

A DYNAMIC MODEL FOR STOCK PRICE PREDICTION
UTILIZING THE STEPWISE MULTIPLE LINEAR REGRESSION TECHNIQUE

by

Jorge H. Gonzalez

B. S., Kansas State University, 1965

A MASTER'S REPORT

submitted in partial fulfillment of the
requirements for the degree

MASTER OF SCIENCE

Department of Industrial Engineering

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1966

Approved by:



Major Professor

LD
2668
R4
1966
G643
C.2

TABLE OF CONTENTS

BALANCE SHEET.....	4
ASSETS.....	7
LIABILITIES.....	9
STOCKHOLDERS' EQUITY.....	10
INCOME STATEMENT.....	12
FINANCIAL RATIOS.....	16
NET WORKING CAPITAL.....	16
CURRENT RATIO.....	17
QUICK ASSETS.....	17
INVENTORY TURNOVER.....	18
BOND RATIO.....	18
LEVERAGE.....	19
NET PROFIT RATIO.....	19
INTEREST COVERAGE.....	20
EARNINGS PER COMMON SHARE.....	21
PRICE EARNINGS RATIO.....	22
COMMON STOCK RATIO.....	22
NET BOOK VALUE OF A SHARE OF COMMON STOCK.....	23
NET BOOK VALUE OF ALL COMMON STOCK.....	24
NET BOOK VALUE OF ALL BONDS.....	24
GROSS NATIONAL PRODUCT.....	25
CONSTRUCTION OF THE MODEL.....	26
SELECTION OF VARIABLES AND CALCULATION OF REGRESSION COEFFICIENTS.....	29
OPERATION OF MODEL AND COMPUTER PROGRAM.....	42

LOGIC OF THE COMPUTER PROGRAM..... 45

OPERATING PROCEDURES..... 50

SAMPLE PROBLEM..... 52

SUMMARY AND DISCUSSION OF MODEL..... 55

CONCLUSION..... 55

ACKNOWLEDGEMENTS..... 57

REFERENCES..... 58

APPENDIX..... 59

TABLE OF PLATES

PLATE I: A Sample Balance Sheet.....	6
PLATE II: A Sample Income Statement.....	14

INTRODUCTION

The stock market in itself is, by nature, a world of uncertainty. In this report, however, it is maintained that some of the stock market's variation can be explained. This report deals with a statistical investigation of the stock market in order to predict stock prices in the future. The statistical method utilized is a multiple linear regression. The key to the success of such a model is the determination of the factors that significantly affect the stock price. Once these are found, they are incorporated into a multiple linear regression equation which constitutes the statistical model.

In using a regression model in the traditional sense one aspect of stock price forecasting is not considered. This is the determination of the timing for buying or selling stock in order to maximize returns on investments. This aspect is instilled in the present model by predicting the high and low stock prices for the next period. Thus, when the stock price reaches the high range predicted, a selling action should be taken. Conversely, if the stock price falls into the predicted low range, the signal is to buy. With this knowledge of the expected high and low prices the market itself will indicate the timing for the buying and selling decisions.

In order to obtain the predicted high and low levels for a

stock price statistical inference is used. Predictions are made for the high and low stock price utilizing the high and low values from previous years as data in the multiple linear regression analysis. The values obtained are the predicted or expected high and low price for the next period. Associated with this expected value is a standard error of estimate. By utilizing the standard error of estimate the probability of the stock price reaching a certain level near the expected value can be calculated. The investor must decide, with a certain probability, at what level he will make a move in the market. For example, the investor wishes the stock to exceed his upper limit and to fall below his lower limit with the probability .975. The limits signaling his moves are calculated by subtracting two standard deviations from the high predicted stock price and adding two standard deviations to the low predicted stock price. When the stock prices exceed these limits, he will buy or sell, knowing with probability .975 that he made the right move.

Thus, this model provides a statistical investment logic. An advantage of this model is that it allows the market to signal the moves. But, as mentioned previously, the key to the success of this model is to determine the factors which influence the stock price. A reasonable choice of factors would be the various financial ratios of the firm in question. These ratios have been used by accountants and investors for many years to test the soundness of a company's operations, and also, the attractiveness of its securities. Thus, the objective in constructing a regression model for each firm is to select the particular set of financial ratios

for that firm for which the forecasted stock prices have the minimum unexplained variance. Therefore, the procedure is to obtain the data necessary from the financial reports for the firm in question, calculate the financial ratios and construct the multiple linear regression model utilizing these financial ratios as independent variables.*

In constructing the model the unexplained variation of the estimate is reduced by introducing the variables into the multiple regression equation. To determine the variable to introduce at each stage, the reduction of variance resulting from the introduction of a variable into the model is calculated. The maximum reduction is found, and the corresponding variable is introduced into the model if the reduction of variance is significant according to a prescribed level. This stepwise procedure of introduction of variables is continued until the maximum reduction of variance by the variables not in the equation is insignificant. When this

* These variables are dependent in a statistical sense, but are independent in a functional sense with respect to the forecasted stock price in the regression function. These variables will be referred to in this entire report as independent variables.

has been achieved, the model is in its final form. A multiple correlation coefficient, R , and an F statistic for the model are calculated for testing the model's significance. The model is then used to predict the dependent variable, the stock price, using the independent variables, the financial ratios, in the model. The two financial statements from which the information for these ratios is obtained are the balance sheet and the income statement. These statements are discussed below.

BALANCE SHEET

The balance sheet represents the financial picture on a particular day, the day which is given on the report, as though the business were momentarily halted. A sample balance sheet is listed in Plate 1. The balance sheet is basically divided into two columns. The assets are shown on the left side and the liabilities and stockholders' equity on the right. Both columns of this statement are always in balance. In the asset column is listed the goods and property owned as well as monies that others owe the firm. Under the liabilities are listed all the debts due, which are the creditors' claims against the assets. Listed under stockholders' equity is the amount of the stockholders' interest in the company.

EXPLANATION OF PLATE I

A sample balance sheet

BALANCE SHEET

Year ended December 31, 1966

ASSETS

Current Assets

Cash		\$XXXX	
Marketable Securities, at Cost (Market Value \$xxx)			XXXX
Accounts Receivable	\$XXXX		
Less: Provision for Bad Debts	<u>XXXX</u>		XXXX
Inventories			<u>XXXX</u>
Total Current Assets			<u>\$XXXX</u>

Fixed Assets

Property, Plant and Equipment

Land	\$XXXX		
Buildings	XXXX		
Machinery	XXXX		
Office Equipment	<u>XXXX</u>		
Total Fixed Assets	<u>\$XXXX</u>		
Less: Accumulated Depreciation			XXXX
Net Property, Plant and Equipment			<u>XXXX</u>
Goodwill, Patents, Trademarks			<u>XXXX</u>
Total Assets			<u><u>\$XXXX</u></u>

LIABILITIES AND STOCKHOLDERS' EQUITY

Current Liabilities

Accounts Payable	\$XXXX		
Notes Payable	XXXX		
Accrued Expenses Payable	XXXX		
Federal Income Taxes Payable	<u>XXXX</u>		
Total Current Liabilities			<u>\$XXXX</u>

Long-Term Liabilities

Total Liabilities			<u>XXXX</u>
			<u>\$XXXX</u>

STOCKHOLDERS' EQUITY

Capital Stock

Preferred Stock (characteristics and number of shares outstanding)	\$XXXX		
Common Stock (par value and number of shares outstanding)	XXXX		
Capital Surplus	XXXX		
Accumulated Retained Earnings	<u>XXXX</u>		
Total Stockholders' Equity			XXXX
Total Liabilities and Stockholders Equity			<u><u>\$XXXX</u></u>

ASSETS

Current Assets is the first entry on the asset side of the balance sheet. Current assets as defined by the American Institute of Accountants are those assets "which in the regular course of business will be readily and quickly realized, together with such additional assets as may readily be converted into cash without impairing the business or enterprise." Current assets may be divided into five main classes, which are cash, marketable securities, accounts receivable, inventories and prepayments.

Cash is the most liquid of all the assets. In this entry is included the currency in the petty cash fund and the money on deposit in the bank.

Market securities represent only highly liquid assets which can readily be converted into cash. The intention of management is an important factor in the determination of the classification of these investments. Management often decides to invest excess or idle cash in stocks, bonds and United States Government securities to earn dividends and interest. The securities must be readily marketable and subject to a minimum of price fluctuation in order to assure the funds, if needed, on short notice. The market value of these securities is commonly shown in parenthesis beside the entry of the securities at cost.

Accounts receivable include the book accounts which have not been collected from customers who have already received the services. It is common practice to show the offsets to this account,

such as allowances for probable losses and outstanding discounts. It is common knowledge that some bills will not be paid because of some disaster or misfortune. Therefore, a provision for bad debts is necessary in order that the figure for accounts receivable be realistic.

Generally, merchandise bought for resale, finished and partly finished goods manufactured for sale and materials and supplies purchased for use in production are considered to be the firm's inventory. Inventory is generally entered in the balance sheet at its cost or the market price, whichever is lower. This method of listing the inventory is conservative since the cost may be influenced by decline in prices, deterioration or obsolescence.

Prepayments are payments which have been made in advance of the receipt and utilization of goods and services. Insurance premium, rent, prepaid wages and salaries and deposits on contracts are illustrative of current assets of this type.

Fixed assets can be subdivided into land, natural resources, buildings and machinery and equipment. These assets are generally not intended for resale and are to be used in the manufacture of the product. These would include furniture, automobiles and trucks and all valuables belonging to the firm which are necessary in the allocation of their goods and services. The generally accepted and approved method of valuation for the fixed assets, with the exception of land, is cost less the depreciation accumulated to the date of the balance sheet. The figure is not intended to be the market value at the present or the replacement cost in the future. This is merely a representation of the money basically

invested in the manufacturing of the product. In this way, once it is clearly recognized that book values do not mean to be a representation of the present values, an intelligent analysis is possible. The sum of the value of land, natural resources, buildings and machinery and equipment constitute the fixed asset entry in the balance sheet.

LIABILITIES

Current liabilities are the debts or obligations which will mature within one year of the date on the balance sheet. Meaningful information can be gained by comparing the current assets to the current liabilities since current assets are the source from which payment of the current debts are made. Advance payments by customers who have not received their goods and services is included in the current liabilities. Thus, the current liabilities can be subdivided into two groups: the payables and the deferred revenue. The sum of these two would be listed as the total current liabilities.

