b}_i$, assume that the disturbance u has a normal distribution. Now the assumptions for u are:

- The mean is zero.
- The variance is a constant s^2.
- The covariance is zero.
- The distribution is normal.

and the assumptions can be compactly written as:

u is $N(0, s^2 I_n)$.

Examining the distribution of $\hat{b}$, one obtains:

$$\hat{b} = (X'X)^{-1}X'y$$
$$= (X'X)^{-1}X'[Xb + u]$$
$$= b + (X'X)^{-1}X'u$$

Hence, any $\hat{b}_i$ is equal to b_i plus a linear function of u which has a multivariate normal distribution. Thus $\hat{b}_i$ has a normal distribution.

The mean of $\hat{b}_i$ is:

$$E(\hat{b}) = E[b + (X'X)^{-1}X'u]$$
$$= E(b) + (X'X)^{-1}X'E(u)$$
$$= b$$

The variance-covariance matrix of $\hat{b}_i$ is:

$$\begin{aligned} E[(\hat{b}-b)(\hat{b}-b)'] &= E[\{(X'X)^{-1}X'u\}\{(X'X)^{-1}X'u\}'] \\ &= E[(X'X)^{-1}X'uu'X(X'X)^{-1}] \\ &= (X'X)^{-1}X'E(uu')X(X'X)^{-1} \\ &= s^2(X'X)^{-1} \end{aligned}$$

The variance of $\hat{b}_i$ is the ith term of the principal diagonal of $(X'X)^{-1}$ multiplied by s^2 (the variance of u_i).

In summary $\hat{b}$ has a multivariate normal distribution specified by:

$\hat{b}_i$ is $N(b_i, s^2a_{ii})$

where a_{ii} is the ith principal diagonal element of $(X'X)^{-1}$. The value $e'e/s^2$ has an X^2 distribution with $n-k$ degrees of freedom. Finally, e and $\hat{b}$ are determined to be independently distributed:

$$\begin{aligned} e &= y - X\hat{b} \\ &= (Xb+u) - X[(X'X)^{-1}X'(Xb+u)] \\ &= u - X(X'X)^{-1}X'u \\ &= [I_n - X(X'X)^{-1}X']u \end{aligned}$$

Substituting the above values gives the independence of e and $\hat{b}$.

$$\begin{aligned} E[e(\hat{b}-b)] &= E[\{I_n - X(X'X)^{-1}X'\}uu'X(X'X)^{-1}] \\ &= s^2X(X'X)^{-1} - s^2X(X'X)^{-1} \\ &= 0 \end{aligned}$$

The t-distribution can be used for testing $\hat{b}_i$ since $\hat{b}_i$ is $N(b_i, s^2a_{ii})$, and $\sum_{t=1}^{n} e_t^2/s^2$ has an independent X^2 distribution with $n-k$ degrees of freedom.

From the definition of t- distribution,

$$t = \frac{(\hat{b}_i - b_i)/s\sqrt{a_{ii}}}{\sqrt{\left(\frac{1}{s^2}\sum_{j=1}^{n} e_j^2\right)\Big/(n-k)}}$$

$$= \frac{\hat{b}_i - b_i}{\sqrt{\sum_{j=1}^{n} e_j^2 \Big/ (n-k)}\sqrt{a_{ii}}}$$

is obtained, which follows the t-distribution with $n-k$ degrees of freedom, where a_{ii} is the ith principal diagonal element of $(X'X)^{-1}$.

A hypothesis that a certain regression coefficient b_i is zero—that is, the independent variable X_i has no effect on the dependent variable Y—can be tested by computing a specific value $t_\diamond$ as :

$$t_\phi = \frac{\hat{b}_i}{\sqrt{\sum_{j=1}^{n} e_j^2 \Big/ (n-k)}\sqrt{a_{ii}}}$$

Assuming a significance level of e percent, and a value of t_e from the t-distribution table with $n-k$ degrees of freedom is obtained, the test of the hypothesis is:

- If $|t_\phi| \geqq t_e$, the hypothesis is rejected.
- If $|t_\phi| < t_e$, the hypothesis is accepted.

It is an often-accepted practice to consider as meaningful only those variables with a t-value of at least plus or minus 2.0.[3]

The dispersion of the values computed by the regression equation in fitting the historical data may be measured by the variance of the disturbance u of the equation. However, since u is not directly observable, the residual e is used. From the sum of squared residuals and the division by its degrees of freedom, the estimated value of the variance of equation S^2 is obtained as:

$$S^2 = (e'e)/(n-k)$$

The positive square root of S^2 is the *standard error of equation*. The smaller the standard error of equation is, the better the regression results. It can be expected, for example, two thirds of the observations may fall within a range of plus or minus one standard error from the estimate of the equation.

