SAS GOES TO WALL STREET: PROC ARIMA AND THE DOW JONES STOCK INDEX

Joseph Earley, Loyola Marymount University
Michael J. Saxby, University of Chicago
Saul E. Sobel, Loyola Marymount University

Introduction

The purpose of this paper is to investigate annual movements of the Dow Jones Industrial stock average using the PROC ARIMA procedure. A univariate ARIMA model of the Dow Jones stock index is developed in the first part of this paper. ARIMA models of the Standard and Poor's and the New York Composite stock indices are also estimated for comparison. Forecasts are then calculated.

In the second part of the paper, an ARIMA Transfer function model is presented, using the money supply M1 as the input series. Integrated to this section of the paper is the use of the inverse correlation function (ICF), available as a model option under PROC ARIMA. While Box-Jenkins methods have been employed extensively using daily and monthly data, this paper uses annual data in order to take advantage of macroeconomic variables which do not become available in the numerator so frequently. The Dow Jones index is illustrated in Figure 1.

History of the Dow Jones Aggregates

The Dow Jones Company has published continuous aggregates for representative stock market averages since 1896. The current "high level of the average is a result of the long duration of the index. The basis of the average has never changed. Although stocks have split over the years, the industrial average has not. Starting on December 21, 1966, the Wall Street Journal began daily publica-
tion of changes from the previous close to the DJ stock average, calculated and expressed as a percentage basis as well as an absolute basis. The purpose of this was to facilitate comparability of the averages over time. The Dow Jones Company frequently emphasizes that the number should not be mistaken for dollars per share prices of stocks. Today the DJ averages are primarily used as market movement indicators, not dollar averages of current market prices. The industrial average first computed on May 20, 1916 included 12 stocks. The list was expanded to 20 in 1916 and finally to 30 in 1928. Substitutions are made whenever it is thought that the stock is not representative of the industrial sector. The thirty stocks included in the Dow Jones industrial average were not randomly selected, but rather were selected to be representative. Editors of the Wall Street Journal take responsibility for changing the compo-
nents of the DJ average. In addition to the industrial average, the Dow Jones Company publishes an average of 30 transportation common stocks (this was called the railroad average until renamed in January, 1976 as the Transportation index) and the DJ utility common stocks (the Utility index). The purpose of publishing these averages is to give the reader a general idea of fluctuations in the securities market over time.

Theory of Stock Price Movements

The primary factors that influence the movement of stock prices and therefore comparison indices are those which affect the discounted values of the cash flows resulting from holding the stocks. Thus the principal economic factors that need to be investigated are the money supply, national income, the rate of inflation, the rate of interest and the federal deficit. There are certainly other factors which may have an effect on stock prices and thus the index. For example other important factors such as OPEC price changes, the international deficit, International arm agreements all affect the stock market as a whole. For the purpose of developing ARIMA transfer function models, we will concentrate on the aforementioned macroeconomic variables. The fact that these are the principal variables which influence expected cash flows implies that they ultimately are the factors responsible for movements in the price of stocks.

Univariate ARIMA Modeling

The initial stage in modeling the movement of the Dow Jones Industrial average involves developing a univariate ARIMA model. For comparative purposes models were also developed for the SP and the DJUC stock indices. A univariate ARIMA (p,q,d) model may be represented as:

\[ \theta(B) L^d Y_t = \phi(B) a_t \]

where:

\[ \theta(B) = 1 - \theta_1 B - \cdots - \theta_p B^p \]

\[ \phi(B) = 1 + \phi_1 B + \cdots + \phi_q B^q \]

\[ a_t \sim N(0, \sigma^2) \quad \sigma > 0 \]

\[ L^d \]

\[ Y_t \sim N(0, \sigma^2) \quad \sigma > 0 \]

\[ a_t \]

\[ Y_t \]

are random shocks (errors) assumed to be normally and independently distributed with mean zero and constant variance.

Figure 1

Dow Jones Industrial Index: 1950 to 1987 (Annual Data)
It was necessary to difference the DJ series to achieve stationarity. The ACF, PACF, PCF, and autocorrelation check for white noise indicated that the difference filter converted each series to both stationarity and a random error process. The autocorrelation check of residuals after the estimation stage confirmed this conjecture. Forecasts with confidence intervals were then calculated based on the model developed in the estimation stage. It was determined that for the DJ series, as well as for the SP and FTSE, a random walk model was a satisfactory representation. The model can be written as:

\[ y_t = y_{t-1} + e_t \]

Transfer Function Models

In the second stage of investigation, explanatory variables are used to improve model performance. A transfer function model is one in which explanatory variables are used as inputs in order to predict future values of the output series. Using ZAD notation, the model is written as:

\[ y_t = \sum_{j=0}^{p} \delta_j y_{t-j} + \sum_{j=0}^{q} \theta_j z_{t-j} + e_t \]

where \( y \) and \( z \) are independent ARIMA time series and the \( \delta_j \) coefficients are called transfer function weights of lagged response weights. The \( \theta_j \) will be restricted to allowable functional forms according to the results of an analysis of the cross-correlation function.
The modeling strategy used is that suggested by SAS:

Modeling Strategy for Transfer Functions

a. Identify and estimate models for each input X.
b. Pre-whiten both X and Y.
c. Identify the appropriate transfer function form using the cross-correlation function.
d. Fit the transfer function followed by residual analysis.
e. Plot the transfer function with noise model.
f. Forecast both X and Y.

