THE IMPACT OF CERTAIN VARIABLES ON COMMON STOCK PRICE FORMATION:

A REGRESSION ANALYSIS

by

PAUL STUART GELLER

B.A., Haverford College
(1983)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREE OF
MASTER OF SCIENCE
IN MANAGEMENT
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
February 1985

© Paul Stuart Geller 1985

The author hereby grants permission to M.I.T. to reproduce and distribute
copies of this thesis document in whole or in part.

Signature of Author: ____________________________ Sloan School of Management

January 18, 1985

Certified by: ____________________________ Terry A. Marsh

Thesis Supervisor

Accepted by: ____________________________ Jeffrey A. Barks

Director of the Master's Program

ARCHIVES
MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

MAR 15 1985
LIBRARIES
THE IMPACT OF CERTAIN VARIABLES ON COMMON STOCK PRICE FORMATION:  
A REGRESSION ANALYSIS

by

Paul Stuart Geller

Submitted to the Sloan School of Management  
on January 18, 1984  
in partial fulfillment of the requirements for the degree of  
Master of Science in Management

ABSTRACT

This paper represents an attempt to determine the importance of a number  
of variables in the formation of common stock prices. The relevance of  
many of these variables is based in theory espoused by institutional  
investors on Wall Street and elsewhere and by individual investors that  
find credibility in these theories. These theories often involve concepts  
such as technical analysis and the non-efficiency of the stock market (at  
least in the semi-strong form, and as technical analysts would assert, in  
the weak form too), and run counter to the precepts of Modern Portfolio  
Theory (MPT).

Nonetheless, because so many investors seem to be active supporters of  
and implementers of the theories backing some of the variables (in using  
these and similar variables in their investments), it was determined that a  
thorough analysis of these variables and their importance (or unimportance)  
in determining stock prices would be useful. The ensuing study used the  
technique of multiple regression, and involved the performance of various  
cross-sectional and time-series regression runs.

The runs of both kinds that had been initially selected, on theoretical  
grounds, to be the runs of choice, showed relatively high significance in  
most regression coefficients and relatively tight regression fits for the  
equation as a whole (as evidenced by a variety of statistics). This was  
especially surprising regarding the final cross-sectional run, which  
included more of those counter Efficient Markets Hypothesis and MPT type  
variables than did the final time-series run. The results seemed to  
indicate that 1. there seems to be a fairly high correlation between many  
of the variables investigated and stock market prices, and 2. knowledge of  
these correlations and the magnitudes of the regression coefficients would  
not necessarily lead to profitable predictive ability.

Thesis Supervisor: Terry A. Marsh  
Title: Assistant Professor of Finance
Table of Contents

TITLE PAGE.................................................................1
ABSTRACT ...........................................................................2
TABLE OF CONTENTS..........................................................3

CHAPTER ONE - THE CROSS-SECTION
Section 1.1 - Introduction..................................................4-8
Section 1.2 - The Model......................................................8-27
Section 1.3 - Methods and Estimation Procedure...................27-29
Section 1.4 - Results.........................................................29-48
Section 1.5 - Summary and Conclusions and
Directions for Future Research.........................................49-56

CHAPTER TWO - THE TIME-SERIES
Section 2.1 - Introduction..................................................57
Section 2.2 - The Model......................................................57-71
Section 2.3 - Methods and Estimation Procedure...................72
Section 2.4 - Results.........................................................72-78
Section 2.5 - Conclusions and Directions
for Future Research......................................................78-81

BIBLIOGRAPHY....................................................................82-87
NOTES.................................................................................88-90

APPENDIX A - Regression Results, Cross-Section..................91-92
APPENDIX B - Interpolation (for levels of significance
of selected coefficients, Cross-Section).........................93-94
APPENDIX C - Industry Classifications, Cross-Section............95-96
Industry Classification Assignments................................97-100
Process Generating Appendix C.......................................101
APPENDIX D - Variance-Covariance Matrix, used in the
computation of t-values for coefficients
whose slopes have been unconstrained via
interaction terms (pp. 33-38).............................................102
APPENDIX E - Alternative Equation formulations..................103
APPENDIX F - Regression Results, Time-Series....................104
APPENDIX G - No Heteroscedasticity Plot............................105
CHAPTER ONE - THE CROSS-SECTION

1.1 Introduction

Various factors determine stock market prices. Speculators, using techniques of fundamental, technical, or quantitative analysis (or some combination thereof), try to exploit these factors by anticipating changes in share prices before they occur. Many academics contend that stock prices follow a random walk; this paper will address that issue in investigating the possible existence of factors with which prices may change in a somewhat predictable way (or, at least, with which prices may be correlated in some meaningful way). The explanatory variables used in the paper's regression analyses of stock market prices are those which seem to the author to be most important to those investors that are largely responsible for price movements (i.e., the investors—institutional for the most part—at the "margin").

However, the magnitudes of the effects of the various factors probably change somewhat rapidly over time with the occurrence of significant world events, changing speculative moods, and the like. Thus, a cross-sectional study of many stocks at one point in time seems like an appropriate start (so that we can determine the "current" levels of the coefficients), and it is expected that such studies carried out over very different time periods will yield differing regression coefficients. This study uses relatively recent data and thus the formulations described apply only to the present market.

(In much of the following, whether individual or institutional investors
are being referred to may not be stated explicitly; hopefully, this should be made clear by the particular context within which the reference is made.)

Much empirical work has been done concerning the market for common stocks. Unfortunately, because so many of the common stock related variables are non-numeric in nature and difficult to include in a regression equation, there has been considerable difficulty in arriving at a model which might bear out this paper's thesis. Researchers have attempted to formulate models that seem, sometimes, to be misguided. Any model attempting to use a dependent variable such as share price should obviously concern itself with those factors that (as nearly as we can tell) directly cause it to change. However, in many of the studies examined, some obscure measures or pieces of financial data, likely known only to a company's accountants or managers, have been used as explanatory variables in regression. It is true that correlations between such variables and stock prices may exist at a given time, but very often only because those obscure variables which are used are related to some other variables which more directly cause price movements. These latter variables are the ones that investors use in making buy and sell decisions, and these are the variables we should seek out in modeling a useful equation since they will obviously have higher correlations with the dependent variable than some proxy which might have had only a coincidental or one period correlation. For example, Marc Nerlove regressed rate of return on "mean retained earnings per dollar of total assets" [1], while the price/earnings ratio might have been a more appropriate regressor (as a more publicized
variable, P/E would be more effective in picking up the perceived cost of earnings of a company). The fact that changes in these variables may cause price movements which seem irrational given some determined or perceived "intrinsic" value or book value of a company does not matter. Granger and Morgenstern write in **Predictability of Stock Market Prices** that "Some problems can perhaps be overcome by making the equations sufficiently complicated. Even then, the knowledge that a company seems underpriced according to one's complicated formula is of little use if the rest of the investing public does not realize it" [2] (the emphasis is theirs).

In other studies researchers have often made the polar mistake of oversimplification. That is, when they have correctly attempted to use explanatory variables that are obviously used by investors they have in almost all cases left other relevant ones out. Some studies have regressed measures of share prices or the rates of return on shares against dividends and price/earnings ratio, but nothing else. Though these variables are certainly often considered when stocks are bought and sold, there are others which are quantifiable and which are also related to price movements that should be included in multiple regression, because if they are not our estimates will certainly be less efficient (and perhaps biased, if the omitted variables are correlated with the independent variables already present) and our overall fit less close. This study, as much as possible, attempts to include such variables.

It must be emphasized that the movement of stock prices is governed by many intangibles and variables that cannot be incorporated (easily) into a regression equation. Unanticipated world events such as wars, oil
embargos, and assassination attempts can affect prices in any number of ways, or some seemingly inexplicable change in speculative mood may occur and disrupt (or enhance) market confidence. Expectations and psychology also play an important role in share price determination, making the specification of a useful equation an even more difficult task. Thus, the goal in undertaking the following study was not the unrealistic one of the achievement of a very accurately predictive equation. Rather, there was sought a fairly general equation, applying over most industries (though only over a limited time period) from which could be obtained reasonable measures of the sensitivity of share price to certain measurable variables (i.e., measures of the regression coefficients).

G.R. Fisher, in his article "Some Factors Influencing Share Prices" suggested that perhaps "the variables used so far are not sufficiently reliable indicators of the market's view of expectations, and this deficiency must be allowed for" [3]. This paper seeks to improve upon the models specified thus far by using a new measure of the dependent variable and by including many of the independent variables that investors seem to react to and which should thus be significant in a regression equation. The random walk theory that many academics expound, which, simply stated, supposes that stock price changes are as random as the outcomes of the flip of a coin, applies where all other factors are held constant. When some or all of the clearly important investment variables change, stock prices should change in a (hopefully) predictable way – the rub is that we may not be able to react quickly enough to reap any profit from this knowledge. Malkiel points out in his A Random Walk Down Wall Street that stock price
changes do have an inherent logic in that they move in predictable ways with various fundamental variables. The problem, of course, lies in specifying a model that can bear this out to any significant degree, which may not be possible if in all cases price adjustment to new information is as fast as the Efficient Market Hypothesis says its is (i.e., virtually instantaneous).

In the most extensive equation specification of this paper, seven quantitative (continuous) variables and three qualitative (dummy) variables were used. The quantitative data were all taken from "Business Day" sections of The New York Times. Information needed for the determination of the states (i.e., "on" or "off") of the dummy variables was gathered from several primary and secondary sources, listed in the bibliography. The method used to collect this information is explained in Appendix C. Fifty-two companies were examined.