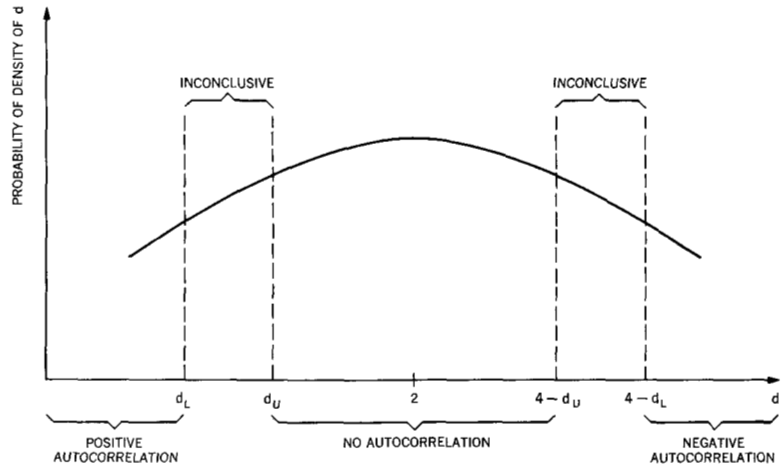
Durbin-Watson d Statistic. One of the assumptions on the disturbance u in the application of the least-squares estimation is that the u has zero variance; that is, the u is non-autocorrelated. If this assumption is not satisfied, the variance of the regression coefficient becomes large so that the use of the regression model for prediction is not justified. To test whether the assumption is satisfied, the *Durbin-Watson d statistic* is calculated from residuals e_t as follows:

$$d = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

The Durbin-Watson d statistic table gives upper (d_U) and lower (d_L) limits of the significance levels of d. The test of the hypothesis that u has no autocorrelation is as follows:

- If $d \leqq d_L$, or $4 - d_L \leqq d$, the hypothesis is rejected.
- If $d_U \leqq d \leqq 4 - d_U$, the hypothesis is accepted.
- If $d_L < d < d_U$, or $4 - d_U < d < 4 - d_L$, the test is inconclusive.

Figure 1 Distribution of d



A diagrammatic representation of the test is shown in Figure 1. As the figure illustrates, a d-value of about 2 is often considered as meaningful to conclude that u has no autocorrelation.

Comparison of actual and computer values. In each estimated equation $y = X\hat{b} + e$, observed values for independent variables $(X_2, X_3, \cdots, X_k)$ are substituted to obtain the estimate of the dependent variable, $\hat{y}$, for each period:

$$\hat{y} = X\hat{b}$$

The discrepancy, or residual e, between the observed value y and the estimated value $\hat{y}$ is computed and examined for each period. This method, sometimes referred to as the *partial method*, is applicable to only one equation—namely, a single-equation model or to each individual equation in a simultaneous-equation model.

forecasting

The period for which data for the estimation is available is called the *sample period.* The period subsequent to the sample period and for which the future values are to be forecasted is called the *forecast period.* If the values of independent variables for the forecast period are available, the forecasted values of the dependent variable can be calculated in the same way as the partial method. The independent values are provided by means of:

- External information.
- Policy or management objectives.
- Other techniques such as extrapolation by growth rates.
- Other models such as a master or sub-model.

Simultaneous-equation regression model

In the following simultaneous-equation model:

$$Y_{1t} = c_{11} + b_{12}Y_{2t} + c_{12}X_{1t} + u_{1t}$$

$$Y_{2t} = c_{21} + b_{22}Y_{1t} + c_{22}Y_{2(t-1)} + u_{2t}$$

Y_{1t} and Y_{2t} are to be forecasted, and they are called *endogenous variables*, $Y_{2(t-1)}$ in the second equation is a *lagged endogenous variable*. The values of X_{1t} are always given from outside the model; hence, X_{1t} is called an *exogenous variable*. The exogenous variables and the lagged endogenous variables are called *predetermined variables*.