The debts which will mature after one year from the date in the balance sheet are listed as long-term liabilities. Illustrative of this type of liabilities are mortgage bonds. In such a transaction the firm receives money from the bondholders and, in turn, are given a certificate, called a bond, as evidence of the loan. The date due and the percentage of interest is recorded on the bond and on the balance sheet. The principle is due on the

date specified, but, the interest is usually payable semi-annually. The bondholders are safeguarded by a stipulated written guarantee that if the company is unable to pay the bond when, and as promised, the bondholders have a claim before other creditors on certain assets which may be sold with the proceeds used to satisfy the debt. Total liabilities is the sum of the current liabilities and long-term liabilities.

STOCKHOLDERS' EQUITY

The total interest that the stockholders have in a corporation is the stockholders' equity. Stockholders' equity may be subdivided into three categories, which are capital stock, capital surplus and accumulated retained earnings.

Capital stock represents the amount contributed to the firm by its stockholders. It accounts for the permanent capital invested by stockholders in exchange for proprietary interest. The ownership claims are confirmed by the corporation by issuing stock certificates to the shareholders. There is considerable variation in the types or classes of shares issued by a corporation.

Preferred stock is any stock which ranks above another issue of the same corporation with respect to income distribution or assets available upon liquidation. Preferred stocks are usually callable or redeemable. Other variations may be converted into other securities or may participate in the residual profits. The characteristic differential is the priority over several optional

stocks offered by the corporation. In order to evaluate the worth of a preferred stock, its particular restrictions imposed on the corporation with respect to the payment of dividends must be considered.

Common stock is the stock which represents the members of the corporation who accept the primary risk and have the ultimate responsibility for management. Sometimes preferred stockholders do not have a voice in the company affairs unless the company fails to pay them dividends at the promised rate. Common stock shareholders have no guarantee on the returns on investment; but, in prosperous times when company earnings are high, dividends may also be high. The preferred stock is a more stable investment since the guarantee is stipulated. But, if the preferred stocks' claims have been met, the earnings may still provide a better return to the common shares.

Capital surplus is the amount paid by shareholders over the par or legal value of each share. Demand can cause the price of the stock to rise beyond the share prescribed by the stock certificate. This difference is entered as the capital surplus figure on the balance sheet.

Accumulated retained earnings or earned surplus consists of appropriated earnings retained by the corporation as reserves for contingencies, sinking funds, self-insurance or retirement of preferred stock. This value represents capital that is necessary for the corporation to have in order to fill these needs.

INCOME STATEMENT

The income statement is a summary picture of the firm's income and expenses over a period of time. It may cover any selected period of time, but, because of income tax requirements, virtually all firms prepare a statement covering operations during the calendar or fiscal year. Most businesses, in addition, also prepare monthly, quarterly or semi-annual income statements.

In essence, the income statement shows whether the firm earned a profit or suffered a net loss during the period covered. A sample income statement is shown in Plate 2. The main logic of the income statement is sales and other revenue minus cost of goods sold equals gross margin; gross margin minus expenses, losses and taxes equals net corporate income. The interest on bonds are subtracted from the net corporate income to obtain the net profit on stock equity; from the net profit on stock equity is subtracted the dividends to obtain the addition to retained earnings or surplus. The addition to surplus figure is added to the cumulative surplus of the previous year to obtain the updated accumulative retained earnings.

From gross sales the sales returns and allowances and cash discounts must be deducted to obtain the net sales value. Under the entry "Other Revenue" is listed as any other income acquired by the firm by investment in marketable securities or other incidentals. The sum of the net sales and all other revenue represents the net amount of revenue out of which the firm will pay for the products and all its expenses.

EXPLANATION OF PLATE II

A sample income statement

INCOME STATEMENT

Year ended December 31, 1966

SALES		\$XXXX	
Other Revenue		XXXX	\$XXXX
Cost of Goods Sold			<u>XXXX</u>
GROSS MARGIN			\$XXXX
EXPENSES			
Selling and Delivering	\$XXXX		
Administrative and General	XXXX	XXXX	
LOSSES		\$XXXX	
TAXES			
Property Taxes	\$XXXX		
Federal Income Tax	<u>XXXX</u>	<u>XXXX</u>	<u>XXXX</u>
NET CORPORATE INCOME			\$XXXX
INTEREST			<u>XXXX</u>
NET PROFIT ON STOCK EQUITY			\$XXXX
DIVIDENDS			<u>XXXX</u>
ADDITION TO SURPLUS			\$XXXX
SURPLUS January 1, 1966			<u>XXXX</u>
SURPLUS December 31, 1966			<u><u>\$XXXX</u></u>

Cost of goods sold is the next entry in the income statement. Cost of goods sold is the difference between cost of goods available for sale and the ending inventory. To obtain the cost of goods available for sale, the beginning inventory is added to the gross purchases during the period.

The expenses are listed in a manner that best suits the firm. The number of different types would be determined by the complexity of the firm's financial statement. Generally, expenses include selling and delivering and administrative expenses.

The losses should be described briefly in the statement with reference to some other document for clarification.

The tax entries should be subdivided into a grouping for levies other than federal income and profits, taxes such as property tax, city tax, etc., and a grouping for federal income and excess-profit taxes. Taxes should always be listed separately from the expenses regardless of how brief the income statement might be.

As mentioned in the discussion of the balance sheet, interest on bonds must be paid either annually or semi-annually. This expense has priority over any of the stock's claims.

Dividends are declared by the corporation with the amount of earnings in mind, but also, regarding the need of the corporation for retained surplus.

FINANCIAL RATIOS

From the financial reports discussed, information is obtained and ratios are formed which are useful in analyzing and evaluating the desirability of a firm as a prospective investment opportunity. These ratios are used as the independent variables in the linear multiple regression. The calculation of these ratios is discussed below.

NET WORKING CAPITAL

This figure is obtained from the balance sheet. It is sometimes also referred to as net current assets or simply working capital. This figure represents the amount of money that would be left free and clear if all current debts were paid. The ability of a company to meet its obligations, expand its volume and take advantage of opportunities is often determined by its working capital. This is the reason for the importance of this figure. Net working capital is the difference between the current assets and current liabilities, thus,

net working capital = current assets - current liabilities.

CURRENT RATIO

A rough test for an industrial company to gauge its ability to meet its obligations is with a current ratio. Current ratio is important in studying the undesirable condition such as over-investment in receivables or inventories.

The minimum safety requirement of this ratio differs for each industry. Generally, companies that have a small inventory and easily collectible accounts receivable can operate safely with a lower current ratio than those companies who have a greater proportion of their current assets in inventory and who sell their products on credit. The interpretation of current ratio in this report is that it is a measure of the company's current liquidity. Current ratio is the ratio of current assets to current liabilities, thus,

$$\text{current ratio} = \frac{\text{current assets}}{\text{current liabilities}} .$$

QUICK ASSETS

Another common method of testing the adequacy of the current position is the quick asset test. Quick assets are the current assets which can be quickly converted into cash. This would exclude merchandise inventories because such inventories have yet to be sold. Quick assets are the difference between current assets

and inventories, thus,

$$\text{quick assets} = \text{current assets} - \text{inventories}.$$

INVENTORY TURNOVER

The appropriate size of the inventory will depend on many factors. Consideration should be given to such factors as the type of industry, the time of year and the cost of product. A relatively slow turnover means an overinvestment in merchandise. Inventory turnover must be considered jointly with the net profit ratio since rapid turnover in itself does not assure profits. It is a measure of the number of times the merchandise is replaced during a given period and is computed as the ratio of sales to inventory, thus,

$$\text{inventory turnover} = \frac{\text{sales}}{\text{inventory}} .$$

BOND RATIO

The bond ratio is a measure of the desirability of a particular stock. Since the bond interest must be paid before dividends, and if the amount of bonds is excessive in relation to the other sources of funds, then the stock of the particular company is not too desirable. Bond ratio is the ratio of the face value of the

bonds to the sum of all the company's securities, thus,

$$\text{bond ratio} = \frac{\text{face value of the bonds}}{\text{bonds} + \text{preferred stock} + \text{common stock} + \text{capital surplus} + \text{accumulated retained earnings}}$$

LEVERAGE

Leverage is a measure of a company's proportion of bonds and preferred stock outstanding in relation to the amount of common stock. When a company is required to obtain some capital by trading on equity, it considers the leverage to determine the type of security to issue. If the interest due on the bonds would be higher than the amount of earnings per share of common stock, stock would then be issued. On the other hand, if the earnings per share would be more than the interest due on the bonds, the bonds will be sold. This is known as trading on equity with consideration on the leverage advantage. Leverage is expressed as the ratio of bonds and preferred stock outstanding to common stock, thus,

$$\text{leverage factor} = \frac{\text{bonds} + \text{preferred stock outstanding}}{\text{common stock}}$$

NET PROFIT RATIO

The net profit ratio is a measure taken from the income state-

ment. As mentioned previously, it is important also to use this statement since the balance sheet only shows the fundamental soundness of a company by reflecting its financial position at a given date, while the income statement shows the record of its operating activities for the whole year. The net profit ratio is compiled by taking the net sales plus all other forms of income less the costs incurred. The costs incurred usually consist of cost of the goods sold, overhead expenses such as wages and salaries, rent, supplies, depreciation, interest on money borrowed and taxes. The ratio of net profit to sales then results in the net profit ratio, thus,

$$\text{net profit ratio} = \frac{\text{net profit}}{\text{sales}} .$$

INTEREST COVERAGE

Another important factor to consider is the ratio of the income and the interest due on bonds. This would indicate the commitment the particular company has to its creditors. The immediate commitment for a corporation is the interest on bonds. The interest on bonds must be paid regardless of the bond value or when the bonds are due. The yearly interest is usually fixed and must be paid from the corporation's earnings. Interest coverage shows whether the total income justifies the borrowed funds. In effect, it is a measure of efficiency of the corporation's use of

its borrowed money. The total income figure used in the calculation of this ratio in the revenue figure before interest and income tax have been deducted. Interest coverage is the ratio of the total income to the interest on bonds, thus,

$$\text{interest coverage} = \frac{\text{total income}}{\text{interest on bonds}} .$$

EARNINGS PER COMMON SHARE

Earnings per common share is as important to the buyer of common stock as are the dividends. The reason being that the earnings per share is a measure of the ability of the company to pay dividends. This factor is important in the prediction of the stock price because earnings are a dominant factor affecting the pricing of common stocks. Earnings per common share is calculated by the ratio of the earnings available to the number of common shares outstanding. Earnings available is the net income for the year less dividend requirements on preferred stock, thus,

$$\text{earnings per common share} = \frac{\text{earnings available}}{\text{number of common shares}} .$$

PRICE EARNINGS RATIO

The price earnings ratio is a means to study the relative demand for a particular stock. This ratio points out the potential return that an investment in the stock can bring. The stock market price represents the amount necessary for investment and the earnings per share represents the return on the investment. The trend of this ratio will, in fact, influence the price of the stock. If investors can obtain a sizeable return on their investment, demand for the stock will cause the price of the stock to rise. Price earnings ratio is the ratio of the stock market price to the earnings per share, thus,

$$\text{price earnings ratio} = \frac{\text{stock market price}}{\text{earnings per share}} .$$

COMMON STOCK RATIO

The common stock ratio represents the percentage of securities issued by the corporation to common shareholders. If the common stock ratio is low, it indicates that a greater percentage of the securities are concentrated for bonds and preferred stock. If this situation exists, the common stock will not be attractive to investors since the bond interest and preferred stock dividends must be paid before common stock dividends.

