PROPOSED U. S. PASSENGER CAR AND LIGHT TRUCK SALES FORECAST MODEL

S. J. Olson, Allied Corporation
Dr. John Janakiraman, Miles Laboratories

INTRODUCTION

My job responsibility within a large automotive supplier company includes forecasting U. S. automotive vehicle sales (specifically passenger car and light truck). It is necessary to prepare an annual and five-year forecast for recommendation to top management for concurrence and approval. This forecast then provides the basis for corporation automotive financial and strategic planning.

Recently, the accuracy of automotive forecasting has been less than desired. Methods used are often highly subjective—based only upon economic service bureau and/or O. E. manufacturer assumptions. Although our forecasts have been conservative by design to others within the industry, they too have been far from reality.

Stimulated by exposure to computer regression analysis and SAS as a student in the Graduate Business Program at Indiana University in South Bend and Dr. Janakiraman’s systems analysis courses, I have spent considerable time developing a regression model as an aid to forecasting this difficult and uncertain market.

AUTOMOTIVE MARKET

Actual sales history of the individual market segments which comprise the dependent variable are depicted in Exhibit A. Using SASGRAPH, the 144 monthly data points were easily plotted to visualize the seasonality, the cyclicity and the divergent trends of the data for the period used to develop the model. During this time significant political, economic and market events occurred, such as the Oil Embargo, Iranian Revolution, Double Digits Inflation, High Interest Rates, Extended Recessions, Import and Light Truck Popularity which have changed the consumer and the structure of the U. S. automotive industry.

Strikes by the United Auto Workers (represented by vertical reference lines for the appropriate period) during 1973 against Ford and 1976 against Chrysler had little effect on Sales. An earlier strike in 1970 against General Motors had little effect on early 1971 sales. The low dealer stocks of new 1971 models at introduction time likely caused postponement of some purchases of GM vehicles until the stocks could be replenished and special orders filled in early 1971.

The “Chicken War” tariff of 1980 on imported trucks and the voluntary Japanese pass car import quotas initiated in 1981 (also represented by reference lines) appear to have had some effect on the sales of these vehicles by the looks of their recent performance to trend. Both of these political influences are expected to continue into the future. Possible domestic content legislation and further fledgling domestic production by foreign manufacturers could further influence and complicate this segment of the market.

MODEL

To reflect the events stated above within a model and develop conclusions which have relevance today, I chose to use monthly data to the extent possible for the period 1971 through 1982. The dependent variable, Total U. S. Passenger Car and Light Truck Sales including imported product, was selected to minimize market complications caused by the shifting consumer preference to light trucks and imports. Approximately 60 researched and derived variables of 144 observations each were chosen for their potential impact on the automotive consumer from an economic, psychological and technical perspective. The following five independent variables were used in the model and are ranked in order of contribution to $R^2$:

- $X_1$ - Consumer Confidence - Conference Board
- $X_2$ - Annual % Change in GNP per quarter - NIPA - U. S. Dept. of Commerce
- $X_3$ - Vehicle Seasonality - U. S. Dept. of Commerce
- $X_4$ - Vehicle Scrappage - R. L. Polk & Co.
- $X_5$ - Average Retail Price of Regular Gasoline - U. S. Dept. of Energy

SAS procedures PROC CORR, STEPWISE (FORWARD, BACKWARD and MAXR), RSQUARE and GLM were utilized to develop the model. SASGRAPH procedure PROC GPLOT was used initially to visualize the data in scatter diagram and linear format to determine the relationship between the various dependent/independent variable combinations. At present the following equation produces the highest $R^2$:

$$Y = 676.42 + 1.03X_1 + 6.54X_2 + 2.02X_3 + 0.06X_4 + 0.62X_5$$

Every effort was made to obtain monthly data for the model; however, some Department of Commerce data, specifically from the National Income and Product Accounts of the Bureau of Economic Analysis, is only available on a quarterly basis. Vehicle Scrappage data is determined annually as a residual of vehicles registered during the period versus vehicles-in-use growth. To use this data quarterly or annually would essentially eliminate other monthly data from the model because of SAS/GLM missing value treatment. Therefore, early Consumer Confidence, GNP and Vehicle Scrappage monthly data were interpolated from bimonthly, quarterly and annual data. Taking this liberty with the data assumes linearity between reporting periods which is not necessarily the case.