The general form of a linear model containing g simultaneous relations (endogenous variables) and k predetermined variables can be written in matrix form as:

$$By_t + Cx_t = u_t$$

where:

B = Coefficient of endogenous variables

$$= \begin{bmatrix} b_{11} & \cdots & b_{1g} \\ \vdots & & \vdots \\ b_{g1} & \cdots & b_{gg} \end{bmatrix}$$

C = Coefficients of predetermined variables

$$= \begin{bmatrix} c_{11} & \cdots & c_{1k} \\ \vdots & & \vdots \\ c_{g1} & \cdots & c_{gk} \end{bmatrix}$$

and y_t, x_t, and u_t are column vectors at time t of endogenous variables, predetermined variables, and disturbance, respectively:

$$y_t = \begin{bmatrix} y_{1t} \\ \vdots \\ y_{gt} \end{bmatrix} \qquad x_t = \begin{bmatrix} x_{1t} \\ \vdots \\ x_{kt} \end{bmatrix} \qquad u_t = \begin{bmatrix} u_{1t} \\ \vdots \\ u_{gt} \end{bmatrix}$$

The model in the form of $By_t + Cx_t = u_t$ is called a *structural form model.*

If the matrix B is nonsingular, we can define the *reduced form* model as:

$$y_t = -B^{-1}Cx_t + B^{-1}u_t$$

The reduced form model expresses the endogenous variables y_t in terms of predetermined variables x_t. The forecast is to obtain the values of y_t given the values of x_t. Thus the reduced form model is convenient for computing the forecast of endogenous values.

estimation

From the reduced form model, it is evident that each element of the endogenous variable vector y_t is related with every element of the disturbance vector u_t. This means if, for example, in the first equation (y_{1t} equation) endogenous variables other than y_{1t} (such as y_{2t}, y_{3t}, and so forth) appear in the equation, those endogenous variables y_{2t}, y_{3t}, and so forth have a dependency on u_{1t}.

This fact that some of the independent variables have dependency on the disturbance violates the assumption of:

$$E(xu') = 0$$

in the least-squares estimation when we apply the least-squares estimation to an equation that contains the endogenous variables as independent variables in the equation.

The *two-stage least squares* (TSLS) technique is a widely accepted method of estimation procedure in a simultaneous equation model containing the estimation problem previously described. The TSLS technique is applied to each individual equation in a simultaneous model. Thus we can consider the ith equation of the model and let it be expressed as:

$$y = Y_1 b + X_1 c + u$$

where:

- y is the column vector of n observations on the endogenous variable to be forecasted by this equation.
- Y_1 is the $n \times g$ matrix of the observations on the other current endogenous variables included in the equation (explaining endogenous variables).
- b is the column vector of g coefficients associated with Y_1.
- X_1 is the $n \times k$ matrix of the observations on the predetermined variables appearing in the equation.
- c is the column vector of k coefficients associated with X_1.
- u is the column vector of n disturbances.

The TSLS technique purges the explaining endogenous variables Y_1 of the stochastic components associated with the disturbance u in two stages. In the first stage, the least-squares estimation is

applied to the parameters of explaining endogenous variables (b) regressing on the predetermined variables appearing in the model (denoted by X), and the computed $\hat{Y}_1$ matrix is obtained.

$$\hat{Y}_1 = X\hat{b}$$
$$= X(X'X)^{-1}X'Y_1$$

This regression is justified since X has no correlation with u. In the second stage, y is regressed on $\hat{Y}_1$ and X_1, and the result of the TSLS estimate is:

$$\begin{bmatrix}\hat{b}\\ \hat{c}\end{bmatrix} = \begin{bmatrix}\hat{Y}_1'\hat{Y}_1 & \hat{Y}_1'X_1\\ X_1'\hat{Y}_1 & X_1'X\end{bmatrix}^{-1}\begin{bmatrix}\hat{Y}_1'y\\ X_1'y\end{bmatrix}$$

evaluation and forecasting

The evaluation and forecasting methods for a simultaneous-equation model are basically the same as those for a single-equation model except for the following two methods that take simultaneity into consideration:

- Total method.
- Final method.

The *total method* applies to all equations in a simultaneous-equation model. In the reduced form equation:

$$y_t = -B^{-1}Cx_t + B^{-1}u_t$$

observed values of predetermined variables x_t are substituted for all periods to estimate values y_t, and the differences (residuals) between actual values and estimated values are computed. This method has the advantage of checking the simultaneous nature of the model, whereas it is impossible to do so with the partial method. However, if the model has lagged endogenous variables to express the dynamic nature of the system, the total method fails to check how well the model can be used as a simulation tool because it uses observed values of lagged endogenous variables instead of values calculated by the model.