The common stock ratio is the ratio of common stock, capital

surplus and accumulated earnings to total capitalization, thus,

$$\text{common stock ratio} = \frac{\text{common stock} + \text{capital surplus} + \text{accumulated earnings}}{\text{total capitalization}}$$

where total capitalization is the sum of the amount of bonds, preferred stock, common stock and capital surplus and the accumulated retained earnings.

NET BOOK VALUE OF A SHARE OF COMMON STOCK

The net book value of a share of stock represents the amount of money each share of common stock would receive if the company were liquidated according to balance sheet values after the preferential liquidation rights of bondholders and preferred stockholders are satisfied. This is measured by taking the ratio of the sum of the common securities less the intangible assets and the number of common shares.

This figure is not as significant as the coverage on the senior securities (bonds and preferred stock). In case of a liquidation the assets disposed of at a forced sale do not bring nearly the values as carried in the books.

The net book value of a share of common stock is the ratio of net book value of all common stock to the number of shares of common stock outstanding. The net book value of all common stock is

the sum of the common stock, capital surplus and accumulated retained earnings less the intangible assets, thus,

$$\text{net book value of a share of common stock} = \frac{\text{net book value of all common stock}}{\text{number of shares of common stock}} .$$

NET BOOK VALUE OF ALL COMMON STOCK

This value is very similar to net book value of a stock, but it is the total book value for all common stock collectively rather than for each stock. The value disregards the effect caused by the number of stock shares. It is a measure used to gauge the trend of the common securities growth. The ratio of net book value of all common stock is the ratio expressed as the intangible assets minus the sum of the common stock, capital surplus and accumulated retained earnings, thus,

$$\text{net book value of all common stock} = \frac{\text{common stock} + \text{capital surplus} + \text{accumulated retained earnings} - \text{intangible assets}}{\text{number of shares of common stock}} .$$

NET BOOK VALUE OF ALL BONDS

The net book value of bonds is used to predict the trend of the particular corporation's commitments. The net book value of bonds can be considered to be the ownership claims of investors in a corporation. The net book value of all bonds is the ratio

of the difference between intangible assets and the sum of the amount of bonds, common stock, capital surplus and accumulated retained earnings, thus,

net book value of all bonds = amount of bonds + common stock +
capital surplus + accumulated re-
tained earnings - intangible assets.

GROSS NATIONAL PRODUCT

This variable is a gauge of the entire economy in the United States. It is used to predict the trend of the economy as a whole. Generally, the future of any company parallels the general economic outlook, thus, it is included as one of the factors.

CONSTRUCTION OF THE MODEL

The forecasting model used to predict the stock prices is a multiple linear regression. Where the independent variables are the financial ratios and the dependent variable is the stock price. It must be noted that these variables are not independent in the statistical sense, but are independent in the functional sense. This model is based on the assumption that the stock price and the financial ratios are jointly distributed normal variables. The particular model used for forecasting the expected high and low stock prices is expressed mathematically as:

$$y_t = b_0 + b_1(X_1 - \bar{X}_1) + b_2(X_2 - \bar{X}_2) + b_3(X_3 - \bar{X}_3) \dots b_{n-1}(X_{n-1} - \bar{X}_{n-1}) \quad (1)$$

where b_0 is a constant and the remaining b_i 's are the regression coefficients, the \bar{X}_i 's are the means of the respective X_i 's which are the observed values of the independent variables.

The objective is to construct a model by selecting the regression coefficients so that the error of estimation is minimized. One approach to the estimation of the regression coefficients is by the method of least squares which result in a best fit regression equation. The least square estimates are obtained by minimizing the square of the error of estimation. The square of the error of estimation is expressed as the sum of the square of the deviations of the forecasted values from the observed values and is called the error function which is expressed as:

$$S = \|e\|^2 = \sum_{t=1}^m \left\{ (y_t - b_0) - \sum_{i=1}^{n-1} b_i (X_{it} - \bar{X}_i) \right\}^2 \quad (2)$$

where $n-1$ = number of independent variables
 m = number of sets of observations.

The coefficients of the least squares regression equation are obtained by minimizing S , equation (2), with respect to the b_i 's. This is done by finding the partial derivatives of (2) with respect to each of the coefficients and setting them equal to zero. The resulting set of simultaneous equations are known as the normal equations, the solution of which yield the regression coefficients.

The first normal equation resulting from finding the partial derivatives with respect to the b_0 is expressed as:

$$\frac{\partial S}{\partial b_0} = 2 \sum_{t=1}^m \left\{ (y_t - b_0) - \sum_{i=1}^{n-1} b_i x_{it} \right\} (-m) = 0 \quad (3)$$

where $x_{it} = X_{it} - \bar{X}_i$.

This equation may also be expressed as:

$$0 = \sum_{t=1}^m y_t - \sum_{t=1}^m b_0 - \sum_{i=1}^{n-1} b_i \left(\sum_{t=1}^m x_{it} \right). \quad (4)$$

From the fact that the sum of the difference between the observed values and the mean is zero, results in,

$$x_{it} = \sum_{t=1}^m (X_{it} - \bar{X}_i) = 0 \quad (5)$$

and as a consequence equation (4) reduces to

$$y_t = m b_0 \quad (6)$$

or

$$b_0 = \frac{y_t}{m} = \bar{y} . \quad (7)$$

Thus from equation (7) it is seen that b_0 is equal to the mean of the dependent variable which is the stock price.

The remaining normal equations are found by taking partial derivatives of equation (2) with respect to the other b_i 's and are as follows:

$$\frac{\partial S}{\partial b_i} = 2 \sum_{t=1}^m \left[\{ (y_t - b_0) - \sum_{j=1}^{n-1} b_j x_{jt} \} \{ x_{it} \} \right] = 0 \quad (8)$$

where $i=1, 2, \dots, n-1$.

The normal equation in (8) may be rewritten as the following set of $n-1$ simultaneous equations:

$$\begin{aligned} \sum_{t=1}^m x_1(y_t - b_0) &= \sum_{j=1}^{n-1} \sum_{t=1}^m \{ (x_{1t})(x_{jt}) \} b_j \\ \sum_{t=1}^m x_2(y_t - b_0) &= \sum_{j=1}^{n-1} \sum_{t=1}^m \{ (x_{2t})(x_{jt}) \} b_j \\ &\vdots \\ \sum_{t=1}^m x_{n-1}(y_t - b_0) &= \sum_{j=1}^{n-1} \sum_{t=1}^m \{ (x_{n-1,t})(x_{jt}) \} b_j \end{aligned} \quad (9)$$

These normal equations are, consequently, solved for the least square estimates of the regression coefficients in equation (1) for a given set of independent variables. The choice of these variables is the discussion of the next section.

SELECTION OF VARIABLES AND CALCULATION OF REGRESSION COEFFICIENTS

Based on the earlier discussion, the variables for the forecasting model are to be selected from the various financial ratios. In this model the number of variables must not exceed the number of sets of data. The reason for this becomes apparent in the computation of the F statistic for testing the statistical significance of the multiple regression model. The degrees of freedom for the F statistic is computed as:

$$D.F. = m - n \quad (10)$$

where m = sets of data
 n = number of dependent and independent variables,

A negative degrees of freedom has no statistical significance, thus, m must be greater than n . The next and most important concern is the selection of the ratios to be used in the model. The selection technique used here is a stepwise multiple regression approach. With this technique the variables are introduced one by one into the regression, and the variables which reduce the unexplained variation significantly are selected for the model.

This method for selecting the factors for a multiple regression model is based on the work of M. A. Efroymsen (8). This approach has the interesting feature that not only is the final solution obtained, but, a "partial regression equation" is obtained at each stage. This regression equation contains only the variables which have been selected up to that point. The insertion of variables into the regression is terminated when either the variables not in the regression do not significantly reduce the reduction of variance or the degrees of freedom have been reduced to 2. The level of significance in the reduction of variance is specified by the user. M. A. Efroymsen's procedure for a stepwise multiple regression analysis is discussed below.

The procedure is to solve the following partitioned matrix to determine which factors go into the regression equations as well as the value of the regression coefficients,

$$\begin{pmatrix} S & T' & I \\ T & Z & D \\ -I & B & C \end{pmatrix} \quad (12)$$

where S, C, I, are $(n-1) \times (n-1)$, T and D are $1 \times (n-1)$, B is $(n-1) \times 1$, and Z is a scalar. Specifically,

$$(S)_{ij} = s_{ij} = \sum (x_{it})(x_{jt}) \quad (13)$$

$$(T)_{1j} = t_{1j} = \sum (x_{jt})(y_t - \bar{y}) \quad (14)$$

$$Z = \sum (y_t - \bar{y})^2 \quad (15)$$

$$(T')_{il} = (T)_{li} \quad (16)$$

$$B = C = D = 0 \quad (\text{initially}) \quad (17)$$

$$(I)_{ij} = \delta_{ij} \quad (18)$$

where $(I)_{ij}$ is the identity matrix and $x_{it} = X_{it} - \bar{X}_{it}$.