1.2 The Model

"In a perfect market with perfect certainty the value of a security would be the sum of discounted future returns. But in reality the capital market is imperfect and uncertainty prevails. Each of these factors will influence price" [5]. That is, at this point in time the market is not capable of completely and instantly processing and utilizing relevant information, or of removing structural inefficiencies resulting from the mechanics of trading and the segmentation of this information. As a
result, it is incapable of completely and instantly discounting the value of important data or news items. Because such inefficiencies exist, knowledge of some of the factors alluded to above may, it has been contended by many, enable us to predict prices to some degree (depending on the rates of adjustment to the various kinds of data and information).

The model should, as delineated in the introduction, take into account as many of these factors as possible. The independent variables used in this attempt were: 1. per cent yield, 2. price/earnings ratio, 3. distance from the 52-week low divided by price, 4. volume, 5. risk, 6. (yield)x(risk), 7. (distance from low/price)x(risk), 8. a Chemicals industry dummy, 9. a Petroleum dummy, and 10. a Natural Gas Utility dummy. The dependent variable used was a measure of standardized share price. We start by defining the following:

\[
y = \text{per cent yield}
\]

\[
P/E = \text{price/earnings ratio}
\]

\[
d/P = \text{the fractional distance from the 52-week low}
\]

\[
V = \text{volume}
\]

\[
R = \text{risk}
\]

\[
YR = (\text{yield})x(\text{risk})
\]

\[
dR = (d/P)x(\text{risk})
\]

\[
C = \text{Chemicals dummy}
\]
P = Petroleum dummy

NG = Natural Gas Utility dummy

O = "Average Industry" dummy

STP = standardized share price

The "Average Industry" dummy is used as the base category and will be described in greater detail later. It brings the number of dummy categories to four, requiring us to enter three dummy variables into the equation. The magnitudes of their coefficients are examined relative to the magnitude of the base category, which is given by the y-intercept. (Admittedly, the fourth dummy category was chosen in kind of a backwards way, after we had already decided on the three categories we wanted to include explicitly).

Since the included dummies are certainly not exhaustive regarding possible industry classifications, the choice of the catch-all "Average Industry" dummy becomes useful as it indicates that we are comparing these three industries to one which represents in some way an "average" industry rather than to one which completes some list of all possible industries (because this would require all possible industries less one to be included in the regression equation). We should, theoretically, be able to use one of the three specific dummies as a base category and enter the base explicitly into the equation (that is, as an on/off variable and not as the intercept which is, in effect, always "turned on" with the value of one). If for some reason this was desired (for example, to see what happens to
the t-values of the other dummies upon the switch), it might be useful to enter this dummy as a series of alternating 0's and 1's. This will also be elaborated upon shortly.

Next, we specify the equation simply as

\[
\text{STP} = f(y, P/E, d/P, V, R, YR, dR, C, P, NG) \quad [1.2-1]
\]

A discussion of all of the above variables follows. Ones that were ultimately eliminated, either on theoretical grounds or because of some technical inability to include them, are discussed in the last section.

1. Per cent yield. This variable is used by many market participants, who often seek returns largely on the basis of dividends that they will receive, perhaps with an eye to some price appreciation and capital gains over time. However, the absolute level of the dividend is not what they examine when considering a stock for purchase; this is merely some dollar amount paid out per share. Rather, on any given day, they look over to the next column in their *The New York Times* or *Wall Street Journal* listings, at "per cent yield". Here the value of the dividend is divided by price per share, resulting in a percentage dividend yield figure for all shares. This is obviously a more useful measure because it can be, to some degree, compared to the market rates of return elsewhere, such as on bonds, T-bills, or money market funds.

In addition to dividends providing an indication of rates of return, Fisher writes that "dividends are the shareholder's income and are indicative at least of a firm's current profitability." Thus, high dividends will not only attract investors to a stock because of the
immediate returns they indicate, but also because they indicate some
measure of the soundness of a company and its ability to keep paying those
dividends. Fisher goes onto say that dividends are "the primary and active
decision variable in most situations" [7].

However, though dividends are very important and nearly all models have
included them, the dividend variable should not be used as the only
independent variable in regression (contrary to some studies which have
unsuccessfully attempted to use it in this way). It is very unlikely that
many market speculators look down the list of stocks on any exchange and
pick out the highest yielding stock for purchase on that basis alone. Many
other variables must be and are taken into account by investors, and they
must be included in multiple regression to give us any chance at a
predictively useful or informative model.

The expected sign for this variable's coefficient is obviously
positive.

2. Price/earning:s ratio. This variable came to mind next, as it is one
often talked about by brokers and analysts and is listed alongside per cent
yield in the newspapers every day, for all to see. It is a measure of
price per share divided by net company earnings for the year per share, and
is said to give some indication of the valuation of a stock and its
earnings by investors. If a stock is appealing, investors will discount
this over time and raise up its price. On any given day, however, a high
P/E may indicate a stock that has already seen a price rise and is ready
for a fall, a stock that has low earnings, or some combination of the two.
Thus, so the reasoning for buying low P/E stocks goes, we might expect a high P/E on one day, all other things being equal, to result (on average) in a fall in price the next day as a result of selling pressure which tends to build on technical grounds as the stock price goes up (and thus, that this variable's coefficient would be negative). (The author does not necessarily subscribe to this reasoning, however. Only because so many in the industry give it credence is this variable investigated here.)

Whether or not P/E does have any of the explanatory power attributed to it by the above reasoning, it may be that this variable takes in other effects which other studies have attempted to include through the use of many, often complicated variables, as did the aforementioned Nerlove study. In all likelihood, the majority of investors rarely use variables such as "rate of growth of net sales" and "mean retained earnings per dollar of total assets" (as he does) because P/E gives them the simple and useful measure of the characteristics of a company they would be seeking, and is thus more justifiably included in a regression that seeks to predict, or at least tell us something about, stock price changes.

Again, P/E cannot be used alone in the regression, or even just with dividends. Other variables must be considered.

3. Fractional distance from low. This variable is a measure of the distance of a stock price on the day of the study from its 52-week low, divided by price. The idea is, if a stock is at or very near its low for the year, it may tend to rise briefly in knee-jerk fashion (a "technical bounce," according to industry jargon). Investors often follow a stock for
a number of weeks before they decide to buy. In fact, many investors look at the stock pages every day, making mental or written notes of purchase prospects. When a stock that has always clustered around a certain level and has always rebounded to that level reaches down toward its 52-week low, it is often thought that the price will rise back up to some more "natural," less overly pessimistic level. Sometimes, it seems, in response to such thinking it is bought up and the prophecy is fulfilled. For this reason it might be expected that if distance from low increased in a given period, price would tend to fall in the following period (and similarly, if the distance from the low was small, buying pressure would increase, raising price). Thus, this variable was expected to be negative.

Nevertheless, my results prompted me to change my theory on this matter, as I discuss in section 1.4. This is another variable that is not exactly solidly based in modern portfolio theory, but, since so many institutional investors do seem to give it or something like it some credence, the author felt it should be investigated.

It should be pointed out that the absolute distance from the low for each stock was divided by its price merely to bring this variable to the same general level for all companies. That is, shares of companies selling at a higher price (for whatever reason) obviously have a greater chance of being further from their lows in absolute terms; by dividing by price, I sought to make distance from lows relative and thus more usefully comparable.

4. Volume. This is the last major piece of information listed in the daily stock pages (for each company). Any information so readily available
will inevitably be used by speculators who are avaricious for useful and easy to obtain data. In addition, this variable is also one often (though not as often as P/E) talked about by brokers and advisors. High volume often indicates active interest (either positive or negative) in a stock, which often breeds more interest, serving to reinforce the effect.

Technical analysts proclaim volume to be a very important determinant variable and use many permutations of it in their research. They say that it, along with price, can tell them something very important about supply and demand that cannot be gleaned from fundamental variables. The author here again parts ground with such prophets (being well-schooled in the erroneousness of technical theory), but includes the variable because of the possibility that so many investors do use the above reasoning that it becomes important, and because the variable may be important for other reasons.

5. Risk. This is the last non-interacting continuous variable that was included. It is an example of one of those factors that seems very sensible to include but that is somewhat more difficult to deal with than the others described because it is non-numeric in nature and because there is no definitive, completely satisfying way of quantifying it. I could have used the volatility implicit in corresponding option prices, but most of the companies in this study do not have options on them trading on an organized exchange. Other measures of risk, such as the standard beta, were ruled out because of unavailability or unsuitability. In an effort to include something in the equations which might serve as a measure for risk, the standard deviation of the price of each company's shares were
calculated for the 49 day period preceding the day of the study. It was hoped that this, in picking up swings in stocks' prices that investors might perceive, would serve as a measure of risk associated with each stock when incorporated into the regression equation. The exact choice of 49 days was somewhat arbitrary (it corresponded to that unbroken period over which I had collected the newspaper business section). Such a period is likely used for examination of prospects before a purchase is made, at least for individual investors (for whom inside tip and quick hunch buys are usually exceptions), and seems to be a reasonable time over which some notions of risk are formed. Although investors do not generally go through the computational exercise this study did, by following a stock's price every day they must certainly get a good feel as to its volatility (again, speaking of individual investors). The standard deviations that were calculated varied considerably, by a range of over an order of magnitude; it was found, in carrying out the calculations, that such differences in volatility were often obvious even from casual examination of the prices over time. Risk averse speculators should thus be able to perceive and stay away from the more volatile ones, and, (since they are the majority), should tend to drive their prices down (making the sign for the coefficient of risk negative). Thus the measure used is one intended to pick up the feel or intuition of investors regarding risk.