The total method enables the model-builder to improve his simultaneous equation model, in which lagged endogenous variables are involved, by comparing the results with those of the following final method. Since the total method regards the lagged endogenous variables as purely exogenous, it cuts off the effects of the time lags of endogenous variables. This makes the calculations so simple that the model-builder could tell the cause of any possible deficiency in his model related with time-lag feedback mechanism.

The *final method* also applies to all equations in a simultaneous-equation model when the model represents time-dependent dynamic nature by including lagged endogenous variables. In the reduced form equation, observed values of exogenous variables

are substituted for all periods. However, actual values of lagged endogenous variables are used only before the initial time period, and the values computed by the model are substituted for subsequent periods. This means that the computed endogenous values are fed back to the model and this method tests the dynamics of the model due to time lag. If the simultaneous-equation model does not contain lagged endogenous variables, the final method gives identical results with the total method.

Identification

From a structural form model having g relations in the form of:

$$By_t + Cx_t = u_t$$

the reduced form model was obtained:

$$y_t = -B^{-1}Cx_t + B^{-1}u_t$$

If the original structural form model is multiplied by a $g \times g$ nonsingular matrix A, the resulting model is:

$$(AB)y_t + (AC)x_t = Au_t$$

If the reduced form is derived from the new model, the result is:

$$\begin{aligned} y_t &= -(AB)^{-1}(AC)x_t + (AB)^{-1}Au_t \\ &= -B^{-1}Cx_t + B^{-1}u_t \quad (\text{for } (AB)^{-1} = B^{-1}A^{-1}) \end{aligned}$$

Hence, both the original and new models give an identical reduced form which means that the models with different values for all the parameters will generate the same distribution of dependent variables conditional upon the values of predetermined variables and disturbances. In this case, the estimated results of the original model and the new model are said to be *observationally equivalent* in that they may have exactly the same implications about observable phenomena. This point is known as the *identification problem* in the context of simultaneous-equation models. The problems is that many different sets of coefficients, (B,C) and (AB,AC), may be obtainable from a set of observations (y_t, x_t) and it is impossible to conclude which coefficients represent the true model.

The general rule for determining the identification status of any given structural equation is derived as follows. The structural equations with estimated coefficients of B and C are:

$$By_t + Cx_t = 0$$

On the other hand, least-squares estimators of reduced form coefficients can be obtained as:

$$y_t = Px_t$$

By substituting for y_t from the reduced form into the structural form, the following results:

$$BPx_t + Cx_t = 0$$

or $BP = -C$

The following are denoted for the ith equation:

g_Δ = number of endogenous variables included in the ith equation

$g_{\Delta\Delta} = g - g_\Delta$ where g is the number of endogenous variables in the simultaneous-equation model

$k_\diamond$ = number of predetermined variables (including a constant term) included in the ith equation

$k_{\diamond\diamond} = k - k_\diamond$ where k is the number of predetermined variables (including a constant term) in the simultaneous-equation model.

We can assume that the coefficients of the ith equation (b_i and c_i) are arranged in such a way that the nonzero elements appear first, being followed by the zero elements. This can be written as:

$$b_i = [b_\Delta 0_{\Delta\Delta}]$$

$$c_i = [c_\Delta 0_{\diamond\diamond}]$$

where:

$b_\Delta = [b_{i1} b_{i2} \cdots b_{ig\Delta}]$ $1 \times g_\Delta$ vector

$0_{\diamond\diamond} = [0\ 0 \cdots\cdots 0]$ $1 \times g_{\Delta\Delta}$ vector

$c_\Delta = [c_{i1} c_{i2} \cdots\cdots c_{ik\diamond}]$ $1 \times k_\diamond$ vector

$0_{\diamond\diamond} = [0\ 0 \cdots\cdots 0]$ $1 \times k_{\diamond\diamond}$ vector

The matrix P can also be partitioned in a corresponding way:

$$P = \begin{bmatrix} P_{\Delta\diamond} & P_{\Delta\diamond\diamond} \\ P_{\Delta\Delta\diamond} & P_{\Delta\Delta\diamond\diamond} \end{bmatrix}$$

and the ith equation can be regarded as:

$$[b_\Delta\ 0_{\Delta\Delta}] \begin{bmatrix} P_{\Delta\diamond} & P_{\Delta\diamond\diamond} \\ P_{\Delta\Delta\diamond} & P_{\Delta\Delta\diamond\diamond} \end{bmatrix} = -[c_\diamond\ 0_{\diamond\diamond}]$$