The S matrix is the variance - covariance matrix of the independent variables, x_i and x_j for all i's and j's. The T' vector contains the covariance of the ith independent variable with the dependent variable. The scalar Z is the sum of the squares of the deviations from the mean for the dependent variable. The S matrix

corresponds to the coefficients of the b_i 's on the right hand side of the normal equations and the T matrix corresponds to the left hand side of the equations. The two identity matrices are attached so that the current inverse of the variance-covariance matrix will be maintained. The application of linear transformations to the matrix A(I,J) in entering or removing a variable will cause non-zero elements to enter the B, C and D positions.

The normal equations (9) which are represented by S are solved by the Gaussian reduction method in order to obtain the regression coefficients. With each successive row reduction of the S matrix a regression equation is obtained with one more variable in the regression equation. This technique enables the variables with the most significant reduction of variance to be inserted in a stepwise procedure.

To remove a variable from the regression, a row elimination on the C matrix is performed, and a regression with one less variable is obtained.

The B matrix will always contain the regression coefficients for the variables in the regression. The D matrix contains the negative transpose of the B matrix. The C matrix contains the inverse of the S matrix for all X_i 's and X_j 's which are in the regression equation at that step. With this procedure, coefficients are calculated for the terms which have been entered or removed from the regression equation at each stage.

This process of entering variables into the regression is terminated when the variance reduction with the addition of any of the remaining variables is not significant at the prescribed level.

A test is made before each variable is entered into the regression to maintain the degrees of freedom greater than one. If the addition of the variable decreased the degrees of freedom to two, the process is stopped. Two is the smallest allowable value for the degrees of freedom in order to have a valid statistical statement.

Using the variance terms

$$ns_i = \sum x_i^2 \quad (19)$$

and the equations of the simple correlation coefficients written as

$$n(s_i)(s_j) r_{ij} = \sum (x_i x_j) \quad (20)$$

the normal equations (9) are rewritten as

$$s_y r_{ij} = \sum_{j=1}^{n-1} (s_j r_{ij}) b_j \quad (21)$$

for $i = 1, 2, \dots, n-1$.

Using the general term

$$B_i = \frac{s_i}{s_y} b_i \quad (22)$$

and dividing the normal equations by s_y , equations (21) may be rewritten as

$$\begin{aligned}
 r_{1,y} &= B_1 r_{1,1} + B_2 r_{1,2} + \dots + B_{n-1} r_{1,n-1} \\
 r_{2,y} &= B_1 r_{2,1} + B_2 r_{2,2} + \dots + B_{n-1} r_{2,n-1} \\
 &\vdots \\
 &\vdots \\
 r_{n-1,y} &= B_1 r_{n-1,1} + B_2 r_{n-1,2} + \dots + B_{n-1} r_{n-1,n-1} .
 \end{aligned} \tag{23}$$

This set of $n-1$ simultaneous equations (23) are known as the standardized normal equations.

With the normal equations standardized the regression equation (1) is also normalized to the form

$$y_t = B_1 x_1 + B_2 x_2 + \dots + B_{n-1} x_{n-1} \tag{24}$$

where $B_i = b_i \frac{s_i}{s_y}$.

In order to obtain the regression coefficients, the constants B_i 's in the matrix B must be transformed with the relationship established in equation (22), thus,

$$b_i = B_i \left(\frac{s_y}{s_i} \right) . \tag{25}$$

In the computer program for the stepwise multiple regression the form used had the S , T , T' and Z matrices standardized. These four matrices together, after being standardized, form the r matrix, the matrix of the simple correlation coefficients, The standard-

ized partitioned matrix is now represented by

$$A = \begin{pmatrix} & & I \\ & r & D \\ -I & B & C \end{pmatrix} \quad (26)$$

The multiple regression analysis begins by calculating the simple correlation coefficients r_{ij} . These are then calculated by use of the equation

$$r_{ij} = \frac{\sum (x_{it})(x_{jt})}{\sqrt{\sum (x_{it})^2 \sum (x_{jt})^2}} \quad (27)$$

The A matrix is formed using the calculated correlation coefficients and the initialization of the auxiliary partitioned matrices B, C, D = 0 and where -I and I are the negative identity and identity matrix respectively.

The process to determine which variable is to go into the regression function begins. This procedure is one of finding the variable which reduces the unexplained variance the most significantly. The reduction of variance as a result of introducing this variable into the regression equation is

$$V_i = \frac{(r_{in})(r_{ni})}{r_{ii}} \quad (28)$$

r_{in} , r_{ni} , and r_{ii} are updated elements from the r partitioned matrix where the subscript i corresponds to the variable in question, and n corresponds to the dependent variable. Before the actual calculation is made, a check is performed on the magnitude of r_{ii} . The magnitude of r_{ii} introduces the possibility of dependents when, because of round-off error, a variable can be expressed as a linear combination of the other variables. To prevent this occurrence r_{ii} is compared to a specified tolerance. Commonly used tolerances are from .001 to .00001. If r_{ii} is less than the tolerance, this variable is neglected. If r_{ii} is larger than the tolerance, the variance reduction (V_i) is then calculated.

The criterion used to determine whether the reduction of variance is significant is an F test. The F statistic for a variable leaving is calculated to be

$$F_{\text{TEST}} = \frac{|V_{\min}| \text{ (D.F.)}}{r_{nn}} \quad (29)$$

where V_{\min} is the smallest reduction of variance caused by a variable in the regression and D.F. is the degrees of freedom at this stage. The F statistic for a variable entering is calculated to be

$$F_{\text{TEST}} = \frac{(V_{\max}) \text{ (D.F.)}}{r_{nn} - V_{\max}} \quad (30)$$

where V_{\max} is the largest reduction of variance caused by a vari-

able not in the regression and D.F. is the degrees of freedom at this stage.

The test is to compare the F statistic calculated with the appropriate F level specified. The user of the program must specify the F levels required to select or to remove a variable from the regression equation. If the F statistic is larger than the specified F level for the removal of a variable, the variable is removed from the regression and the matrix is transformed. If the removal criterion is not satisfied, the entering criterion is then tested. If the calculated F statistic is larger than the F level specified for entering a variable, the variable is entered, and the matrix is transformed. If the entering criterion is not satisfied, the regression equation is in its final form.

The F tabular values must be read into the computer by the user. These F (f_1, f_2) values have $f_1 = 1$, $f_2 = m - v$ degrees of freedom where m is the number of sets of data, and v is the number of variables in the regression including the dependent variable. Therefore, if a model utilized ten years of data, $F(1,8)$, $F(1,7)$, $F(1,6)$, $F(1,5)$, $F(1,4)$, $F(1,3)$ and $F(1,2)$ would be needed. The upper limit for f_2 is $m - 2$ because when the first variable is introduced, there are two variables in the regression, the dependent and the independent variable just introduced. The lower limit is always two because of the valid statistical inference; this, then, is the smallest degrees of freedom allowed by the program. The f_1 degrees of freedom is always 1 because of the test on the reduction of variables for the single variable in question. Therefore, the number of F tabular values necessary will be determined

by the sets of data to be used in the model.

The alpha level for the F tabular values may be designated by the user. The alpha level is the critical region or the probability of rejecting a variable as insignificant when it is, indeed, significant. It is suggested that a small alpha is used because the least number of variables in the regression equation will give the model the greatest possible degrees of freedom. With a larger number of degrees of freedom a better statistical inference can be obtained.

The criteria used to select the x_i variables to be added or removed from the regression are as follows:

(a) The x_i corresponding to the minimum V_i for negative V_i 's is removed from the regression if the variance contribution of x_i is insignificant at this step. If no variable is to be removed, then rule (b) is applied.

(b) The x_i corresponding to the maximum V_i for positive values of V_i is added to the regression if the variance reduction due to adding x_i is significant.

Applying either rule above, the a_{ii} corresponding to the variable selected is used as the key element. The Gaussian reduction is used to transform the matrix in order to obtain the regression coefficients. The key element which corresponds to the variable leaving or entering the regression equation is used as the pivot element. Its corresponding subscript is equated to "k". The Gaussian reduction can be generalized to be:

$$\begin{aligned} \text{for } i \neq k \text{ and } j \neq k \text{ then } a_{ij} \text{ (new)} &= a_{ij} - \frac{(a_{ik})(a_{kj})}{a_{kk}} \\ \text{for } i = k \text{ and } j \neq k \text{ then } a_{kj} \text{ (new)} &= \frac{a_{kj}}{a_{kk}} \\ \text{for } i \neq k \text{ and } j = k \text{ then } a_{ik} \text{ (new)} &= \frac{-a_{ik}}{a_{kk}} \\ \text{and for } i = j = k \text{ } a_{kk} \text{ (new)} &= \frac{1}{a_{kk}} . \end{aligned} \tag{31}$$

This general form of the Gaussian reduction applies to the entire matrix A; the element may have an r, b, d or C designation. The designation of elements are made by the following rules.

(a) $a_{ij} = r_{ij}$ when both x_i and x_j are not in the regression at this step. The dependent variable is never considered to be in the regression.

(b) $a_{ij} = B_{ij}$ when x_i is in the regression and x_j is not in the regression. B_{ij} is a standardized regression coefficient of x_i on x_j adjusted for any other variable that is in the regression.

(c) $a_{ij} = d_{ij}$ when x_i is not in the regression, and x_j is in the regression.

$$d_{ij} = -B_{ji}$$

(d) $a_{ij} = c_{ij}$ when both x_i and x_j are in the regression. The matrix with elements c_{ij} is the inverse of r_{ij} matrix for all i 's and j 's that are in the regression at this step.

With this algorithm the entire matrix is transformed each

time after a variable is to be removed or inserted into the regression equation. In the computer program after the matrix is transformed, the number of the variable which was removed or entered is printed with the appropriate regression coefficient, the F level, the standard error of y , the constant (b_0) and the standard error of the coefficient. Then the investigation is conducted again to find out whether there is another variable to be entered or one to be removed. This procedure continues until there is no longer a need to remove or to enter any variables.

Once the regression equation is in its final form, the prediction model may be formulated. The prediction model consists of the sum of the products of the variables in the regression at the final stage and their appropriate coefficients, updated at the final stage.

To predict the value of the stock price from the model, the final values for the variables in the regression are substituted into the model. The predicted value is calculated by adding to the constant, b_0 , the sum of the products of the coefficients and the final values for the variables in the regression.