To these five variables were added different combinations of interaction terms and qualitative (or "0,1") variables. Regarding the latter, it was thought that it would be useful to include industry dummies, since they serve to recognize the different reactions of stock prices of companies of
different industries to changes in the same explanatory variables. In addition to this rationale for using such dummies (that inherent differences in reactions to changes in explanatory variables may exist among industries), we may also draw upon the reasoning that there are often periods when certain industries or groups of stocks are in favor or disfavor (with analysts, the media, and eventually the general public). Such groups will have residuals that are either considerably positive or negative with respect to the sample regression line, in which case inclusion of dummies should improve the overall fit.

Nerlove used industry dummy variables in his cross-sectional study of firms listed on the New York stock Exchange, and attained a high value of $R^2$ relative to similar studies that didn't use dummies. Using them leads to a better predicting equation since with the lower standard error of the estimate (i.e., standard error or s.e.e.) that is probably obtained the regression coefficients are probably closer to the mark for any given study (since lower s.e.e. leads, other things being equal, to lower variances for the regression coefficients).

Many different possible industry classifications were considered, and various combinations of ten of these were tried explicitly in regression (see appendices). Rationalizing most of these on theoretical grounds seemed arbitrary at best. This is because theory would have to be geared very specifically toward determining the nature of things on the day of the study in order to determine which ones ought to be included. However, it is very difficult to say which industry effects should be operant on a
given day and to know if these effects will remain operant. Thus, the ten 
(out of an original list of over 30) which seemed to have the best chance 
for improving the equation (e.g., a petroleum dummy and an atomic energy 
dummy) were tried. Of the combinations tried, most did not seem to add 
very much to the regression fit. In fact, some worsened the equation, not 
by lowering the $R^2$ but by lowering the coefficients of the five original 
variables, relative to regression on them alone. Thus, most of these 
dummies probably did not belong (at least for the day of the study, to 
which the equation most closely applies).

In retrospect, the results described above (and elaborated upon, along 
with a further description of the dummy selection process, in section 1.4) 
are not very surprising. The market, it seems, often moves in waves, 
responding almost entirely to one piece of news. For example, if in the 
face of rising interest rates the Fed makes evident new policy directives 
or major banks change their prime lending rates substantially, 1500 out of 
1900 stocks may, for the most part, may seem to ignore other information 
and react almost solely to this (at least regarding direction of change). 
In fact, almost every day we read in the newspapers something to the effect 
of "...Analysts cited <some occurrence> for the broad-base drop in the 
Dow..." Thus, it may not contradict expected theory that, for example, the 
atomic energy dummy (which was initially expected to be potent, given the 
general level of sensitivity toward this industry) turned out to be quite 
insignificant and detracted from the equation in general, since some piece 
of news not relevant to atomic energy may have just hit the market and 
caused sweeping and relatively uniform price changes.
However, the inclusion of one combination of three dummies did seem to improve the fit substantially. The dummies were: a Chemicals Industry dummy, a Petroleum Industry dummy, and a dummy for Natural Gas Utilities. An adequate discussion of why these dummies might have had explanatory power on the day of the study would require an in depth analysis of these industries for that day. However, with the benefit of hindsight, some possible explanations will be proposed in section 1.4. The rationale for including variables that were screened in this way (in opposition to standard econometric practice) is further discussed in a later section.

As has been indicated, it was also thought that the regression fit might be improved by "unconstraining" the equation via various interaction terms. The dummy variables were ignored in this regard, since, as indicated, these variables are somewhat difficult to rationalize in the first place (and thus, refining, their relatively rough-edged specifications seemed inappropriate) while all (six) possible continuous variable interaction terms were considered. Again, however, theory became somewhat elusive when it came to justifying the inclusion of specific variables. It seemed that any of the variables might be usefully unconstrained, in that the effect of any variable on prices might not remain constant as the level of any other variable increased (or decreased) indefinitely. On the other hand, for some or all of the variables this might in fact not be the case. Either way, without some mechanical manipulation of the data (as with the dummies) it is hard to guess what will be true for any of the variables regarding their interaction with others on theoretical grounds only.
So, as said, all of the possible continuous variable interaction terms were considered, namely: (yield)x(risk), (yield)x(d/P), (risk)x(d/P), (d/P)x(P/E), (P/E)x(yield), and (P/E)x(risk). As with the dummies, only a portion of these proved to improve the fit (via lowered s.e.e., increased $R^2$, and increased t-values for the regression coefficients). These were (yield)x(risk) and (d/P)x(risk). It is perhaps of note that risk is involved in both of these terms. Again, with the benefit of hindsight we will be able to discuss why these two interaction terms, of the six attempted, might have been important.

With the explanatory variables now outlined, the dependent variable should be described. It has been indicated that some measure of share price was desired. However, since prices of different companies are at different levels to begin with, using straight price would not give meaningful (P/X)s. It would tell us only, for example, that companies with a high level of dividends tend, on average, to be selling at high prices. This is something we naturally expect and doesn't help us in clearly revealing the cause and effect relationship between some changing independent variable and some changing standardized level of price. What is needed is a measure that will give generalizable results, so that a given (P/X) could apply to any company. Various possible measures were considered (three of these will be discussed shortly); the one that was ultimately chosen was the one that had been the most intuitively pleasing and the one which, thankfully, yielded the best results. The prices of shares of all companies were standardized in the following way.
First, from the price of each given stock on the day of the study was subtracted the mean price of the stock as measured over the previous 49 days. This mean was intended to represent a measure of the "usual" or average price of each company's shares in the same manner that standard deviation was supposed to reflect volatility. The difference (between $P_t$ and $P$) was then divided by the standard deviation for each of the companies under examination. The result is a measure of the distance of a stock from its "usual" price corrected for by the volatility of the stock. We say it is corrected for because a very volatile stock would not surprise us by being far from its mean, and the standardized price, diminished by a larger denominator, would take this into account. Thus, by using the transformation we arrive at a generalizable measure of relative levels of stock prices. To expand upon this: if a stock is at its mean level then its standardized price is zero, and a rise in dividends or some combination of movements of the independent variables might be expected to make it positive. If the stock has had a history of jumping around a lot, we don't want to misinterpret this as a high level of standardized price relative to other stocks, so we correct by dividing by volatility. Thus, it may still have a high standardized price (which will be referred to as STP from now on), but not quite as high as it would have, or should have been if it were not very volatile. Put another way, if a stock that rarely moves is very far from its mean on any given day due to changes in the explanatory variables, we want it to have a higher STP than a stock that is at the same relative distance from its mean and is responding to the same changes in the explanatory variables, but moves up and down a lot very often.
It should be mentioned that the STP independent variable that was ultimately used was, as said earlier, not the only one considered. Also tried, with reservations, were price/dividends, fractional change in price from the period \( t=1 \) to \( t=2 \) (i.e., \( \frac{(P_{i2} - P_{i1})}{P_{i1}} \)), and similarly, \( \frac{(P_{i1} - P_{i1})}{P_{i1}} \) as measures of standardized price. None of these were as intuitively pleasing as STP and, as shall be described, all turned inferior results.

The use of price/dividends as a measure of standardized price was based on the rationale that dividends will in general be higher with higher priced shares and that division by them should bring all prices to a more generalizable level. However, this rationale seemed weak to me, and in regressing price/dividends against my lagged independent variables less than useful results were expected for a number of reasons. First, the division of price by dividends was somewhat arbitrary, serving simply as a device that would mechanically level out the price measure but lacking in any theoretical underpinnings (unlike STP, where dividing by \( s \) was critical, as described in section 1.2). Dividends don't necessarily standardize price - some companies with very high priced shares don't even pay dividends. In addition, because the yield variable is equal to dividends/price, it was quite certain that using price/dividends as the regressand would result in high negative correlation between the dependent variable and its reciprocal independent variable and would turn per cent yield into a dominant variable (making all other estimates less efficient). This is indeed what happened, as a very significant and
negative coefficient for the per cent yield variable was obtained, with the
other variables all insignificant at the 95% level and three out of four
with wrong signs.

To make the above test of the price/dividends regressand more valid the
same regression was run but this time without the per cent yield data.
Price/earnings was the only variable that turned out significant here, but
it had the wrong sign, as did two out of the other three variables.

\[ R^2 = .2492 \] and \[ F = 3.236 \] values were much lower than in
the final runs. It is interesting to note that the Fisher study alluded to
earlier, with a similarly low average \( R^2 \) of .2792, had used this price
transformation. In any event, as a result of this testing and relatively
weak theory supporting it, price/dividends was eliminated as a possible
measure for the dependent variable.

As indicated, the second measure considered was that of fractional change
in price from some period \( t=1 \) to \( t=2 \). It wasn't clear whether \( \frac{\alpha - R}{R} \) should
be regressed against similar fractional changes in the independent variables
from the period \( t=1 \) to \( t=2 \) (in which case we say that explanatory variable
effects are transmitted to price during the period \( t=1 \) to \( t=2 \), without an
overnight lag between the changes of the independent and dependent variables),
against similar fractional changes in the independent variables from the
period \( t=0 \) to \( t=1 \) (the overnight lag exists), or against the absolute levels
the independent variables in period \( t=1 \). I tended to support the latter two
(and the existence of an overnight lag) over the first (and no overnight lag),
but, in the end, none of the regressions utilizing any of these measures of
the dependent variable worked out at all anyway. One reason was the nature of
the data that resulted from the transformations. Very often from one day to
the next \( \frac{\bar{P}_n - \bar{P}_i}{\bar{P}_i} \) was zero, while the explanatory variables, for this one given
change, had some high or low level not consistent with zero change. The
transformation of price in this manner was thus constraining the data so that
a reasonable fit became unlikely. Also, differencing in this way and then
dividing by price led to small magnitudes for the values of the dependent
variable and magnified the size of the error in comparison. Thus, \( \frac{\bar{P}_n - \bar{P}_i}{\bar{P}_i} \) was
also eliminated as a possibility.