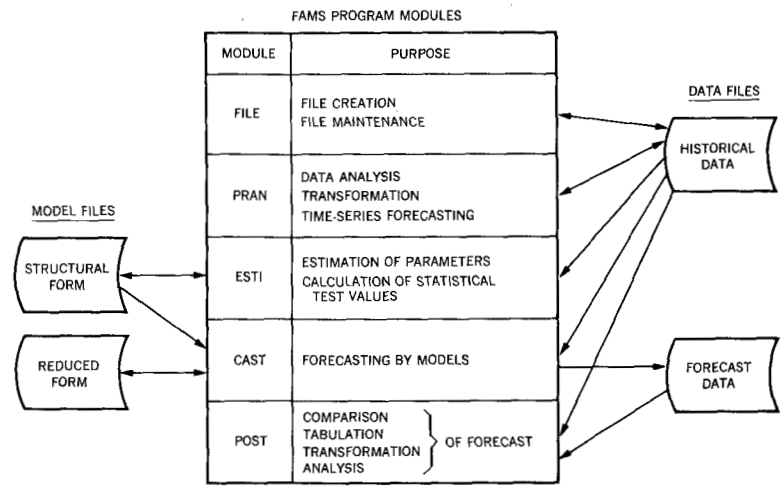
From this, the following are obtained:

$$b_\Delta P_{\Delta\diamond} = -c_\diamond \cdots\cdots (a)$$

$$b_\Delta P_{\Delta\diamond\diamond} = 0_{\diamond\diamond} \cdots\cdots (b)$$

If (b) can be solved for b_Δ, then $c_\diamond$ can be solved from (a). The vector b_Δ contains $g_\Delta - 1$ unknown coefficients since one of the b's in the structural equation is unity.

Figure 2 FAMS program modules and files



A necessary condition is that in order to obtain a solution for the $g_\Delta - 1$ unknowns in b_Δ, the number of equations in (b) must be at least $g_\Delta - 1$, namely:

$$k_{\diamond\diamond} \geqq g_\Delta - 1$$

In other words, the number of predetermined variables excluded from the equation must be at least as great as the number of endogenous variables included less one.

A necessary and sufficient condition for identification is that the number of independent equations in (b) is $g_\Delta - 1$, namely:

$$\text{rank } (P_{\Delta\diamond\diamond}) = g_\Delta - 1$$

In summary, the identification problem can be stated as reducing the values of parameters of the structural form (B and C) from a knowledge of the reduced form parameters (P). The identification problem is associated with each structural equation in a model.

A forecasting system

An example batch-mode forecasting system for System/360 and/370 Disk Operating System (DOS) and System/360 and /370 Operating System (OS) is the IBM program product, Forecasting And Modeling System (FAMS).[1] This program is a collection of statistical and data-handling routines based on the previously described forecasting techniques:

- Time-series analysis.
- Single-equation linear model.
- Simultaneous-equation linear model.

Table 2 Example partition sizes required for PAMS

Number of equations	*Number of variables*	*Number of observations*	*Partition size DOS*	*OS*
20	30	14	52K	70K
50	90	30	84K	102K
100	160	35	184K	202K

The equation (single or simultaneous) is based on regression or given by definition. The linear model (single or simultaneous) includes nonlinear combination of predetermined variables, and log-linear transformation.

FAMS provides a capability for creation and maintenance of data files, analysis and transformation of data, qualification of forecasting models, forecast of future values, and analysis of forecasted results. It also provides for updating of models, statistical tests, and summaries and comparisons to analyze and evaluate the models and their forecasted results.

There are five program modules in FAMS:

- Data file (FILE).
- Pre-analysis (PRAN).
- Estimation (ESTI).
- Forecast (CAST).
- Post-analysis (POST).

Their functions are depicted in Figure 2. Also shown are the four different kinds of permanent user files (model and data files) from which the information on the data and model is transferred to relevant module functions. Specific functions and features of the program modules are described in the appendix. The size of the model depends on the available partition size of main storage. Table 2 shows some example partition sizes in *K* bytes.