In order to test the model, two values are calculated. The first value calculated is the multiple correlation coefficient $R_{y.k}$, where k is the set of variables in the regression equation. This value represents the linear relationship between the variables in the regression and the dependent variable. The multiple correlation coefficient is calculated by taking the square root of the sum of the products of the standardized constants and the simple correlation coefficients of the independent variables in the re-

gression equation and the dependent variable. Therefore, if the array of the variables in the regression is represented by "k", R may be represented by

$$\sum_k \sum_{i=1}^{IC} (B_{Ni,k}) (r_{N,i}) \quad (32)$$

where IC is the number of variables in the array and N is the subscript for the dependent variable. For development of this formula, see Hays (5).

The other measure for the model is an F statistic. This statistic is calculated to show the significance of the model. The F statistic is calculated using the multiple correlation coefficient R. It is obtained by

$$F = \left(\frac{R^2}{1-R^2} \right) \left(\frac{N-IC}{IC-1} \right) \quad (33)$$

For development of this statistic, see Hays (5).

OPERATION OF MODEL AND COMPUTER PROGRAM

In this model financial ratios are used as independent variables in a multiple linear regression to predict the stock price, the dependent variable. To calculate these ratios the necessary data is obtained from the balance sheet and income statement. A stepwise procedure is used to insert only the variables which significantly reduce the variance of estimate. When the model is in its final form, a prediction is made.

Models are formed for high, low and average stock price prediction. Using the predicted values for the high and low stock price and their respective standard errors of estimate, an investment logic can be formed. A discussion of the computer program to form this model is discussed below.

The computer program is written in FORTRAN II and takes approximately 54000 core locations. With this program, the values from the balance sheet and income statements are read and the appropriate ratios are calculated to use as variables in the stepwise multiple regression analysis.

The first variable which must be read in to the computer is the number of years of data to be used excluding the set of data used to predict. This variable is designated by the term NNN in the computer program, and a limit is set for it. The number of sets of observations must be within the range of not more than ten, but not less than five years. The reason for the lower limit imposed is that four or less years of data would not be sufficient to enable a significant inference statement. The higher limit is

based on the belief that data beyond ten years contributes little to the present market trend. The format for this variable is an I2. Therefore, the value must be punched right justified in columns 1 and 2.

The input variables which must be read in next are:

1. Current Assets
2. Current Liabilities
3. Inventory
4. Sales
5. Value of the Bonds
6. Preferred Stock Outstanding
7. Common Stock Outstanding
8. Capital Surplus
9. Accumulative Retained Earnings
10. Net Profits
11. Total Income
12. Interest on Bonds
13. Earnings per Share
14. Shares of Common Stock
15. Gross National Product
16. Intangible Assets
17. Stock Price.

These variables must be in this prescribed order. Two cards must be used to read all the data for each of the variables, except in the case when only five years of data are used. For each

variable five values should be punched on the first card and the rest on the second card. The format specified for these variables is an F 10.5. This format provides for a five decimal point accuracy. The values should begin with the earliest year used, followed by consecutive years to the present year. If some of these variables do not exist for the model in question, blank cards must be read in the position for the missing variable. Therefore, there must always be 34 data cards of this type read in to the computer. As mentioned previously, the only exception would be when only five years of data are used, and then only 17 cards would be necessary.

It is suggested that an auto-multiple regression be used for the data. An auto-multiple regression consists of leading the stock price one year to the rest of the data. In other words, if the first year used was 1955, then the corresponding first value of the stock price would be the price in 1956. This would enable the values for the variables of the last year to predict the stock price in the year(s) to come.

Another method of obtaining the appropriate values to put in the predicting model is to have a straight multiple regression, i.e., stock price value from the same year as the variable values, and then by either speculation or some other method, predict the values for the variables in the regression equations. These would then be used in the model to obtain the predicted stock price. A simple correlation may be used to predict the variable values for the coming year.

The first method of the auto-multiple regression is the one

recommended and used in this report. If prediction is to be performed on the computer with this program, the last year's data must be read in so that the ratios may be calculated. All the ratios are calculated, but only the ratios corresponding to the variables in the regression are used in the prediction. It must be pointed out that in the set of data for the last year, since the stock price is leading by one year, it is obviously not known. Therefore, it is only necessary to read in the first sixteen variables. Since one of the ratios utilizes the stock price in its calculation (Price Earnings Ratio), the last value of the stock price will be used in this calculation. Since the program will utilize this last value, it is, therefore, unnecessary to read any value for this variable.

It is important in this technique that the data must be scaled. The data which is read in must be in the range of 100. to .1. This is to insure that good results are obtained. If numbers scaled larger than 100. to .1, were used, numerical round-off error could possibly occur which would affect the solution of the matrix by the Gaussian reduction.

LOGIC OF THE COMPUTER PROGRAM

A brief discussion will be given to explain the steps in the program so that a better understanding might be obtained. The first important function performed is to read the input variables (values from the balance sheet and income statement). The ratios

are then calculated from this data in order to form the regression variables (financial ratios already discussed).

The next function is to test SENSE SWITCH 1 to see if prediction will be made. If SWITCH 1 is on, the last year's data will be read and the ratios calculated and retained. If SWITCH 1 is off, reading the last year's data is skipped and the next function is then performed.

The next calculations performed are the initialization of subscripted variables and the calculation of sums of both the variables and their squares, the means of the variables, sums of the cross product of the variables and standard deviations for each variable. These values are necessary for the calculation of the simple correlation coefficients.

Only the upper triangular section of the correlation matrix is calculated. This is the only portion necessary since the matrix is symmetrical with all the diagonal elements equal to 1 after the matrix is normalized. SENSE SWITCH 2 will control the punching of these simple correlation coefficients. If SWITCH 2 is on, the upper triangular matrix of correlation coefficients will be punched. If SWITCH 2 is off, this punching will be suppressed. If the punching was desired, each correlation coefficient will be punched in a card with its appropriate subscripts. The SENSE SWITCH is tested after each correlation coefficient is punched, therefore, if only the first portion is desired, the SWITCH may be turned off and the remaining correlation coefficient will not be punched.

After these calculations have been made, the second part of the program, the stepwise multiple regression, is begun. The first

function is to set up the matrix $A(I,J)$. This matrix is the $(2n-1)$ by $(2n-1)$ partitioned matrix discussed in the first portion of this report. First the B, C and D matrices are initialized to zero. The I and -I matrices are then set up with 1's and -1's respectively in their diagonals. The r matrix is then completed. The diagonal of the r matrix contains the correlation coefficient of each of the variables with themselves, i.e., r_{ii} , therefore, these correlation coefficients are always 1. The simple correlation coefficients of variables with the dependent variable are retained. This is necessary since these values are needed to calculate the multiple correlation coefficient (method described previously in this report, page (41)).

With the matrix in its initial form the calculations are begun to find the best variable to insert into the regression. The variable which will reduce the variance the most is then selected. The reduction of variance is tested to see if it is significant. If it is significant, the variable is inserted into the regression. If the first variable to be entered with the largest variance reduction is not significant, then the model cannot be formulated with the F level specified. It is an obvious fact that a model must have at least one variable in order to predict the dependent variable. Therefore, if the largest variance reduction is found to be significant, a simple correlation model is formed. The multiple correlation coefficient, R, and the F statistic do not apply for a simple correlation model. Therefore, if there is only one variable in the model only the simple correlation coefficient is calculated.

After each variable is introduced into the regression, the matrix $(A(I,J))$ must be revised. The method utilized (mentioned previously in this report) is the Gaussian reduction. Each time a variable is inserted into the regression, the variable number, the updated standard error of estimate, the coefficient, the F level of its reduction of variance and the updated constant are printed.

The matrix $(A(I,J))$ must be revised if a variable is to be removed from the regression model. When the variable is removed, the coefficient for this variable is zero. When a variable is removed from the regression model, the variable number, the updated standard error of estimate, the updated constant and the F level of its reduction of variance are printed. The coefficient is not printed because it will always be zero for a variable leaving the model.

After a variable has been either entered or removed, calculations are performed to investigate the possibility of significance of the reduction of variance of another variable. Therefore, the reductions of variance are calculated with the now updated matrix. These variance reductions are sorted for the maximum reduction caused by a variable not in the regression and the minimum reduction caused by a variable in the regression. It is important to note that the removal of a variable is tested before testing the variance reduction for entering a variable. This is based on the premise that the least amount of variables give greater number of degrees of freedom and therefore, of greater statistical significance. There is only one variable in transition at each

step and the priority is therefore, given to a removal.

This process of search is continued until there are no more variables which can either be removed or entered. There is one exception to the exit from this search, an exit is executed if the degrees of freedom have been reduced to 2. It is believed that this is the lowest value which should be allowed in the number of degrees of freedom. An exit from the search indicates that the model has been completed.

After the model is complete, the variable numbers, the coefficients and the standard error of the coefficients of the variables in the regression are printed. The method utilized in determining the variables in the regression is to test the B vector in the segmented matrix. If the coefficient for the variable is zero, the variable is not in the regression. If the element is nonzero, the corresponding variable is in the regression.

Having the final model constructed, a predicted value is calculated if SWITCH 1 is on. If there are two or more variables in the regression model, the F statistic and the R-multiple correlation coefficient are calculated for the model and printed out. If there is only one variable in the model, only the simple correlation coefficient will be printed.

With these values the program has terminated the calculation of the model. If another model is to be loaded into the computer with simply depressing the start button, a branch will be executed to the beginning of the program. The input variable data may then be read in for the calculations for another model.

OPERATING PROCEDURES

A brief explanation will be given for the users of this program that are not familiar with the operating procedures of the IBM 1620 computer.