Four of the model's independent variables have a direct relationship with the dependent variable and one has an inverse. Each relationship seems to follow expected logic. As Consumer Confidence, GNP, Vehicle Seasonality, and Vehicle Scrappage increase so do New Vehicle Sales. The Retail Price of Regular Gasoline, as might be expected, negatively...
affects vehicle sales. Further analysis of the independent variables with components of the dependent variable (Domestic Pass Car, Imported Pass Car, Domestic Lt. Truck and Imported Lt. Truck Sales) revealed some interesting and conflicting relationships that, at first, were thought to be a potential problem with the model. Probably because of the diverse sales trends and the recent popularity of high mileage imported vehicles, opposing relationships occurred between the independent variables and these dependent variable components. Models were tested using these dependent variable components as separate dependent variables assuming that a multi-equation model for New Vehicle Sales would prove most accurate. However, this did not prove to be the case. The single equation model produced a higher $R^2$ value than any of the component models and it is much easier to administer requiring fewer independent variables to forecast.

Exhibits B thru F are comparative linear plots of the dependent/independent variable combinations used in the model. These serve as a useful tool to detect synchronous movement in cycles which may be useful in determining lead/lag relationships. Experimentation with the SAS/LAG function proved interesting but of little value to improve this model appreciably.

Numerous other independents were tried because of high correlation and an intuitively logical relationship with the dependent variable. Change in vehicle miles driven monthly holds promise for a future independent variable addition to the model. It must be studied further because the change is calculated from the month for the previous year and seasonality prevents the change being calculated simply from month to month. Because an apparent lag relationship exists it would seem that New Vehicle Sales could be an excellent independent variable for forecasting vehicle miles traveled per vehicle. The prime interest rate, the average retail new automobile selling price, commercial bank loan interest rates for new automobiles and average length of new automobile loan maturity also seemed to have a logical relationship with new vehicle sales. However, $T$ tests for variable significance proved otherwise and in addition they are highly correlated one to another and with the Gasoline Price variable which has significance and greater correlation with New Vehicle Sales. This could be a function of a relatively new phenomenon in the automotive market: The Rebate -- which has been viewed by some in the industry as a mixed-blessing. Although the overall effect of this form of price reduction was included with an insignificant variable, average retail new automobile selling price, its impact upon the consumer does seem to be highly significant. I have not found data regarding the rebate in a format readily adaptable to this model. Perhaps some sort of derived data could be generated from the period, amount and vehicle information that is available. I do believe that it is a factor which must be dealt with both for its stimulative and borrowed future sale effect to improve the accuracy of this model.

An ex post forecast of New Vehicle Sales was generated using actual independent variable data for the final year of the model: 1982 (see Exhibit I). Comparison of the actual monthly data with predicted values follows (units in thousands):

<table>
<thead>
<tr>
<th>Month</th>
<th>Actual</th>
<th>Forecast</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan</td>
<td>697.9</td>
<td>619.0</td>
</tr>
<tr>
<td>Feb</td>
<td>825.5</td>
<td>733.0</td>
</tr>
<tr>
<td>Mar</td>
<td>1,033.5</td>
<td>969.3</td>
</tr>
<tr>
<td>Apr</td>
<td>857.1</td>
<td>874.0</td>
</tr>
<tr>
<td>May</td>
<td>1,001.9</td>
<td>912.2</td>
</tr>
<tr>
<td>Jun</td>
<td>849.9</td>
<td>936.5</td>
</tr>
<tr>
<td>Jul</td>
<td>809.8</td>
<td>809.9</td>
</tr>
<tr>
<td>Aug</td>
<td>793.3</td>
<td>776.6</td>
</tr>
<tr>
<td>Sep</td>
<td>875.3</td>
<td>752.4</td>
</tr>
<tr>
<td>Oct</td>
<td>825.8</td>
<td>830.9</td>
</tr>
<tr>
<td>Nov</td>
<td>960.1</td>
<td>833.3</td>
</tr>
<tr>
<td>Dec</td>
<td>824.1</td>
<td>888.3</td>
</tr>
<tr>
<td>Total</td>
<td>10,346.2</td>
<td>9,645.4</td>
</tr>
</tbody>
</table>

On an annual basis the forecasted monthly results for this period varied from actual by 5.8%. This indicates that with perfectly forecasted independent variable data, model error is approximately 7%.