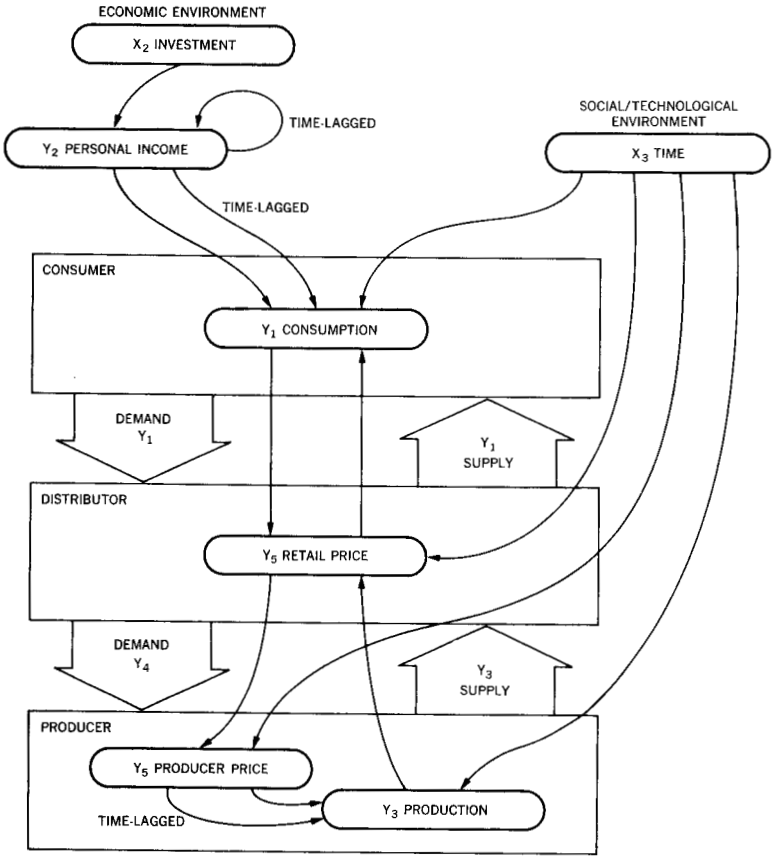
Example forecasting applications

Two examples of the application of functional relations in forecasting are now discussed.

demand analysis of a consumer product

A forecasting model can perform demand analysis of a consumer product. For example, the model could analyze the demand of food products. This model, whose flow is shown in Figure 3, represents the demand and supply between consumer and distributor, and between distributor and producer. The following equations comprise the model:

Figure 3 Demand analysis model



$$(1)\quad Y_1 = b_{11}Y_2 + b_{12}Y_5 + c_{11}Y_{2(-1)} + c_{12}X_3 + c_{10}$$

$$(2)\quad Y_1 = b_{21}Y_3 + b_{22}Y_5 + c_{21}X_3 + c_{20}$$

$$(3)\quad Y_2 = c_{31}X_2 + c_{32}Y_{2(-1)} + c_{30}$$

$$(4)\quad Y_3 = b_{41}Y_4 + c_{41}Y_{4(-1)} + c_{42}X_3 + c_{40}$$

$$(5)\quad Y_4 = b_{51}Y_5 + c_{51}X_3 + c_{50}$$

where:

Y_1 = food consumption per capita

Y_2 = disposable income per capita

Y_3 = production of agricultural products

Y_4 = production prices (received by farmers for food products)

Y_5 = retail prices of food products

X_2 = net investment per capita

X_3 = time

Demand function from consumer to distributor (1). Determining factors of this function are the retail price and personal income. As for the income, both the values of current period and those of previous period are considered. The values of the previous period explain an inertia of consumption behavior. The changes in habits and taste of consumers are introduced by an exogenous variable of time X_3.

Supply function from distributor to consumer (2). Assuming the demand and supply are balanced, the supply function can be represented by using the same variable Y_1 (consumption) as the dependent variable. The supply is explained from the retail price and the production amount. The time X_3 denotes changes in fabrication and marketing.

Income function (3). The income of the previous period and investments can explain the current level of income. This equation has a characteristic of statistical definition equation.

Supply function from producer to distributor (4). The supply to distributors is explained from two main factors: a quick response to the current production prices and one-period delayed response to the price. The time X_3 represents a trend of the increase of people engaged in production.

Demand function from distributor to producer (5). The demand from distributor to producer is considered to be measured by knowing how much of the retail price is received by producers.