- I. Clear Core Storage
 - A. Depress Instant Stop Key
 - B. Depress Reset Key
 - C. Depress Insert Key
 - D. Type the instruction 160001000000
 - E. Depress Release Key and then the Start Key (or depress the RS Key on the upper left side of the typewriter's keyboard)
 - F. After approximately four seconds core should be cleared to zeros.
 - G. Depress Reset Key

- II. Prepare Punch (If it is to be used to punch the correlation coefficients)
 - A. Pick up cards in the punch hopper
 - B. Depress the Non-Process Run OUT button on the 1622
 - C. Place some blank cards in the punch hopper
 - D. Depress the Punch Start button on the 1622

- III. Set SENSE Switches
 - A. Set SENSE Switch 1 ON if prediction is to be made
 - B. Set SENSE Switch 2 ON if simple correlation coefficients are to be punched; OFF if this punching is to be suppressed

IV. Load Program

- A. Place the program (object deck) in the Reader hopper
- B. Depress the yellow Load button
- C. To read the last two cards, depress the Reader START button

V. Load Subroutines

- A. The message LOAD SUBROUTINES will be typed
- B. If the subroutines are with the program and already in the Reader hopper, simply depress the START key on the 1620
- C. If the subroutines are not with the program, place the subroutines in the Reader hopper. Depress the START key on the 1620 followed by depressing the Reader START button on the 1622
- D. To read in the last two cards depress the Reader START button on the 1622

VI. Load Data

- A. Place the data in the Reader hopper
 1. Control card with number of years of data
 2. Data for Input variables
 3. If prediction is to be made, place prediction data, one variable per card, after input data
 4. F levels to enter and remove
- B. Depress the START Key on 1620
- C. Depress the Reader START button on 1622
- D. To read in the last two cards, depress the Reader START button on the 1622

SAMPLE PROBLEM

A prediction model has been formed with actual data from Ford Motor Company. This financial information was obtained from the Standard & Poor's Corporation Records for 1965 (10) and Moody's Industrial Manual for 1965 (7). Information for any major corporation can be found in either of these two references.

The input data used is listed on pages 67, 68, 69 in this report. The code numbers on the right margin correspond to the numbers given for the input variables on page 43 of this report.

The computer output for this input data is listed on page 70 of this report. These results will be discussed in the following paragraphs.

In order that the variable numbers can be identified a list of the variable numbers and the corresponding variable names will be given.

VARIABLE NUMBER	VARIABLE NAME
1	Net Working Capital
2	Current Ratio
3	Quick Assets
4	Inventory Turnover
5	Bond Ratio
6	Leverage
7	Net Profit Ratio
8	Interest Coverage

Sample Problem (continued)

9	Earnings Per Common Share
10	Gross National Product
11	Price Earnings Ratio
12	Common Stock Ratio
13	Net Book Value of Stock per Share (Common)
14	Net Book Value of Stock
15	Net Book Value of All Bonds
16	Stock Market Price

Upon testing these variables in the program the resulting statistical model for the low stock price was:

$$y_t = -42.55859 + 1.51753(X_{10}) .$$

Refer to page 70 in this report.

The variable X_{10} corresponds to the gross national product. With the 1964 value for this variable inserted into this regression model the predicted low stock price value for 1965 is 51.87736. The standard error of estimate S_y is 1.3918. With this information the low limit for this stock price is calculated to be:

$$51.87736 - 2(1.3918) = 49.09376.$$

The model for the average stock price is expressed as:

$$y_t = 17.63129 + 2.08561(X_3).$$

The X_3 in this model represents the quick assets variable. With the 1964 value for the quick assets inserted into the prediction model the predicted average stock price for 1965 is 53.34393. The standard error of estimate for this model is 1.5448.

The prediction model for the high stock price value is expressed as:

$$y_t = 21.95971 + 2.43349(X_3) .$$

As in the average stock price model the variable in the model is X_3 , the quick assets. With the 1964 value for the quick assets inserted into the prediction model the predicted high stock price for 1965 is 63.62910. The standard error of estimate for this model is 2.1691. The high limit for this stock price is calculated to be:

$$63.62910 + 2(2.1691) = 67.9573 .$$

Since the models were set to predict the 1965 stock price values, the predicted values may be compared to the actual values in 1965.

	HIGH	AVERAGE	LOW
Actual Values	62.9	56.5	50.1
Predicted Values	63.63	53.34	51.88

It appears that the model constructed fits the market situation fairly well.

SUMMARY AND DISCUSSION OF MODEL

In the model presented in this report financial ratios were used as independent variables in a multiple linear regression to predict the stock price, the dependent variable. The ratios are calculated by the computer after the necessary information from the financial statements has been read in. The variables are entered into the regression equation by a stepwise procedure entering only the variables which significantly reduce the variance of estimate. When the model is in its final form, a prediction will be made.

In order to form an investment logic models were formed for high, low and average predicted stock prices. The standard errors of estimate were used to calculate the limits which signal the action to be taken in the market. Having the limits calculated, the market will then signal the timing for the moves.

CONCLUSION

This method has several advantages over most prediction methods. In this method each model is specifically for the individual situation of prediction. Each prediction for a different company or for the high, low or average stock prices is considered as a separate model in order to obtain the variables which are significant. Another advantage is that a statistical investment logic may be formed from the information obtained. With this logic the timing

for buying or selling will be signaled by the market. This is a desirable characteristic that many prediction models do not have.

This method should always be supplemented by the investor's judgment and common sense. The lack of this consideration may be the greatest pitfall of this model.

ACKNOWLEDGEMENTS

My sincere thanks go to my major professor, Dr. Frank A. Tillman, Department of Industrial Engineering, for his instruction during the preparation of this report and for his time and energy he devoted for counsel.

REFERENCES

- (1) Bennett, C. A. and Franklin, N. L. "Statistical Analysis in Chemistry and the Chemical Industry." New York: John Wiley & Sons, Inc., 1954.
- (2) Freund, J. E. "Mathematical Statistics." Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1962.
- (3) Fryer, H. C. "Concepts and Methods of Experimental Statistics." Boston: Allyn & Bacon, 1966
- (4) Hald, A. "Statistical Theory with Engineering Applications." New York: John Wiley & Sons, Inc., 1952.
- (5) Hays, William L. "Statistics for Psychologists." New York: Holt, Rinehart and Winston, Inc., 1963.
- (6) "How to Read A Financial Report." New York: Merrill Lynch, Pierce, Fenner and Smith, Inc., 1962.
- (7) "Moody's Industrial Manual." New York: Moody's Investor Services, Inc., 1965.
- (8) Ralston, A. and Wilf, H. S. "Mathematical Methods for Digital Computers." New York: John Wiley & Sons, Inc., 1960, pp. 190-202.
- (9) "Selection of Variables for Fitting Equations to Data." Technometrics, Richmond, Vol. 8, No. 1, February, 1966.
- (10) Standard & Poor's -- Corporation Records. New York: 1965.
- (11) Stanton, W. J. "Fundamentals of Marketing." New York: McGraw-Hill Book Company, 1964.
- (12) Snedecor, G. W. "Statistical Methods." Ames, Iowa: The Iowa State University Press, 1956.
- (13) Spiegel, M. R. "Statistics." New York: Schaum Publishing Co., 1961.
- (14) Wyman, D. G. "Stepwise Multiple Linear Regression Analysis for the IBM 1620 (CARD)." IBM 1620 Program Library Abstract, 6.0.007.

APPENDIX

COMPUTER PROGRAM, DATA AND OUTPUT


```

**      MODEL TO PREDICT STOCK PRICES WITH STEPWISE
C      MULTIPLE LINEAR REGRESSION BY JORGE GONZALEZ
      DIMENSION A(31,31),X(15,16),V(18,16),SB(19),VA(19),B(19)
      DIMENSION XBAR(18),SIG(25),SAC(16),FI(8)
95:    LJC=0
      LBJ=0
      DO 252 I=1,16
252    SAC(I)=0.
C
C      INPUT VARIABLES ARE READ
C
      READ 122,NNN
122    FORMAT(I2)
      DO 101 I=1,17
      READ 2,(V(I,J),J=1,5)
      IF(NNN-6)101,111,111
111    READ 2,(V(I,J),J=6,NNN)
101    CONTINUE
      2 FORMAT(5F10.5)
C
C      CALCULATION OF RATIOS FROM INPUT VARIABLES
C
      DO 103 J=1,NNN
241    X(J,1)=V(1,J)-V(2,J)
      IF(V(2,J))30,31,30
31     X(J,2)=0.
      GO TO 32
30     X(J,2)=V(1,J)/V(2,J)
32     X(J,3)=V(1,J)-V(3,J)
      IF(V(3,J))33,34,33
34     X(J,4)=0.
      GO TO 35
33     X(J,4)=V(4,J)/V(3,J)
35     X(J,5)=V(5,J)/(V(5,J)+V(6,J)+V(7,J)+V(8,J)+V(9,J))
      IF(V(7,J))36,37,36
37     X(J,6)=0.
      GO TO 38
36     X(J,6)=(V(5,J)+V(6,J))/V(7,J)
38     IF(V(4,J))39,40,39
40     X(J,7)=0.
      GO TO 41
39     X(J,7)=V(10,J)/V(4,J)
41     IF(V(12,J))42,43,42
43     X(J,8)=0.
      GO TO 44
42     X(J,8)=V(11,J)/V(12,J)
144    X(J,9)=V(13,J)
147    X(J,10)=V(15,J)
43     IF(V(13,J))49,50,49
50     X(J,11)=0.
      GO TO 51
49     X(J,11)=V(17,J)/V(13,J)
51     CARD=V(5,J)+V(6,J)+V(7,J)+V(8,J)+V(9,J)
      X(J,12)=(V(7,J)+V(8,J)+V(9,J))/CARD

```

```

52 X(J,14)=(V(7,J)+V(8,J)+V(9,J))-V(16,J)
   IF(V(14,J))55,56,55
56 X(J,13)=0.
   GO TO 57
55 X(J,13)=X(J,14)/V(14,J)
57 X(J,15)=V(5,J)+X(J,14)
103 X(J,16)=V(17,J)
    IF(SENSE SWITCH 1)244,245
C
C   READ IN VARIABLES FOR PREDICTION
C
244 IF(LJC)105,105,106
105 LJC=1
    J=NNN+1
    DO 108 I=1,16
108 READ 240,V(I,J)
240 FORMAT(F10.5)
    V(17,J)=V(17,NNN)
C
C   GO TO CALCULATION OF PREDICTION RATIOS
C
GO TO 241
C
C   RETAIN PREDICTION RATIOS
C
106 J=NNN+1
    DO 242 I=1,16
242 SAC(I)=X(J,I)
C
C   INITIALIZATION
C
245 DO 4 J=1,16
    B(J)=0.
    SB(J)=0.
    VA(J)=0.
    SIG(J)=0.
    4 XBAR(J)=0.
    AY=NNN
C
C   CALCULATE SUMS OF X AND SUMS OF X SQUARE
C
DO15 J=1,16
DO15 I=1,NNN
B(J)=B(J)+X(I,J)
15 SB(J)=SB(J)+X(I,J)*X(I,J)
C
C   CALCULATE MEANS
C
DO113 J=1,16
113 XBAR(J)=B(J)/AY
M=1
C
C   INITIALIZATION FOR SUM OF CROSS PRODUCTS
C