A further test could be performed for predicted dependent variable sensitivity to independent variable forecast error. Comparison of the error in forecasted New Vehicle Sales resulting from an introduction of known error to alternate independent variables would indicate how stable or robust the model is. Ideally, the change in error of the resultant forecast should be less than the amount of error introduced in the independent variable.

MODEL CRITIQUE

Generally, I am pleased with the initial results of the model. The $R^2$ or coefficient of determination of .356 can be interpreted to mean that the five independent variables selected explain about 85% of total New Vehicle Sales. T tests for individual independent variable significance indicates that all are highly significant -- ranging from 3.17 to 17.17 standard deviations from zero. The least significant variable has significance to a 99% confidence level.

A degree of autocorrelation exists as is indicated by the Durbin-Watson Statistic of 1.01 and is confirmed by the pattern in the plot of the model's residuals (see Exhibit G). This is somewhat unsatisfactory but a probable consequence of using time-series data. The data could be manipulated to eliminate this undesirable characteristic by using the SAS/DIF function but this would likely reduce the $R^2$ and perhaps not significantly improve the accuracy of the model.
CONCLUSION

As with any regression analysis the ability to accurately forecast the dependent variable is contingent upon the ability to accurately forecast the independent variable(s). Most of the independent variables in this model are perhaps as difficult to forecast as the dependent variable. However, they have high recognition within automotive industry management as probable determinents of New Vehicle Sales and indicate the level of economic activity in general. While the ultimate goal of any forecast is always accuracy, in a market where the only certainty is uncertainty, a more realistic and attainable goal may be to develop a forecast that is consistent with the assumptions on which it was based. I believe that this model will quantify New Vehicle Sales reasonably well in relation to the independent variables chosen and can be an extremely useful tool within a forecasting system.

While my model's accuracy in terms of R^2 leaves room for improvement, this experience and SAS have helped me to better understand the many variables impacting the automotive industry. The relatively user friendly SAS software package places an extremely powerful computer within easy reach of a computer neophyte. It allows the knowledgeable user to experiment within his field of expertise, communicating directly with a computer and obtain significant results that may not be achievable through a third party system/program approach.

S. J. Olson
Allied Corporation
Allied Automotive Sector
Bendix Chassis and Brake Components Division
401 N. Bendix Dr.
South Bend, IN 46620
Tele Area (219) 237-2262

PROPOSED U.S. PASS CAR AND LT TRUCK SALES FORECAST MODEL

ACTUAL VEHICLE SALES PER MONTH
DOMESTICS AND IMPORTS
PROPOSED U. S. PASS CAR AND LT TRUCK SALES FORECAST MODEL

VARIABLE PLOT

VEHICLE SALES VERSUS REGULAR GASOLINE – RETAIL PRICE

PROPOSED U. S. PASS CAR AND LT TRUCK SALES FORECAST MODEL

VARIABLE PLOT

VEHICLE SALES VERSUS CONFERENCE BOARD – CONSUMER CONFIDENCE SURVEY

PROPOSED U. S. PASS CAR AND LT TRUCK SALES FORECAST MODEL

VARIABLE PLOT

VEHICLE SALES VERSUS QUARTERLY CHANGE IN GROSS NATIONAL PRODUCT – INTERPOLATED MONTHLY

PROPOSED U. S. PASS CAR AND LT TRUCK SALES FORECAST MODEL

VARIABLE PLOT

VEHICLE SALES VERSUS VEHICLE SCRAPPAGE – IN MILLIONS