This paper has two objectives. First, it seeks to highlight the crucial steps in econometric model building and their successful execution wholly within a SAS environment. Second, it proposes a way to exploit collinearity among explanatory variables with the aid of factor analysis. The model building exercise was undertaken to provide forecasts and market analysis for a major consumer product line. Figure 1 depicts the resulting sequence of events. Step one was the creation of a SAS data base containing (a) the dependent variables (members of the product line) and (b) the independent variables such as retail sales, retail prices and retail inventories which were outside the control of the firm, and promotional frequency, price discounts, etc., which were clearly subject to company marketing policy.

The other steps fell into the following categories:
- Data Analysis
- Construction and Validation of Product Models
- Dynamic Models of Independent Variables
- Application of the Model

DATA ANALYSIS

Since the goal of the project was a search for a mathematical relationship between each product in the line and variables suspected to have caused its historical movements, a first step was to generate plots, scatter diagrams and correlation matrices. The SAS procedures GMM and CORR proved helpful at this point. An additional procedure NPAIRS examined the variables for serial correlation. Both the graphical and correlation analyses showed that the dependent and independent sets of variables were strongly correlated. The correlation among the independent set, however, was enough to lead to possible parameter estimation problems. We felt that this would be prevented if we first identified the mathematical functions, if they existed, for the factor or factors giving rise to these high correlations. Once identified, we planned to use the factors in the estimation routines and to regard the original variables as expressible thus:

\[ V_j = a_{j1} F_1 + a_{j2} F_2 + \ldots + a_{jn} F_n \]  

Using the procedure FACTOR we proceeded as follows:

1. With the correlation matrix of the standardized independent variables as input, initial factors were obtained. In the pattern matrix of Figure 2, three mutually orthogonal factors are shown. The entries in the matrix are the regression coefficients in (1). Notice that some variables have large coefficients (loadings) on more than one factor.

2. Since we were searching for factors that represented a subset of related variables, the former were rotated so that variables obtained high loadings generally on only one factor. The rotation was not orthogonal (note the interfactor correlations) since the resulting factors were desired to be as interpretable as possible. The factor structure matrix in Figure 2 shows the correlations among factors and original variables after rotation. The scoring coefficient matrix expresses each factor as a linear combination of the variables:

\[ F_j = C_{j1} V_1 + C_{j2} V_2 + \ldots + C_{jn} V_n \]
In Figure 2, we give an example of the factor analysis options used. The generation of initial factors followed the principal components method. The keyword PROMAX indicates the selection of an oblique rotation.

CONSTRUCTION AND VALIDATION OF PRODUCT MODELS

The construction of models for the products was accomplished via the multiple regression routines of the procedure SYSLIN. The specification of each model used the results of the data analysis. Where appropriate, factors were introduced so that the formulation amounted to restricting the regression coefficients of an original variable to a value that depends on the correlation between it and the factor it helps define.

The standard tests for goodness of fit, such as $R^2$, $F$, etc., were employed for each model estimated. The $t$-tests showed that the parameters of the final models were all statistically significant. The signs on all coefficients were economically reasonable. No significant first order auto-correlation was present, that is, all D-W statistics were good. Residual analysis performed suggested randomness of calculated errors. To more completely validate the models, tests were conducted to determine how stable the coefficients were (test for structural drift). Ex-post analyses generated forecasts of recent history. Such forecast were compared with the actual data and the Mean Absolute Percent Error obtained.

$$\text{MAPE} = \frac{1}{k} \sum_{t} \frac{|A_t - F_t|}{F_t}$$

where $A_t$ is the actual at time $t$ and $F_t$ the forecast at $t$. The regression was run on $n-k$ observations and $k$ periods were forecast. A sample regression output is shown in Figure 4.

DYNAMIC MODELS OF INDEPENDENT VARIABLES

To generate forecasts of exogenous variables we identified and estimated dynamic models using the Box-Jenkins methods of the ARIMA procedure. As in the case of the products models, these were all subjected to diverse validation tests. Once exogenous forecasts were obtained, the product models were "ex-post validated" to determine whether the dynamic models were forecasting well. The results were positive when the method of equation (3) was applied.

SUMMARY AND CONCLUSION

The forecasting system has been successfully transferred to the user. Although the number of observations in each variable in the data base has grown by as much as 30-35 percent, estimation using the full set of data does not result in
significant changes in the coefficients. The system has passed the test of time. Currently the forecast performance is monitored by management through reports as shown in Fig. 5. Our recent experience demonstrated that the ongoing vigil is an important activity in the inventory control process as the forecast errors are used to adjust inventory levels.
A forecasting system draws from a conglomerate of analytical skills in business, economics, statistics and econometrics supported by software that allows the modeler, systems designer, and the user to work with maximum effectiveness and minimum loss of confidence. Our experience has convinced us that SAS goes a long way in that direction.

REFERENCES