```

```

      DC26 I=1,15
      M=M+1
      DC26 J=M,16
26  A(I,J)=0.
      M=1
C
C   CALCULATE SUM OF CROSS PRODUCTS
C
      DC 7 I=1,15
      M=M+1
      DC 7 J=M,16
      DC 7 K=1,NNN
      7  A(I,J)=A(I,J)+X(K,I)*X(K,J)
C
C   CALCULATION OF THE STANDARD DEVIATIONS
C
      DC 8 J=1,16
      DC 8 I=1,NNN
      8  VA(J)=VA(J)+(X(I,J)-XBAR(J))**2
      DC 90J=1,16
      90 SIG(J)=SQRT(VA(J)/AY)
C
C   CALCULATION OF SIMPLE CORRELATION COEFFICIENTS
C
      M=1
      DC 102I=1,15
      M=M+1
      DC 102J=M,16
      V(I,J)=AY*A(I,J)-B(I)*B(J)
      X(I,J)=SQRT((AY*SB(I)-B(I)**2)*(AY*SB(J)-B(J)**2))
      IF(X(I,J))701,702,701
      702 A(I,J)=0.
      GO TO 10
      701 A(I,J)=V(I,J)/X(I,J)
      10 IF(SENSE SWITCH 2)104,102
      104 PUNCH 16,A(I,J),I,J
      16  FORMAT(F18.5,3X13,3X13)
      102 CONTINUE
C
C   F VALUES TO ENTER AND REMOVE ARE READ
C
      NCM=NNN-2
      DC 732 I=2,NCM
      732 READ 29,FI(I)
      29  FORMAT(F7.3)
C
C   NO. OF VARIABLES
C
      N=16
C
C   INITIALIZATION
C
      DC 66 I=1,N
      VA(I)=0.

```

```

75 SB(I)=0.
65 B(I)=0
C
C   SIZE OF MATRIX = L
C
L=(2*N)-1
NE=N+1
NA=N-1
C
C   SET UP MATRIX (INITIALIZE B, C,D)
C
DC 47I=N,L
DC 47 J=N,L
47 A(I,J)=0.
C
C   INITIALIZE -I
C
DC 143 I=NE,L
DC143 J=1,NA
143 A(I,J)=0.
I=N
DC 44 J=1,NA
I=I+1
44 A(I,J)=-1.
C
C   INITIALIZE I
C
DC145 I=1,NA
DC145 J=NE ,L
145 A(I,J)=0.
J=N
DC146 I=1,NA
J=J+1
146 A(I,J)=1.
C
C   SET UP DIAGONAL ELEMENT OF R MATRIX
C
DC 65 I=1,N
65 A(I,I)=1.
C
C   SET UP BOTTOM PORTION OF R MATRIX
C
MM=1
DC142 I=1,NA
MM=MM+1
DC142 J=MM,N
142 A(J,I)=A(I,J)
C
C   RETAIN THE R WITH THE DEPENDENT VARIABLE
C
DC 221 I=1,NA
221 X(1,I)=A(I,N)
JJ=1
DF=NNN-1

```

```

C
C   STEPWISE MULTIPLE REGRESSION
C
C   INITIALIZATION
C
107 VMIN=9999.99
    VMAX=0.
    NMIN=0
    NMAX=0
    F=FI(NCM)
C
C   STANDARD ERROR OF ESTIMATE = SY
C
    SY=SIG(N)*(SQRT(A(N,N)/DF))
    DC 3 I=1,NA
C
C   TEST FOR INTERDEPENDENCE
C
    IF(A(I,I)-.0001)3,3,17
C
C   CALCULATE REDUCTION OF VARIANCE
C
17  VA(I)=(A(I,N)*A(N,I))/A(I,I)
    IF(VA(I))1,3,9
C
C   FIND MAX REDUCTION OF VARIANCE
C
    9  IF(VA(I)-VMAX)3,3,222
222 VMAX=VA(I)
    NMAX=I
    B(I)=0.
    GO TO 3
    1  II=I+N
C
C   STANDARD ERROR OF COEFFICIENT = SB(I)
C
    SB(I)=(SY/SIG(I))*SQRT(A(II,II))
C
C   FIND MIN REDUCTION OF VARIANCE
C
    IF(ABS(VA(I))-ABS(VMIN))13,3,3
13  VMIN=VA(I)
    NMIN=I
    3  CONTINUE
C
C   TEST TO SEE IF VARIABLE IS TO BE REMOVED
C
    FTE=(ABS(VMIN)*DF)/A(N,N)
    IF(FTE-F)18,6,6
18  K=NMIN
    DF=DF+1.
    NCM=NCM+1
    PRINT 888

```

```

888 FORMAT(1H )
PRINT150,K
150 FORMAT(16HVARIABLE LEAVING,3X13)
PRINT151,FTE
151 FORMAT(7HF LEVEL,5XF9.4)
JJ=-1
GO TO 5

C
C TEST TO SEE IF VARIABLE IS TO BE ENTERED
C
6 FTE=VMAX*(DF-1.)/(A(N,N)-VMAX)
IF(FTE-F)22,22,20
20 K=NMAX
JJ=1
DF=DF-1.
NCM=NCM-1

C
C PRINT INFORMATION OF VARIABLE ENTERING
C
PRINT 888
PRINT 153,K
153 FORMAT(17HVARIABLE ENTERING,3X13)
PRINT154,FTE
154 FORMAT(7HF LEVEL,5XF9.4)

C
C GAUSSIAN REDUCTION
C
5 DO 23 I=1,L
DO 23 J=1,L
IF(I-K)59,23,59
59 IF(J-K)60,23,60
60 A(I,J)=(A(K,K)*A(I,J)-A(I,K)*A(K,J))/A(K,K)
23 CONTINUE
DO 85 I=1,L
IF(I-K)24,85,24
24 A(I,K)=-A(I,K)/A(K,K)
35 CONTINUE
DO 76 J=1,L
IF(J-K)25,76,25
25 A(K,J)=A(K,J)/A(K,K)
76 CONTINUE
A(K,K)=1./A(K,K)

C
C CALCULATE STANDARD ERROR OF ESTIMATE, COEFFICIENT
C AND STANDARD ERROR
C
II=K+N
SY=SIG(N)*SQRT(A(N,N)/DF)
B(K)=A(II,N)*SIG(N)/SIG(K)
SB(K)=(SY/SIG(K))*SQRT(A(II,II))

```

```

C
C      UPDACE COEFFICIENTS
C
      DC 86 I=1,NA
      II=I+N
      IF(A(II,N))77,86,77
77 B(I)=A(II,N)*(SIG(N)/SIG(I))
86 CONTINUE

C
C      CALCULATE UPDATED CONSTANT
C
      BC=XBAR(N)
      DC 12 I=1,NA
12 BC=BC-B(I)*XBAR(I)

C
C      PRINT VARIABLE INFORMATION
C
      PRINT155,SY
155 FORMAT(19HSTANDARD ERROR OF Y,F8.4)
      PRINT156,BC
156 FORMAT(8HCONSTANT,F12.5)
      IF(JJ)99,99,98
      98 PRINT157,B(K)
157 FORMAT(11HCoefficient,F12.5//)
      IF(DF-2.)22,22,99
      99 GO TO 107

C
C      PRINT INFORMATION FOR FINAL MODEL
C
      22 PRINT 61
      61 FORMAT(3HVAR,10X5HCoeff,10X12HSTD ERR COEF)
      DC 97 I=1,NA
      97 VA(I)=0.
      IC=0
      R=0.

C
C      CALCULATE R SQUARE AND RETAIN COEFFICIENTS FOR PREDICTION
C
      DC 62 I=NE ,L
      IF(A(I,N))63,62,63
63 II=I-N
      R=R+A(I,N)*X(1,II)
      PRINT 64,II,B(II),SB(II)
64 FORMAT(1HX,I3,5XF10.4,5XF9.5)
      IC=IC+1
      VA(IC)=B(II)
      B(IC)=SAC(II)
62 CONTINUE
      IF(IC-1)263,260,261
263 PRINT 270
270 FORMAT(38HINVALID DATA NO VARIABLE IN REGRESSION)
      GO TO 277
260 PRINT 272
272 FORMAT(33HONLY 1 VARIABLE IN THE REGRESSION)

```

```

PRINL 273
273 FORMAT(37HTHEREFORE NO R OR F CAN BE CALCULATED)
PRINT 274,X(1,K)
274 FORMAT(48HFOR THIS MODEL SIMPLE CORRELATION COEFFICIENT = F8.6)
LBJ=1
261 IF(LJC)250,250,251
C
C   CALCULATE AND PRINT PREDICTED VALUE
C
251 PRED=BC
DO 93 I=1,IC
93 PRED=PRED+B(I)*VA(I)
PRINT 888
PRINT 91,PRED
91 FORMAT(18HPREDICTED VALUE = ,F12.5)
C
C   CALCULATE AND PRINT F STATISTIC FOR MODEL
C
250 IF(LBJ)255,255,277
255 VK=IC
D1=VK-1.
D2=NNN-IC
F=(R/(1.-R))*(D2/D1)
C
C   CALCULATE AND PRINT R FOR MODEL
C
R=SQRT(R)
PRINT 888
PRINT 220,R
220 FORMAT(35HMULTIPLE CORRELATION COEFFICIENT = F11.8)
PRINT 888
PRINT 219,F
219 FORMAT(14HF FOR MODEL = F10.4)
PRINT 218,D1,D2
218 FORMAT(5HWITH F4.0,4HAND F4.0,18HDEGREES OF FREEDOM)
PRINT 888
277 PRINT 217
217 FORMAT(35HTO LOAD ANOTHER MODEL DEPRESS START)
PAUSE
PRINT 888
GO TO 952
END