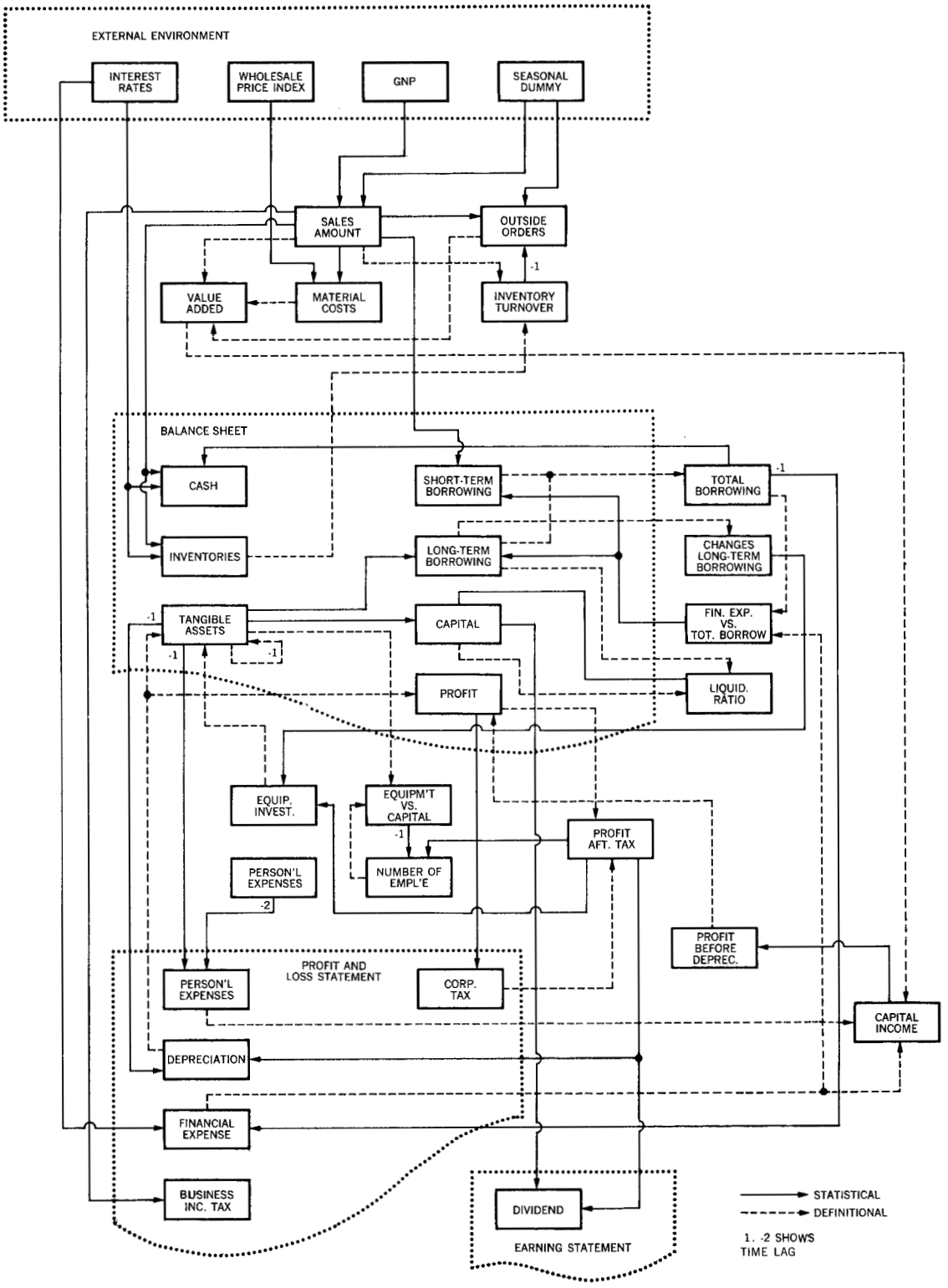
financial forecasting

A second example is a financial forecasting model, depicted in Figure 4. The general framework of the model consists of a given, external, usually uncontrollable environment such as interest rates, price index, Gross National Product (GNP), and a seasonal dummy (a variable representing a temporal effect of different seasons). The model also contains major items of financial statements (balance sheet and profit loss sheet) that are forecasted, taking into account the relationships between major items of financial statements as well as other relevant factors. This model has many uses, some of which are:

- To investigate a financial behavioral structure.
- To estimate future values of major financial items.
- To analyze profit versus owned capital ratio.
- To investigate the relation between profits and expenses.
- To evaluate a potential competitiveness.
- To compare financial structures between companies or industries.