The final measure of the dependent variable that was tried was \( \frac{\bar{P}_n - \bar{P}_i}{\bar{P}_i} \).
However, this possibility was not very pleasing because it had some of the
same drawbacks as the other two. As with price/dividends, dividing by \( P \)
it
served merely to level out values of the dependent variable with little theory
supporting it, and as with \( \frac{\bar{P}_n - \bar{P}_i}{\bar{P}_i} \) this type of differencing was sure to reduce
the size of the estimated STP's relative to the residuals and magnify the
observed error. Runs using this variable were made and turned in poor
results.

As a result of all this testing and the theory cited regarding STP, STP
was used as the measure for the dependent variable.

We are now ready, with the dependent and independent variables in hand,
to elaborate upon equation 1.2-1.

One of the first things that was decided regarding this equation was
that the continuous independent variables should be lagged. It was
believed the most sensible period for any lags would be one day (as
indicated earlier in the discussion of possible dependent variables), since investors can (and will) respond to changes in the levels of the explanatory variables this quickly (i.e., from one day to the next). Thus we have

\[
STP = f[(y_{i,t-1}, (P/E)_{i,t-1}, (d/P)_{i,t-1}, (V)_{i,t-1}, (R)_{i,t-1}, (YR)_{i,t-1}, (dR)_{i,t-1}, C_{i,t-1}, P_{i,t-1}, NG_{i,t-1}] \quad [1.2-2]
\]

By not lagging the dummies we ask whether or not a given company was in one of these industries on the day of the study.

Naturally, this equation could apply only over very limited ranges of all variables, namely over the ranges that could realistically exist on the chosen day for study. Any extreme levels of a variable, such as a 10,000 per cent yield, would render the model useless as reactions to such a level would not be generalizable and would thus be of little informational or predictive value. By such reasoning, the y-intercept (value of STP when all independent variables are zero), except for its role as a benchmark against which we compare the magnitudes of the three dummy coefficients, should not be very meaningful in this model.

When these regressions were initially run, without interaction terms or dummies, the initial idea was that the relationships were linear and that the variables should remain untransformed. Thus, the original model (presented in Appendix E as E-1) postulated that arithmetic (as opposed to logarithmic) changes in all variables were reflected in arithmetic changes in STP over all (useable) ranges. After an initial run of this, the above the above assumption was reconsidered regarding some of the variables.
Specifically, it was thought that as volume reached higher and higher levels diminishing marginal increases in price might occur, and therefore that the log of volume might be a more appropriate variable. This seemed likely for risk and yield also, though it was still believed that the P/E and d/P variables to be approximately linearly related to price.

Thus, another possible formulation was

\[
\text{STP} = \left( P \cdot \frac{P_E}{s} \right) = b_0 + b_1 \log(y) + b_2 (P/E) + b_3 (d/P) + b_4 \log(V) + b_5 \log(R) + b_6 (YR) + b_7 (dR) + b_8 (C) + b_9 (P) + b_{10} (NG) + e_{i,t-1}
\]

It probably wouldn't make sense to add interaction terms to a log-log form of the equation (i.e., a form of the equation which takes the logs of the dependent and independent variables) because such an equation already provides for interaction among the variables (as with, for example, the Cobb-Douglas production function in which capital and labor interact multiplicatively, not additively). Runs were performed with with dummies and logged forms of some of the independent variables (see Appendix A and accompanying output), but the dummies and the interaction terms were added together only to the linear model (which initially didn't give very different results from the semi-log equations anyway).

Thus, the final formulation was equation 1.2-3. It was expected that the regression coefficients would be characterized in the following way: 1. \( y \) - positive sign, high in magnitude, quite significant; 2. \( P/E \) - negative, uncertain magnitude, also quite significant; 3. \( d/P \) - positive, moderate in
magnitude, perhaps not quite so significant; 4. V – positive, low in magnitude, not as significant as first two; and, 5. R – negative, high in magnitude, uncertain significance. All these specific characteristics remain even less certain for the interaction terms and dummy variables. Possible alternative results exist regarding the signs of P/E and V. Regarding the former, among other things it is conceivable that if high P/E is due to high price (as opposed to low earnings) that investors would view a stock as strong rather overpriced, tending to shift this variable in the positive direction somewhat. Regarding the latter, the bandwagon effect on the downside might be stronger in magnitude than has been postulated, rendering this variable either negative or insignificant.

1.3 Methods and Estimation Procedure

As indicated earlier, one of the foremost intentions was to get as generalizable results as possible. Thus, no controlled method of selection was used in deciding which companies to include in the study. Rather, 52 companies listed on the New York Stock Exchange were randomly chosen, the only requirement being that they have positive earnings. Day t, the day from which STP values were calculated, was April 23, 1984, and day t-1, the day from which the levels of the independent variables were taken, was April 22; s and P were calculated for each stock by going back 49 days to February 16. (As indicated earlier, this was the longest continuous stretch over which I had collected newspapers with the data I needed).

Price data and the distance from the low variable were transformed as
described in section 1.2, and were then increased by a factor of ten so that STP and d/P would be on the same general level as the other independent variables. Regressions were then run in an attempt to confirm the postulated equation and to find out something about the magnitudes of the coefficients involved.

The proposed model is relatively straightforward and involves only a single equation. Ordinary Least Squares (OLS) is appropriately used because we accept the following assumptions as valid [8]: that the expected value of the residuals, given the independent variables, is zero \(E(u_i)=0\) for each \(i\); that the residuals \(u_i\) and \(u_j\) are uncorrelated, i.e., there is no autocorrelation \([\text{cov}(u_i, u_j)=0, i=j]\); that there is equal variance of \(u_i\) for each \(x\), i.e., the residuals are homoscedastic \([\text{var}(u_i)=0\) for each \(i]\); that the residuals \(u_i\) and explanatory variable \(X\) are uncorrelated \([\text{cov}(u_i, X_j)=0]\); and, finally, that there is no (extreme) multicollinearity among the independent variables \([X_j + X_{ji}=0]\). Because we do accept these assumptions, simple OLS becomes the method of choice. If, however, we were dealing with a simultaneous equations system, and had one or more dependent variables on the right hand side of the equals signs of the structural form equations, OLS would not be satisfactory. This is because the endogenous (dependent) variable on the right hand side, as an exogenous (independent) variable in another equation, would necessarily be correlated with the error term of the first equation (violating the OLS assumption of...
no correlation between $u$ and $X$). If we tried to impose an OLS fit on the
data we would incur "least squares bias," or simultaneous equations bias.
In such a case we would have to use methods such as two-stage least
squares, indirect least squares, or instrumental variables to eliminate the
bias. Two-stage least squares can also be used to find the regression
coefficients of an "overidentified" equation, i.e., one in which more than
one numerical value can be obtained for some of the parameters of the
structural equations (the number of reduced form equations is greater than
the number of reduce form unknowns — unique estimation of parameters is not
possible). For the purposes of this study, however, with its assumption
that a single equation model is valid and that additional equations would
be for the most part superfluous, classical linear regression (OLS), or
"naive least squares," is suitable. If at some time it was realized that
this assumption were incorrect and that some other equation should be added
to the single one initially postulated (such as one with some possibly
important macroeconomic variable on the left hand side, as the dependent
variable), it would then become prudent to be wary of the estimation
problems that might result and to keep in mind the aforementioned remedial
measures.

1.4 Results

Estimations of the regression coefficients and other information
regarding some of the equations are given in the table of Appendix A. Run A
represents the initial linear run with five continuous variables; run B
represents the initial log run with five continuous variables (logging, specifically, yield, volume, and risk); Run C represents this run plus two dummy variables; and Run D represents the final run, namely, Run A plus three dummy variables and two interaction terms.

Of the runs with only continuous variables, six of the ten coefficients are significant at at least the 99\% level (i.e., \( \approx 1\% \)), two at least the 90\% level, one at exactly the 83.5\% level, and one at exactly the 87.0\% level.

Run C, which includes the five untransformed continuous variables and two dummy variables (the statistical significances of these latter are ambiguous, as will be discussed shortly) had three of its five quantitative variables significant at at least the 99.95\% level, one at at least the 99.75\% level, and one at at least the 90\% level. In Run D, which is linear in all the continuous variables and uses dummies and interaction terms, one of eight quantitative variable coefficients was significant at at least the 99.95\% level, one at at least the 99.75\% level, one at at least the 99.5\% level, two at least the 97.5\%, one at exactly the 89.6\% level, and one at exactly the 82.8\% level. Again, the coefficients of the three dummies should not be addressed so unequivocally, for reasons soon to be explained.

All in all, the relatively high significances of the coefficients in these regressions seems somewhat surprising, given the lack of rigor of the theory backing a number of the variables. Perhaps the results are artifacts of some econometric flaw, or perhaps we ought to lend more credence to some of the included variables, for whatever reason they may warrant it, than we otherwise would have.
There is one piece of data that stands out in the chart of Appendix A, namely, the coefficient of risk in Run D. Upon first examination, it appears to be the wrong sign. However, this is an incorrect conclusion to reach, as will be elaborated upon shortly.

As said in section 1.2, the only way we can adequately assess the correctness or incorrectness of the inclusion of the dummy variables or interaction terms used is with the benefit of hindsight, since they were chosen in large part by how well they worked in the equation and only in small part by pre-formed theory.