```

 *
 INPUT DATA FOR LOW STOCK PRICES
 *

			NO. OF SETS	OF OBSERVATIONS	
9					
9.67177	10.91328	11.02571	13.57686	11.98247	1A
21.24328	24.90469	26.69125	29.90519		1B
6.83104	6.56233	5.32084	5.28485	5.60554	2A
12.45216	15.42614	15.61675	17.92321		2B
5.97494	6.53595	4.74300	5.20413	5.27743	3A
8.90804	9.89184	10.88025	11.93669		3B
66.096	56.555	68.393	52.809	66.488	4A
67.976	67.094	80.895	87.425		4B
.58615	2.5000	2.500	2.49460	2.39435	5A
2.58429	2.71430	2.33132	2.08647		5B
.	6A
.	6B
2.69968	2.72427	2.73810	2.53516	2.74583	7A
2.74583	2.75089	2.75780	2.76536		7B
2.72089	2.79639	2.84059	2.87536	2.88819	8A
2.88819	2.97253	3.04738	3.13652		8B
2.795	1.186	1.633	.068	2.979	9A
2.633	2.447	2.825	2.898		9B
4.542	2.482	2.944	1.162	4.514	10A
4.279	4.096	4.087	4.885		10B
7.	7.139	8.604	5.005	1.1884	11A
11.087	10.652	12.636	12.867		11B
.313	.24	.323	.415	.533	12A
.420	.235	.214	.179		12B
4.25	2.30	2.70	1.06	4.12	13A
3.90	3.72	4.36	4.42		13B
.53462	.54014	.54486	.54762	.54862	14A
.54917	.55018	1.10312	1.10614		14B
39.74	41.92	44.28	44.44	48.27	15A
50.26	51.87	55.62	58.39		15B
16.86066	18.42297	16.73313	15.36259	15.15720	16A
19.99293	20.97044	21.40163	22.80317		16B
25.75	17.88	18.63	25.38	30.25	17A
31.38	36.13	42.00	48.88		17B

 *
 DATA FOR PREDICTION
 *

31.56058					1P
19.82335					2P
14.43730					3P
96.708					4P
2.27496					5P
.					6P
2.77223					7P
3.21801					8P
2.843					9P
5.056					10P
12.546					11P
.134					12P
4.56					13P
1.10890					14P
62.23					15P
25.8756					16P

 *
 F LEVELS
 *

198.					2 F	LEV.
55.55					3 F	LEV.
31.33					4 F	LEV.
22.78					5 F	LEV.
18.64					6 F	LEV.
16.24					7 F	LEV.

 *
 INPUT DATA FOR AVERAGE STOCK PRICES
 *

			NO. OF SETS	OF OBSERVATIONS	
9					
9.67177	10.91328	11.02571	13.57686	11.98247	1A
21.24328	24.90469	26.69125	29.90519		1B
6.83104	6.56233	5.32084	5.28485	5.60554	2A
12.45216	15.42614	15.61675	17.92321		2B
5.97494	6.53595	4.74300	5.20413	5.27743	3A
8.90804	9.89184	10.88025	11.93669		3B
66.096	56.555	68.393	52.809	66.488	4A
67.976	67.094	80.895	87.425		4B
.58615	2.5000	2.500	2.49460	2.39435	5A
2.58429	2.71430	2.33132	2.08647		5B
.	6A
.	6B
2.69968	2.72427	2.73810	2.53516	2.74583	7A
2.74583	2.75089	2.75780	2.76536		7B
2.72089	2.79639	2.84059	2.87536	2.88819	8A
2.88819	2.97253	3.04738	3.13652		8B
2.795	1.186	1.633	.068	2.979	9A
2.633	2.447	2.825	2.898		9B
4.542	2.482	2.944	1.162	4.514	10A
4.279	4.096	4.087	4.885		10B
7.	7.139	8.604	5.005	1.1884	11A
11.087	10.652	12.636	12.867		11B
.313	.24	.323	.415	.533	12A
.420	.235	.214	.179		12B
4.25	2.30	2.70	1.06	4.12	13A
3.90	3.72	4.36	4.42		13B
.53462	.54014	.54486	.54762	.54862	14A
.54917	.55018	1.10312	1.10614		14B
39.74	41.92	44.28	44.44	48.27	15A
50.26	51.87	55.62	58.39		15B
16.86066	18.42297	16.73313	15.36259	15.15720	16A
19.99293	20.97044	21.40163	22.80317		16B
28.59	24.06	22.00	36.06	38.44	17A
45.31	46.94	50.50	55.56		17B

 *
 DATA FOR PREDICTION
 *

31.56058					1P
19.82335					2P
14.43730					3P
96.708					4P
2.27496					5P
.					6P
2.77223					7P
3.21801					8P
2.843					9P
5.056					10P
12.546					11P
.134					12P
4.56					13P
1.10890					14P
62.23					15P
25.8756					16P

 *
 F LEVELS
 *

198.					2 F	LEV.
55.55					3 F	LEV.
31.33					4 F	LEV.
22.78					5 F	LEV.
18.64					6 F	LEV.
16.24					7 F	LEV.

```

*****
*
*INPUT DATA FOR HIGH STOCK PRICES*
*
*****

```

			NO. OF SETS	OF OBSERVATIONS	
9					
9.67177	10.91328	11.02571	13.57686	11.98247	1A
21.24328	24.90469	26.69125	29.90519		1B
6.83104	6.56233	5.32084	5.28485	5.60554	2A
12.45216	15.42614	15.61675	17.92321		2B
5.97494	6.53595	4.74300	5.20413	5.27743	3A
8.90804	9.89184	10.88025	11.93669		3B
66.096	56.555	68.393	52.809	66.488	4A
67.976	67.094	80.895	87.425		4B
.58615	2.5000	2.500	2.49460	2.39435	5A
2.58429	2.71430	2.33132	2.08647		5B
.	6A
.	6B
2.69968	2.72427	2.73810	2.53516	2.74583	7A
2.74583	2.75089	2.75780	2.76536		7B
2.72089	2.79639	2.84059	2.87536	2.88819	8A
2.88819	2.97253	3.04738	3.13652		8B
2.795	1.186	1.633	.068	2.979	9A
2.633	2.447	2.825	2.898		9B
4.542	2.482	2.944	1.162	4.514	10A
4.279	4.096	4.087	4.885		10B
7.	7.139	8.604	5.005	1.1884	11A
11.087	10.652	12.636	12.867		11B
.313	.24	.323	.415	.533	12A
.420	.235	.214	.179		12B
4.25	2.30	2.70	1.06	4.12	13A
3.90	3.72	4.36	4.42		13B
.53462	.54014	.54486	.54762	.54862	14A
.54917	.55018	1.10312	1.10614		14B
39.74	41.92	44.28	44.44	48.27	15A
50.26	51.87	55.62	58.39		15B
16.86066	18.42297	16.73313	15.36259	15.15720	16A
19.99293	20.97044	21.40163	22.80317		16B
31.63	29.75	25.38	46.75	46.38	17A
58.88	58.00	59.00	62.25		17B

HIGH
HIGH

```

*****
*
*DATA FOR PREDICTION*
*
*****

```

31.56058					1P
19.82335					2P
14.43730					3P
96.708					4P
2.27496					5P
.					6P
2.77223					7P
3.21801					8P
2.843					9P
5.056					10P
12.546					11P
.134					12P
4.56					13P
1.10890					14P
62.23					15P
25.8756					16P

```

*****
*
*F LEVELS*
*
*****

```

198.					
55.55					2 F LEV.
31.33					3 F LEV.
22.78					4 F LEV.
18.64					5 F LEV.
16.24					6 F LEV.
					7 F LEV.

```

*****
*ANSWERS FOR LOW PREDICTED STOCK PRICE*
*
*****

```

```

VARIABLE ENTERING      10
F LEVEL                41.8730
STANDARD ERROR OF Y   1.3918
CONSTANT              -42.55859
COEFFICIENT           1.51753

```

```

VAR          COEFF          STD ERR COEF
X 10         1.5175         .23451
ONLY 1 VARIABLE IN THE REGRESSION
THEREFORE NO R OR F CAN BE CALCULATED
FOR THIS MODEL SIMPLE CORRELATION COEFFICIENT = .925619
PREDICTED VALUE =      51.87736

```

```

*****
*
*ANSWERS FOR AVERAGE PREDICTED STOCK PRICE*
*
*****

```

```

VARIABLE ENTERING      3
F LEVEL                45.8676
STANDARD ERROR OF Y   1.5448
CONSTANT              17.63129
COEFFICIENT           2.08561

```

```

VAR          COEFF          STD ERR COEF
X 3          2.0856         .30795
ONLY 1 VARIABLE IN THE REGRESSION
THEREFORE NO R OR F CAN BE CALCULATED
FOR THIS MODEL SIMPLE CORRELATION COEFFICIENT = .931447
PREDICTED VALUE =      53.34393

```

```

*****
*
*ANSWERS FOR HIGH PREDICTED STOCK PRICE*
*
*****

```

```

VARIABLE ENTERING      3
F LEVEL                31.6713
STANDARD ERROR OF Y   2.1691
CONSTANT              21.95971
COEFFICIENT           2.43349

```

```

VAR          COEFF          STD ERR COEF
X 3          2.4334         .43241
ONLY 1 VARIABLE IN THE REGRESSION
THEREFORE NO R OR F CAN BE CALCULATED
FOR THIS MODEL SIMPLE CORRELATION COEFFICIENT = .904979
PREDICTED VALUE =      63.62910

```

A DYNAMIC MODEL FOR STOCK PRICE PREDICTION
UTILIZING THE STEPWISE MULTIPLE LINEAR REGRESSION TECHNIQUE

by

Jorge H. Gonzalez

B. S., Kansas State University, 1965

AN ABSTRACT OF A MASTER'S REPORT

submitted in partial fulfillment of the
requirements for the degree

MASTER OF SCIENCE

Department of Industrial Engineering

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1966

This report deals with a statistical investigation of the stock market in order to obtain a dynamic forecasting model for predicting future stock prices. Financial ratios for the company are used as independent variables in a multiple linear regression to predict the dependent variable, the stock price.

The prediction is made for the high and low prices of the stock. This information provides a built-in indicator for timing movements within the market.

The financial ratios which are used in this model are the common ones used by many investors to gauge the desirability of a company's securities. The wide spread use of these ratios by investors, in turn, affects the market prices since the prices are determined by investor demand.

The ratios are calculated for the company in question, and a stepwise procedure is used in selecting the variables to form a regression model for predicting stock prices. From the predicted prices the appropriate investment moves may be made.