Thus the future financial figures of a company or of an industry can be analyzed and forecasted on a scientific basis. In addition,

Figure 4 A financial forecasting model



the cause of changes in profit ratio provides a method of control to increase profit. These two characteristics, scientific forecasting and method of control, make forecasting a more reliable assumption for better decisions in present business planning.

Concluding remarks

Computer-assisted forecasting techniques have been presented. Time-series extrapolation techniques are based upon the historical pattern of the item to be forecasted. Of particular importance are causal (functional) models which take influencing factors into consideration. Regardless whether the causal model comprises a single relation or multiple relations, one must be aware of the various tests and evaluating methods in the stages of model building in order to obtain better forecast models. Building and analyzing a forecasting model normally requires data manipulation and calculation. Unless these tasks are performed in a systematic manner, a heavy burden is placed on the model-builder. Productivity in forecasting activities is increased by using a systematic computer programming system oriented for forecasting. Introduced as an example forecasting program was FAMS.

Further important steps forward are flexible linkage of techniques, linkage of models, and linkage of planning applications. By realizing these, forecasting will lay a foundation of management science applications and provide a framework for timely and sound decision making.

Appendix

The functional capability and features of each of the five program modules of FAMS are as follows:

FILE module

- Functions
  Creation
  Deletion
  Insertion
  Value modification
  Header modification
  Merge
  Sort
  Retrieval
  List

- Features
  Time-series data
  Cross-section data

PRAN module

- Functions
  Plotting data series
  Transformation
  Simple correlation
  Exponential smoothing
  Polynomial regression
  Seasonal decomposition

- Features
  Combination of functions
  Data (historical) file update (option)
  Card input for data (option)
  Symbolic data reference

ESTI module

- Functions and features presented in this paper
  Ordinary least squares
  Stepwise least squares
  Two-stage least squares
  Combination of estimation techniques
  Statistical measurements (regression coefficient, standard error of coefficient, t-value, coefficient of determination, standard error of equation, F-value, Durbin-Watson d statistic)

- Other features
  Inclusion of identity (or equation by definition)
  Maintenance of structural form model file
  Card input for data (option)
  Symbolic data reference

CAST module

- Functions and features presented in this paper
  Partial method
  Reduced-form derivation
  Total method
  Final method
  Specification of endogenous variable, which allows:
  1. Designation of endogenous variable among the right-hand-side explaining variables.
  2. Interchange of endogenous and exogenous variables
  Initial value tests
  "Identification" information of the model

- Other features
  Card input for data (option)
  Symbolic data reference
  Designation of alternative equation (to form a "complete" model)

POST module

- Functions
  Forecast accuracy analysis
  Growth rate and growth amount
  Multiplier analysis (impact in response to exogenous change, or "sensitivity" analysis)
  Residual summary
  Forecast summary
  Forecast tabulation
  Forecast comparison
  Forecast transformation
  Forecast transformation and comparison
  Case comparison (for different exogenous assumptions or "what-if" simulation)

- Features
  Comparison with plotting up to 3 series of data
  Transformation of any combination of endogenous and predetermined variables
  Symbolic data reference

ACKNOWLEDGMENT

The author wishes to acknowledge M. A. Girshick and T. Haavelmo for the information on the demand analysis model.

CITED REFERENCES

1. *Forecasting And Modeling System(FAMS) General Information Manual* (Program No. 5736-XS4 (DOS) or 5734-XS7 (OS)), Form No. GH19-4000, IBM World Trade Corporation, New York, New York, or IBM Corporation, Data Processing Division, White Plains, New York.
2. J. C. Chambers, S. K. Mullick, and D. D. Smith, "How to choose the right forecasting technique," Harvard Business Review **49,** No. 4, 45–74 (July–August 1971).
3. G. G. C. Parker and E. L. Segura, "How to get a better forecast," Harvard Business Review **49**, No. 2, 99–109 (March–April 1971).

GENERAL REFERENCES

C. F. Christ, *Econometric Models and Methods*, John Wiley and Sons, Inc., New York, New York (1967).

A. S. Goldberger, *Econometric Theory*, John Wiley and Sons, Inc., New York, New York (1964).

J. Johnston, *Econometric Methods*, second edition, McGraw-Hill Book Company, New York, New York (1972).

J. Kmenta, *Elements of Econometrics*, The Macmillan Company, New York, New York (1971).

E. Malivaud, *Statistical Methods of Econometrics*, North-Holland Publishing Company, Amsterdam, The Netherlands (1970).