Though the dummies were ultimately chosen in this less than desirable way, they were, in fact, screened to some extent before their inclusions (as indicated in section 1.2). In the first level of filtering industries that only took in one or two of the companies were eliminated because inclusion of them would surely serve to assign part of the residuals' magnitude to that of the dummy; but this seeming increase in the explanatory power of the equation would likely only be mechanical, and not dependent upon the presence of some logical dummy-type quality except for the fact that the observation for the dummied company is at a different level than others. The list thus obtained was narrowed further by discarding industry classifications that were not thought to be considered very heavily by investors in their gathering of information. On these grounds, such industries as "Printing & Engraving" and "Parking Lots" (as opposed to, e.g., the "Atomic Energy" industry, which was mentioned in the earlier discussion) were removed from consideration. Once the list of
feasible dummies was reduced to ten it was not possible to distinguish among the remaining dummy possibilities any further, and they were all thrown into Runs A and B, and the results analyzed. At this point various combinations of six, four, three, and two dummies were tried and the combination of the three included in Run D was found to work at least somewhat better than the others. The same sort of trial and error method was used with the interaction terms, in narrowing that list from six down to two.

In general, such data manipulation is not valid econometric practice if one is seeking a meaningful regression equation. However, it seemed that the inclusion of such variables might improve the fit for valid reasons, but for valid reasons which could not be specified adequately beforehand. The variables chosen do not represent random pieces of data thrown into an equation until a nice fit resulted — they existed among larger sets of data which all seemed rational as possible variables but which couldn't be easily separated beforehand, on strong theoretical grounds, into variables that would justifiably work in the regression equation.

As indicated earlier, we could probably look back and find some explanation for the fact that Natural Gas Utility companies and companies that dealt with Chemicals both tended to have positive residuals while companies that were in the Petroleum industry tended to have negative residuals on the day of the study. For example, perhaps some announcement had just been made regarding deregulation for one of the industries, or perhaps a new technology or increased earnings had just been announced. Any such announcement or event considered important by investors could have
produced the observed effects.

Similarly, we can, looking back, rationalize the significance of the two interaction terms included, \((d/P)x(\text{risk})\) and \((\text{yield})x(\text{risk})\). Since both include risk we might say that investors' reactions regarding other variables often tend to change with changing levels of risk, and that this was so more with risk than the others because risk is a more subjectively evaluated variable.

As intimated, such observations are tainted to some degree by the fact that we did not state them previous to regression tests which confirmed them but merely in response to these regression tests. They are cited in an attempt to give examples of possible justifications for including the variables and to show that such justifications may actually exist, even though they were not stated beforehand.

Some further elaboration should be made at this time regarding the attempted use of dummy variables in this cross-sectional analysis.

The dummy variables were introduced as a means of detecting systematic differences among share prices attributable to industry classes, or, put another way, to explain industry differences not being explained by the five continuous variables. But since these continuous variables form the basis of the paper's hypotheses, it is important to present regression results "obtained without the use of the dummy variables, or at least

\[ \text{note} \text{ the [change] in } R^2 \text{ which would have been obtained had the dummy variables been deleted} \] [9]. The first part of this suggestion has been
taken care of (see Appendix A). To be complete, the second part should now be addressed by calling attention to the change in $R^2$'s. As shown in the chart (Appendix A), the $R^2$'s did rise substantially with the inclusion of dummy variables, by about 43% and 21% respectively (by more if we compare to the runs with the interaction terms). The continuous variables still, however, do explain most of the variation in the dependent variable.

It is probably more valid, when we are interested in the increment to the explanatory ability of an equation from a set of variables such as industry dummies, to examine the $F$ value of the regression before and after the addition of the set of variables rather than the incremental $R^2$. This is true for two reasons. First, the $R^2$ will almost always rise as variables are added, though the significance level of the equation, as revealed by the $F$ test, may fall (especially if an added variable has a low t-statistic). Second, as Ronald Wippern states, "the magnitude of the increment to the coefficient of determination will vary widely with the order in which the variables are entered, and misleading indications of the relative explanatory powers of the sets of variables may be derived...[because of the] non-orthogonality among the independent variables, [which is] an unhappy but ubiquitous fact of life in economic research" [10]. Thus, the relevant $F$ statistics are also supplied in the chart. Again, they indicate that the added variables (the dummies and the interaction terms) add more to the explained variation in the dependent variable than to the unexplained. (Another useful statistic regarding the
above discussion would be $R^2$. For reference to these, see supplied output).

As was mentioned earlier, there is also an important point to be made regarding the ambiguity of statements that can be made about the significance of the coefficients of the individual dummy variables. These coefficients can be interpreted as "shift parameters estimating systematic differences between the respective classifications for which dummy variables are included in the regression and and that classification for which the dummy variable is omitted" [11] (underlining is used here rather than italics). As Brown points out, the predicted values for the dependent variable and the coefficient of determination ($R^2$) are identical regardless of which classification is omitted (i.e., regardless of which industry becomes the base category), but the coefficients of the dummy variables and their estimated standard errors for the included classifications are in fact functions of the classification which is omitted, and, as a result, "particular classifications of dummy variables may test 'significant' if one classification is omitted, but not significant if some other one is left out" [12]. So, unless there is a strong a priori reason for the omission of some particular classification, statements about the significance of a particular dummy variable's coefficients are likely rendered ambiguous, and use of significance tests on coefficients of dummy variables becomes less worthwhile. The results of these tests are presented, along with this qualification, in Appendix A.

The regression analysis of the dummy variables, in this paper, is
essentially an analysis of variance comparison between the base category industry and all other industries. The significance tests reveal, then, not whether each industry differs significantly from each other, but whether the included industries differ significantly from the omitted one.

The choice of which industry dummy should act as a base category in the actual regression is even more unclear and theoretically difficult than the original choice of dummies themselves. For this reason and the reason of a shortage of useable categories, the rather contrived industry, the "Average Industry" industry, was chosen as the base. As mentioned earlier, the definition of this base would be that every other company belongs in this classification, resulting in alternating 0's and 1's for observations of this variable. The magnitudes of the more explicitly included dummy classifications, then, would imply deviations from this, the "typical" industry. In any event, the t-values and the values of the coefficients of the continuous variables will not change with the choice of base category.

It should be noted that the ultimate decision concerning the elimination of specific industry classifications was not made on the basis of low t-values precisely because of the just described ambiguity of these measures, but rather, on the basis of what they did to the s.e.e. and F statistics, which tell us about the regression equation as a whole.

Regarding interaction terms and dummies, it remains to be shown how the coefficients for the variables involving interaction terms were obtained, and how the associated t-values for these terms were obtained. In addition, some interpretations of the equations with the included dummy
variables and interaction terms should be provided.

First, we write out the equation (without the time subscripts here for simplification):

\[
\text{STP} = b_0 + b_{1y} + b_{2(P/E)} + b_{3(d/P)} + b_{4V} + b_{5R} + b_{6(y \times R)} + b_{7[(d/P) \times R]} + b_{8C} + b_{9P} + b_{10(NG)}
\]

Using partial differentiation we can say

\[
\frac{d\text{STP}}{dy_i} = \hat{b}_i + \hat{b}_j R
\]

Since the level of risk varies with STP, we arbitrarily decide to use the mean of risk over all companies to get a measure that applies to the whole equation. If we wanted a value of a particular company, we would simply choose the corresponding level of risk and evaluate the above expression.

\[
\bar{R} = .99647, \text{ so that }
\]

\[
= 1.2876 + -.8167(.99647)
\]

\[
= .4738
\]

This represents a slight diminution from the value of .528 obtained for the linear constrained equation, indicating that with it we overstate the effects of changes in yield on STP.

The number .474 represents the change in STP that occurs with a unit change in yield at the mean of risk. It ranges from -1.015 to 1.168 over the range of risk (.1459 -- 2.8195). This indicates that a high level of risk makes people less inclined to have a given level yield persuade them to buy a stock, while a low level does the opposite. It also indicates
that, at the mean level, with the two variables not constraining each other, .474 should be a more accurate estimate than the originally obtained .528.

(So far in this analysis we have been tacitly assuming that the relevant F test will indicate that unconstraining the equation will explain more variation than it leaves unexplained. This will be verified shortly).

Similarly,

\[
\frac{\text{astp}}{\text{R}^{2}_{i}} = \hat{b} + \hat{b}_{3} \bar{R} = 9.0564 + -2.0214(.99647)
\]

\[
\frac{\text{astp}}{\text{R}^{2}_{i}} = 7.0421
\]

This value compares with the value of .56475 for the unconstrained linear equation. Similar arguments can be made for the interacting effects here as were made with yield.

Finally,

\[
\frac{\text{astp}}{\text{R}_{i}} = \hat{b} + \hat{b}_{5} \bar{y} + \hat{b}_{6} \left( \frac{d}{P} \right) + \hat{b}_{7} \left( \frac{d}{P} \right) + \hat{b}_{7} \left( \frac{d}{P} \right)
\]

\[
= 0.7689 + (-0.8167 \times 5.75577) + (-2.0214 \times 1.64819)
\]

\[
= -7.2635
\]

The fact that the value of \( \hat{b} \) is positive should not alarm us, as indicated earlier, because this number does not measure the whole effect of risk. As with the previous two calculations, we must add in the interacting effects we hypothesized. In a sense, .7689 is just a constant which corrects for the number we obtain by adding the two products on the far right of the equals sign. The result is a new measure of risk, which
takes into account the interacting effects of yield and d/P.

The value of \(-7.2635\) obtained above compares with the value of \(-2.1122\) for the unconstrained linear equation, indicating that by constraining we underestimate the effect of risk on STP. That is, at the mean of yield and fractional distance from low, a unit increase in risk will drop STP by 7.26 rather than by only 2.11 units.