Schematically this strategy may be illustrated as:

Explanatory Variables

\[ I \frac{w(b)}{\delta(b)} \frac{X}{Y} \]

White Noise

\[ e_t \]

\[ (\hat{Y}_t) \]

\[ \hat{X}_t \]

\[ \hat{Y}_t \]

\[ e_t \]

where:

\[ w(b) = \sum_{i=0}^{n} w_i b^i \]

\[ \delta(b) = \sum_{i=0}^{n} \delta_i b^i \]

SAS computes transfer function models using a ratio of two finite polynomials in the backshift operator as:

\[ \epsilon_t = \frac{w(b)}{\delta(b)} X_t b^{-\theta} + \epsilon_t \]

Following are the results of modeling the BJ series using the moving average (MA) as an input series.

Initially we develop a univariate ARIMA model for X. A first difference removes an adequate filter for this series. Next, we pre-whiten the values of BJ and X using the difference operator, pre-whitening eliminates trend and seasonal components from the series and BJ allows us to determine the true relationship between BJ and X. It should be noted that some authors prefer to pre-whiten each series using the own ARIMA model. SAS pre-whitens the output series using the filter developed for the input series.

Cross Correlation Function

The key statistic necessary for identifying the form of a transfer function is the cross-correlation function. The CCF allows us to identify the relationship between series in the same manner as the autocorrelation function allows us to determine the relationships within a series. The cross correlation function is defined as:

\[ \rho_{xy}(k) = \frac{C_{xy}(k)}{\sigma_x \sigma_y} \]

where:

\[ \rho_{xy}(k) \]

is the cross correlation between \( x \) and \( y \)

which are then estimated using the sample cross covariances and cross correlations.

By observing sample patterns which approximate theoretically derived patterns, we can select the appropriate form for the transfer function.

Calling BJ pre-whitened output series: \( \hat{X}_t \), and the pre-whitened input series: \( \hat{Y}_t \), we write the estimated model as:

\[ \epsilon_t = u_t + \frac{w(b)}{\delta(b)} \frac{X_t b^{-\theta}}{\delta(b)} + \epsilon_t \]

where:

\[ w(b) \] is lag where we encounter lags-six in the sample cross covariance function.

\[ \delta(b) \] is an unknown scale-correlation function.

\[ \epsilon_t \] is the error component.

\[ \epsilon_t \] is the stationary BJ series.

and \( \hat{X}_t \) is the stationary BJ input series, \( \hat{Y}(b) \) and \( \hat{Y}(b) \) are called the numerator and denominator factors.

The SAS output for the transfer function between the output series BJ and the input series X is shown below:

<table>
<thead>
<tr>
<th>TRANSFER FUNCTION MODELING 362 OF BJ SERIES</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUTOREGRESSIVE MODELS</td>
</tr>
<tr>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>COUTH</td>
</tr>
<tr>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>INVERSE AUTOCORRELATIONS</td>
</tr>
<tr>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>PARTIAL AUTOCORRELATIONS</td>
</tr>
<tr>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>TRANSFER FUNCTION MODELING 362 OF BJ</td>
</tr>
<tr>
<td>AUTOCORRELATION TABLES FOR WHITE NOISE</td>
</tr>
<tr>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>PARAMETER ESTIMATE</td>
</tr>
<tr>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>CONSTANT ESTIMATE = 2.724664</td>
</tr>
<tr>
<td>AUTOCORRELATION CHECK OF RESIDUALS</td>
</tr>
<tr>
<td>---------------------------------------------</td>
</tr>
</tbody>
</table>
Conclusion

The general conclusion reached after estimating and forecasting with numerous transfer functions is that the transfer function ARIMA model is superior to the univariate ARIMA model as a forecasting tool. For example, it is interesting to note that the transfer function model forecasted large increases in the M2 series for 1984 and 1985 and a large decline for 1986. This seems consistent with the large changes in the money supply from 1984 to 1985 and that the subsequent drop from 1984 to 1986. The univariate random walk with drift ARIMA model was not able to incorporate the effect of the money variables on the M2 series. It should be noted that the general ARIMA transfer function methodology is also capable of modeling interventions and, as a special case, performing regression analysis with the error structure specified. PERA AUTOREG would not be an appropriate procedure since it does not allow for differencing.

References


Following are the SAS code and the data set used in the transfer function model.

FILE: B1B2A 345
PREL DATA ANALYSIS
IDENTIFY VAR = M2
ESTIMATE VAR = Cetary_Change + M1 + H0RNI_Change
ESTIMATE INPUT = [Cetary_Change + M1 + H0RNI_Change] * 4
FORECAST LEADS OUT = B DINTERS;
PROC MVEAR DINTERS = [Cetary_Change + M1 + H0RNI_Change] * 4;
DINTERS;
PROC MVEAR DINTERS = [Cetary_Change + M1 + H0RNI_Change] * 4;
DINTERS;

FILE: B1B2A 345
FORECAST FUNCTION MODELING = 3 SJ OF MVEAR + 3
PROC MVEAR DINTERS = [Cetary_Change + M1 + H0RNI_Change] * 4;
DINTERS;

The section above by John Mckelvey is in the introduction to Pierce: The Dow Jones Averages, 1885-1980.