At this point, it should be shown how the t-values associated with the composite regression coefficients calculated above were obtained. To get these t-values, we need to get the appropriate standard errors for the terms. Usually, the relevant variance is simply of the form \(\text{var}(ax) = a^2 \text{var}(x)\), but here, since the variance of the coefficients depends on two terms (and in the last case three), the relevant variance is of the form

\[
\text{var}(ax + by) = a^2 \text{var}(x) + b^2 \text{var}(y) + 2ab \text{cov}(x,y).
\]

For yield:

\[
\text{var}(\hat{\beta} / \hat{\gamma}_i) = \text{var}(\hat{b} + \hat{b} \bar{R}) = \text{var}(\hat{b}) + 2\bar{R}\text{cov}(\hat{b}, \hat{b}) + \bar{R}^2 \text{var}(\hat{b})
\]

Getting the appropriate numbers from the variance-covariance matrix (see output) we solve:

\[
\text{var}(\hat{\beta} / \hat{\gamma}_i) = .3199 + 2(.99647)(-.09151) + (.99647)(.08351)
\]

\[
= .2225
\]

\[
\text{std. error}(\hat{\beta} / \hat{\gamma}_i) = \sqrt{.2225} = .4717
\]

This calculation implies a t-value of \(4738/.4717 = 1.004\), which is significant at the 82.8% level. (Actually, it is customary to give significance levels as the value of \(\alpha\): the fraction of the t-distribution cut off in the
tail; here, for example, this would be .172 or 17.2%. However, within the
text significance levels have been described in terms of 1-\(\alpha\), because this
conveys significance levels in a manner that is slightly easier to comprehend.
The more precisely correct distinction between \(\alpha\) and 1-\(\alpha\) is made in
Appendix A, where results are presented).

We perform similar calculations for (d/P) and for risk, which are also
considered interactively:

\[
\text{var}(\overset{\text{stat}}{\bar{y}_i}) = \text{var}(\hat{b}_i + \hat{b} \bar{R}) = \text{var}(\hat{b}_i) + 2\bar{R}\text{cov}(\hat{b}_i, \hat{b}_j) + \bar{R}^2 \text{var}(\hat{b}_j)
\]

\[
= 4.524 + 2(.99647)(-.07874) + (.99647)(2.472)
\]

\[
= 6.8303
\]

\[
\text{std. error}(\overset{\text{stat}}{\bar{y}_i}) = \sqrt{6.8303} = 2.613
\]

\[
t = \frac{7.0421}{2.613} = 2.6945
\]

This is significant at the 99.5% level.

Finally, we have

\[
\text{var}(\overset{\text{stat}}{\bar{y}_i}) = \text{var}(\hat{b}_i + \bar{b} \bar{y} + \hat{b} (d/P))
\]

Paralleling the equation for the variance of a sum of two terms, we see that
the relevant variance is now of the form:

\[
\text{var}(ax + by + cz) = a^2 \text{var}(x) + b^2 \text{var}(y) + c^2 \text{var}(z) + 2\text{cov}(x,y) + 2\text{cov}(x,z) + 2\text{cov}(y,z)
\]

so that

\[
\text{var}(\overset{\text{stat}}{\bar{y}_i}) = \text{var}(\hat{b}_i) + \text{var}(\bar{y}) \text{var}(\hat{b}_i) + \bar{y}^2 \text{var}(\hat{b}_i) + \bar{y} \text{cov}(\hat{b}_i, \hat{b}_j) + \bar{y}(\frac{d}{P}) \text{cov}(\hat{b}_i, \hat{b}_j) + \frac{\bar{y}(d/P)}{d} \text{cov}(\hat{b}_i, \hat{b}_j)
\]

\[
\text{var}(\overset{\text{stat}}{\bar{y}_i}) = \text{var}(\hat{b}_i) + \text{var}(\bar{y}) \text{var}(\hat{b}_i) + \bar{y}^2 \text{var}(\hat{b}_i) + \bar{y} \text{cov}(\hat{b}_i, \hat{b}_j) + \bar{y}(\frac{d}{P}) \text{cov}(\hat{b}_i, \hat{b}_j) + \frac{\bar{y}(d/P)}{d} \text{cov}(\hat{b}_i, \hat{b}_j)
\]
\[ = 20.05 + (5.76)(.08351) + (1.65)(2.472) + 2(5.76)(- .4621) + \\
2(1.65)(- 5.794) + 2(5.76)(1.65)(.02079) \]
\[ = .5614 \]

\[ \text{std. error}(\hat{\beta}_i) = \sqrt{.5614} = .7493 \]
yielding a t-value of \(- 7.26 / .7493 = - 9.689\), which is significant at the 99.95% level.

Finally, at this point, we should in fact confirm that by unconstraining the equation we add more to the explained variation than to the unexplained.

The relevant formula is

\[ F_{m,n-k} = \frac{\text{UESS} - \text{CESS} / m}{\text{URSS} / (n-k)} \]

where \( \text{UESS} = \text{Explained sum of squares (unconstrained equation)} = (\sum y_i^2)_u \)

\( \text{CESS} = \text{Explained sum of squares (constrained equation)} = (\sum y_i^2)_c \)

\( \text{URSS} = \text{Residual sum of squares (unconstrained equation)} = (\sum e_i^2)_u \)

\( k = \text{number of constants in the equation} \)

\( m = \text{number of constraints in the equation} \)

If adding the interaction terms raises the explained sum of squares, CESS will be less than UESS and the F statistic will be large, indicating that we should reject the null hypothesis that imposing the constraint made no difference.

Another way to write the above equation is

\[ F_{m,n-k} = \frac{(\hat{\kappa}_u^2 - \hat{\kappa}_c^2) / m}{(1 - \hat{\kappa}_u^2) / (n-k)} \]

From the run of the five independent variables and three dummies with no
interaction terms (see accompanying output) it is known that

\[ R^2 = 0.4702 \]

The unconstrained equation, Run D, has

\[ R^2 = 0.5684 \]

Thus,

\[
F_{2,41} = \frac{(0.5684 - 0.4702)/2}{(1 - 0.5684)(41-11)} = 3.413
\]

This is significant at the 95% level (\( \alpha = 5\% \)), indicating that we have in fact explained significantly more variation than we have left unexplained by using the unconstrained equation. We thus see that the restricted and unrestricted regressions are different and reject the hypothesis that the parameters in question obey the linear restrictions imposed.

To summarize a bit, we may write out three specific (and arbitrarily chosen) possibilities for our final equation (again leaving off subscripts for simplicity):

1 - The company is in the "Average Industry" industry:

\[
STP = -5.51 + b_{11} y + b_{12} (P/E) + b_{13} (d/P) + b_{14} V + b_{15} R + b_{16} (YR) + b_{17} (dR) 
\]

2 - The company is a producer of petroleum and owns a natural gas utility:

\[
STP = -3.89 + b_{21} y + b_{22} (P/E) + b_{23} (d/P) + b_{24} V + b_{25} R + b_{26} (YR) + b_{27} (dR) 
\]

3 - The company is in the Chemicals industry:

\[
STP = 2.34 + b_{31} y + b_{32} (P/E) + b_{33} (d/P) + b_{34} V + b_{35} R + b_{36} (YR) + b_{37} (dR) 
\]

Thus, according to the final equation, if a company were involved with petroleum in some capacity we might have expected its values of STP to have
been 1.62 units higher than the typical industry's firms, while a company in the Chemicals industry would have, on average, STP higher by 7.85 units. Other cases could be delineated, all of which have the five original continuous variables plus two interaction terms freeing up some of the slopes of variables while levels of other variables change.

The regression results in general are fairly uniform for all of the runs. Because it was realized that very high R²'s were not likely, the ones that were achieved were in fact somewhat surprising. Two other regression studies that were, to some degree, comparable to this one were found, the Fisher and Nerlove studies mentioned earlier. Fisher regressed a different measure of price on dividends and undistributed profits and over 36 trials (varying time and stock group) got an average R² of only .2792 [13] (A likely reason for this was his inadequate dependent variable.) Nerlove regressed rates of return on eight explanatory variables and a number of industry dummies and got an average R² of .5293 [14]. Using two dummies and the five original variables this study obtained a similar R² of .5124. Using the same equation but adding interaction terms (and a third dummy) led to an R² of .5684.

The addition of interaction terms and dummies from Run A to Run D improved the s.e.e.'s relative to the estimated STP's (originally, s.e.e./STP = .918, subsequently s.e.e./STP = .843, an improvement of over 8%), but may still be considered high. As indicated earlier, though, this
was not unexpected because of the many known (and unknown) econometrically messy variables (e.g., some of the intangibles alluded to earlier) that affect stock prices but which are not easily included in the regression equation.

The Durbin-Watson statistics are all very close to 2, indicating that we cannot even nearly reject the null hypothesis of no autocorrelation. (Speaking less precisely, we could say that we accept the hypothesis of no autocorrelation or that we reject the hypothesis of autocorrelation; neither are as correct as the first, more conservative statement, which uses the double negative). This eliminates the need for corrective measures such as the Cochrane-Orcutt procedure, which is not surprising for such a cross-sectional study.

The F test indicates that in all runs significantly more variation is explained by the independent variables (together) than is left unexplained.

No multicollinearity was expected for these runs and none was indicated by the printed diagonals of the \([X'X]\) matrices, as none of their elements were very close to zero. Examination of the R 's, F 's, zero-order correlations, partial regression coefficients, and partial correlation coefficients as suggested by Gujarati [15] also gives no indication of multicollinearity among the independent variables. Further, no heteroscedasticity seems to be indicated by the plot of the squared residuals against \(\hat{STP}\) (included for Run B). We could go on to plot \(e_i^2\)
against any of the explanatory variables in an attempt to find specific indications of heteroscedasticity that may have been lost in the more general plot performed. But since there are no strong theoretical reasons to believe that variance of STP should change to any significant degree with changing levels of any of the independent variables, and since such an exercise would be very time consuming, this has not as yet been followed through.

More standard estimates of the sensitivities given by the regression coefficients can be made by computing the elasticities of STP with respect to each of the independent variables. This is done for Run B below.

For all equations, \( \eta \) (elasticity) = \( \frac{\Delta \text{STP}}{\text{STP}} \frac{\Delta X}{X} = \frac{\Delta \text{STP}}{\Delta X} \left( \frac{X}{\text{STP}} \right) \) [where \( X \) is some independent variable]

For all variables that enter the equation linearly, \( \frac{\Delta \text{STP}}{\Delta X} \) = the coefficient of \( X \), so that \( \eta_{\text{STP}}^{\text{P/E}} \) (elasticity of STP with respect to P/E) = \( \hat{\beta}_2 \left( \frac{\Delta \text{P/E}}{\Delta \text{STP}} \right) \)

\( \eta_{\text{STP}}^{\text{P/E}} = \hat{\beta}_2 \left( \frac{\Delta \text{P/E}}{\Delta \text{STP}} \right) \)

If we evaluate the quotients \( \left( \frac{\Delta \text{P/E}}{\Delta \text{STP}} \right) \) and \( \left( \frac{\Delta \text{P/E}}{\Delta \text{STP}} \right) \) at the means of the variables, we get

\( \eta_{\text{STP}}^{\text{P/E}} = -0.3577 \left( \frac{6.112}{9.173} \right) = -0.3397 \rightarrow \) a 1% rise in P/E leads to a 0.34% fall in STP,

\( \eta_{\text{STP}}^{\text{P/E}} = 6.0841 \left( \frac{1.471}{9.173} \right) = 1.1278 \rightarrow \) a 1% rise in P/E leads to a 1.13% rise in STP

For variables that enter the equation in log form, the calculations are a little different:

\( \frac{\Delta \text{STP}}{\Delta X} = (\text{coefficient of } X)(1/X) \)

\( \eta = (\text{coefficient of } X)(1/X)(X/\text{STP}) = (\text{coefficient of } X)(1/\text{STP}) \)

This indicates that with the logged variables, as STP increases declines, but that \( \eta \) does not increase with the level of the independent variable.

Again, we evaluate \( \eta \)'s at the mean of STP:

\( \eta_{\text{V}} = 3.4940/9.173 = .381 \)

\( \eta_{\text{STP}}^{\text{ST}} = 2.192/9.173 = .239 \)
\[ \eta^{nc}_C = -5.7372/9.173 = .625 \]

It is perhaps of value to note that some preliminary runs were performed with various combinations of independent variables excluded, in order to test the correctness of their inclusions. This testing was different from the data manipulation described regarding the interaction terms and dummies because it came after theory was set down and the equation specified.

One variable that was questioned at the outset was volume. It was thought that it might actually be a dependent variable on the right hand side of the equation (implying the need for at least one more equation, and the possibility of simultaneous equations bias in the model as specified). It seemed that this was probably not the case, as it was ultimately felt that volume is generally not determined, for the most part, by the other independent variables or by price. Rather, it is a more a reflection of speculative interest in a company, of the size of a company, and the like, and (when used) is used as an independent variable by investors as is \( y \) or P/E. In any event, just to see what would happen, a regression run with volume (which, by the way, always came up very significant) deleted was tried using, as in all runs except the dependent variable trial runs, \( \text{STP}^2 \) as the regressand. The s.e.e. went up, the R halved, and some significance was lost in all of the regression coefficients, all indicating that volume probably belongs.

Another variable that seemed to warrant some testing was risk, because of some of the likely troubles with it outlined earlier. To repeat, risk is simply not (generally) calculated by investors and analysts.
mathematically, but is evaluated mentally in day to day tracking (though, of course, many companies and services, such as the Value Line Investment Survey, do publish companies' betas). Some confirmation was desired that the measure of risk used was indeed picking up the volatility that it was thought people would recognize and use in their valuations of the variable. So, in an effort to justify its inclusion, a run of STP against the four independent variables with risk deleted (even though \( \hat{B}_5 \) came out significant in virtually every run) was tried. The same thing happened as with the deletion of volume — \( R^2 \) halved, \( F \) fell, and significance was lost in all of the regression coefficients.

Even one run with both volume and risk deleted was tried, and things got even worse. If volume or risk had been irrelevant variables then introducing them should have introduced "noise" into the equation and raised s.e.e. Since the opposite happened we conclude that these are relevant variables and are usefully (necessarily) included in the regression equation.

While the theory outlined earlier seemed to indiacte that P/E and percent yield be included in the equation, it did not necessarily do so for the same for the third variable, d/P. So, even though this is the variable that came out most significantly in every single run and is the only one which never changed sign when the regressand was one of the three really unsuccessful ones, a run with the (d/P) data excluded was tried. These results were very poor; \( R^2 \) plummeted to .1762, s.e.e. went up to 9.763,
and all but one of the regression coefficients became quite insignificant.

A point mentioned earlier should at this time be addressed. In section 1.2, it was indicated that because of the effects of presumed selling (due to "profit-taking", according to Wall Streeters) when d/P is high and expected capital gains from purchases when d/P is low, the coefficient of \( \hat{b}_3 \) might be expected to be negative. As said, it was positive in every run made and the most significant of all the variables in every run made. Obviously, there was some flaw in the original theory. It was decided that the analytical error stemmed from the author's being a risk "lover" and one who likes to gain or lose quickly; as a result, in modeling this variable capital gains was given too much attention relative to dividend income. Apparently, it seems when distance from the low is high, implying a high price, most investors do not look to cash in but rather hold on, hoping, in conservative fashion, that the company will remain stable and that price rises will continue. Similarly, nearness to the low makes most investors wary of further declines rather than creating dreams of possible capital gains. Investors are, it would seem, interested (for the most part) in the soundness of a company and its ability to continue to pay dividends – they will view high or rising share prices as a signal of this and low or falling prices as an opposite signal, of instability. Thus, a large d/P will indicate that a company is doing well and will induce buying and price rises, and the opposite will occur when d/P is low, rendering \( \hat{b}_3 \) positive.
1.5 Summary and Conclusions and Directions for Future Research

If, as some contend, the stock market is indeed completely efficient regarding its utilization and incorporation of the information which governs stock price movements, then it would be true that stock prices, following a random walk, are completely unpredictable based on knowledge of any of this information. This would truly have far-reaching implications, because it would negatively affect those industries and individual investors who make their living trying to find "cheap" stocks and to "beat" the market. Investors would realize that any information that they could come to possess regarding a stock would be discounted into its price and that there would be no better than average expectations of making capital gains on that specific information. As a result, interest in the secondary market would diminish. And, in a similar fashion, interest in new issues would wane, as there would be no such thing as "a bargain" (as some of the recently issued high-tech stocks were viewed) — if for any reason that could presently be known a new issue would be expected to rise in price, that rise would be effected immediately, eliminating the possibility of quick gains (later capital gains could be made on unforeseen changes in the situation surrounding a stock, but capital losses would be just as likely as all the present relevant information would be accounted for and there would be nothing special about a stock that was not already reflected in a higher or lower price). The only reason to hold stock would be for the income derivable from dividend payments, but these are generally lower than returns available from competing instruments precisely because of the possibility of added return from capital gains (and the prospect of higher
future dividends). Worse yet, many smaller companies, or companies just starting out, or growth companies that reinvest most of their revenues, often do not pay out any dividends. Thus, new issues of such stock would suffer, and investment would drop off as companies would find it difficult to raise initial or additional capital. Only big name, established companies that could offer high enough dividends now or in the future would be able to attract much interest with a stock offering, and only stocks of these companies would be actively traded subsequent to issue. The repercussions would eventually go beyond the loss of capital investment, as brokerage and related industries would suffer.

Such a scenario will doubtless ever become reality in the extreme, as there are inefficiencies regarding the dissemination and incorporation of stock market information that will always exist, or at least be perceived to exist, to some degree. Because of these inefficiencies speculation will continue to occur, giving rise to further speculation, and leaving the "gambling" aspect of the stock market intact as a strong attraction for stock market participants. Eventually, some claim, patterns manifest themselves (resulting from market participants trying to outguess each other in repetitive ways) and become detectable to those who study market occurrences and the activities of their fellow investors most closely. Whoever is able to acquire the most knowledge and the most accurate intuition regarding the stock market will be most successful in buying and selling at the right times. Whether this or is not possible in actuality, if the perception of the possibility exists, there will continue to be interest in the equity markets and the dislocations spoken of above should
be minimized. (It is interesting to note that, at least very currently, money is going out of the stock market and into more conservative instruments like zero-coupon bonds, money market funds, etc. – perhaps the increased efficiency of the stock market is indeed beginning to frustrate investors.)

Stock price movements are predictable to the extent that the variables which lead people to create them can be discovered and quantified before they are incorporated into prices. This first chapter has considered this possibility in its cross-sectional analysis of 52 firms listed on the New York Stock Exchange for the trading day of April 23, 1984. The model, it is contended, has made some improvements upon previous ones which were either too simple or were perhaps misguided regarding the transmission process of independent variable information to price change, and thus included erroneous variables. It must be realized though that for further improvements to be implemented and the equation to become better predictively, we must make and understand choices regarding the specific market we want to study. That is, for example, we must decide whether we will examine the New York Stock Exchange or the London Stock Exchange or the Tokyo Stock Exchange, etc., because it is likely that the explanatory variables operating will differ to some degree from one market to another, and ones that do overlap will likely have different magnitudes for their regression coefficients. In fact, this is likely true over specific groups of stocks on each exchange, making it advisable to narrow our scope of
study even more.

Additionally, a time period must be chosen, one that is as close as possible to the period for which the model is intended to be used predictively.

As indicated in the Introduction, even this paper's model, which was intended to be as general as possible, really only applies to the very current market of the New York Stock Exchange for 52 randomly chosen stocks. One would have to be very cautious in extrapolating this paper's equation or regression results to other markets for stocks (or similar securities), other groups of stocks listed on the New York Stock Exchange, or to distant periods.

Of the multitude of factors considered by investors, many change over time, place, or grouping, for reasons that are not always readily comprehensible, making these changes difficult to correct for. The more general one wants his model to be, the less predictive power his equations will hold. For example, if one wants an equation which determines how any stock price on any exchange will respond to a change in per cent yield at any time, he will undoubtedly get only a very rough measure of the sensitivity of share price to changes in (yield). On the other hand, if he wants to make a prediction regarding 30 Electric Dog Polisher companies listed on the Podunk Stock Exchange for a specific time period, and he gears all his intuition in choosing variables and specifying his equation toward this one period and stock group (i.e., by including variables which might be peculiar to this group or variables that may have come into
operation only very recently) he can probably get much more accurate predictions. The problem here of course is that this equation will need to be reformulated very often to be useful over time and won't be applicable in its exact form to other groups of stocks.

So, with the choice of market to be studied made and realizations about the potential usefulness of our equations clearly in mind, improvements can be suggested with the knowledge that some may be difficult to implement and that others, if successfully implemented, maybe helpful only to the extent that they lead to a sacrifice in generality.

One potential way to improve the equation would be to seek out dummies which make qualitative distinctions other than the industry distinctions already made (perhaps in an effort to account for some of the quantitatively difficult to include variables alluded to earlier). If such variables do indeed belong in the equation, they improve the regression fit and the equation's predictive power by reducing the misspecification bias that inevitably results if they are excluded. In addition, they would not lead to a sacrifice in generality of our equation when used in multiple regression. They serve to hold each of their now included effects constant while we examine changes in each other independent variable. On the other hand, such a sacrifice in generality would occur if, for example, we actually divided our cross-section up into smaller ones, each using only stocks from given industries (as has been done). Each equation thus obtained is valid only over the industry it was calculated for (as opposed to an equation utilizing industry dummies, whose results regarding all industries are valid).
It might also be useful to try various autoregressive models, although, as said, strict random walkers would disapprove of the use of any measure of share price on the right hand side of the equation (since they believe that price in this period is not in any way correlated with price in any other period). Or perhaps an Almon (polynomial) lag could be used for some variable if it is found that cycles (such as the political business cycle) do actually exist in the stock market.

One simple way to improve the study (which was intended to take in most firm types), given the time and resources, would be to increase the number of firms under examination.

It would also be interesting to carry out tests for the stability of the regression coefficients over time and over different groups of stocks by performing several cross-sectional studies, so that we could learn just how much we are limited by our specific model regarding predictions outside of it. We could use the Chow test to see if all of the coefficients together are significantly different from one equation to the next.

Finally, it might be useful to try to conduct some time-series analyses. Many new variables would have to be included, especially macroeconomic ones. Examples would be GNP (or some personal income variable), the level of prices (as measured by the GNP deflator or the CPI), a measure of expected inflation, a measure of perceived interest rates and/or perceived expected interest rates, perhaps some money supply measure (though we might expect such a variable to be multicollinear with any of the interest rate or inflation variables), the level of the nation's
leading economic indicators (which investors are constantly made aware of by the media), the price of gold and/or other available real assets, perhaps some corporate bond rate (again, though, we would have to be wary of multicollinearity with any interest rate or very similar variable), etc. The dependent variable would probably be some broad-based market index, such as the Dow Jones 30 Industrials Average, Standard & Poor's index of 500 stocks, or the New York Stock Exchange Composite Index, any of which might also be lagged and used as an independent variable.

A time-series study has in fact been performed; this is the topic for Chapter Two. When modeling the time-series, various possible pitfalls had to be kept in mind, and considered again before any final equations were specified. For example, if we decided to use as an independent variable a measure of the nation's leading economic indicators, as suggested above, our dependent variable would probably be among these indicators; if this indeed turned out to be the case we might have to specify additional equations or else risk simultaneous equations bias. In addition, such a study would have to use much larger time periods than the single day postulated as most appropriate in my cross-sectional study to pick any significant changes over time for many of the variables mentioned (such as GNP). Thus, our data would limit our ability to use the most satisfying equation specification, since we would not be able to run regressions utilizing the most intuitively pleasing period of transmission of information from explanatory variables to dependent variable. The equation would also, as a result of the increase in period definition, necessarily go back very far in time so that enough data could be used to attain a
reasonable number of degrees of freedom and statistical confidence. It is very possible that over such an extended time period structural or psychological changes affecting some of the variables would occur, altering some of the more short-term type regression coefficients (e.g. risk, and other such cross-sectionally important variables) so that only averaged out effects of these variables might be arrived at.

At any rate, as indicated, our job is not complete. To gain a fuller understanding of the stock market we must also consider companies as a group and run some time-series regressions. After we do this we will be in a much better position to draw conclusions regarding real world stock market phenomena and the importance of various variables regarding common stock price formation.
CHAPTER TWO - THE TIME-SERIES

2.1 Introduction

The discussion of the cross-sectional analysis took place at a much more micro level than will the discussion of the time-series. The former study dealt with stocks of individual companies and their characteristics on a given trading day. Changes in the levels of the dependent variable resulted from the actions of distinct decision-makers, who responded to different levels of the independent variables among different companies on the same day. For the latter we will be interested in the response of a broad-based market index, over long periods of time, to changing levels of macroeconomic (and other macro) variables such as a measure of the nation's money supply. In this case changes in the level of the dependent variable result from the actions of a single large body of decision-makers, whose actions change with the levels of the independent variables over time.

For both analyses, however, it is an aggregate of individual and institutional investors that determine the movement of the dependent variable, and thus for both we must keep in mind their motivations and behavior patterns when modeling variables.

2.2 - The Model

The independent variables used in the time-series model were the following: 1. real personal income, 2. lagged Dow Jones Industrials Average, 3. first-differenced, logged money supply, 4. absence of risk, 5.
weighted GNP deflator, and 6. real, weighted market interest rate. The
dependent variable was the Dow Jones Industrials Average. We define the
following:

\[
\begin{align*}
\text{PI} &= \text{real personal income} \\
& \quad t-1 \\
\text{DOW} &= \text{lagged Dow Jones Industrials Average} \\
& \quad t-(1/3) \\
\text{MS} &= \text{money supply} \\
& \quad t \\
\text{AR} &= \text{absence of risk} \\
& \quad t \\
\text{def} &= \text{weighted GNP deflator (transformation to be described)} \\
& \quad t \\
i &= \text{real, weighted, market interest rate (transformation to be described)} \\
& \quad t \\
\text{DOW}_t &= \text{current Dow Jones Industrials Average} \\
& \quad t
\end{align*}
\]

Thus, the simplest version of the model is

\[
\text{DOW}_t = f(\text{PI}_{t-1}, \text{DOW}_{t-(1/3)}, \text{MS}_t, \text{AR}_t, \text{def}_t, i_t)
\]

Now, to describe the above variables:

1. Real Personal Income. This variable was chosen instead of real GNP
because it was thought to be more directly bearing upon individuals' 
financial situations (though GNP would probably have been almost as useful,
especially since institutions examine this as well as personal income 
figures), and thus to be more directly responsible for their buy/sell
decisions. GNP is a more removed measure, and is thus not as desirable.

This is also one of the few variables that was deemed to operate on 
prices more due to its economic effect than to its announcement effect
(though, of course, the economic effects and announcement effects are important for all variables). For example (as will be discussed), although changes in the money supply are obviously important in how they effect the economy and via this economic effect, in how they effect the stock market, the way this variable was modeled its announcement impact on prices is stressed more. That is, it is postulated that people react to what they hear before they actually feel it. This was thought to be partially true for other variables, such as the measure of inflation that is used.

Because the economic effect was deemed to be more important than the announcement effect this variable was lagged by one period so that by the next period, with its effects having worked throughout the economy, resulting changes in the dependent variable might be manifested. If the announcement of, e.g., a rise in personal income was all that was needed to make an impact, we wouldn't have to lag the variable as its effects would be immediate and wouldn't need to be actually felt economically by investors to cause a change in the level of the dependent variable.

2. Dow \( t-(1/3) \). The time-series is performed with quarterly data. Some measure of the dependent variable was desired on the right hand side of the equation, as it was thought that people would react, to some extent, to its past levels in buying or selling stocks. However, a full quarter - 3 months - seemed to be too long of a period over which investors might react. It seemed much more reasonable to suppose that people might react to the Dow of one month past. Therefore, Dow \( t-(1/3) \), which is a measure of Dow lagged one month (or 1/3 of a period), was used.
3. Rate of change of money supply. This variable provides an index of the current direction of Fed policy (whereas the GNP deflator, the inflation variable used, may have implications for future monetary policy since inflation is an important Fed concern). As indicated before, it was thought that this variable would achieve its major effects through announcements of changes in its levels. Later, undoubtedly, it would affect investors decisions through its effects on personal income, etc. (variables are included to take care of this), but its initial and most powerful effect would come about upon announcements such as "Volker to Loosen Reigns," etc. This variable is logged and first-differenced, to attain the desired measure of rate of change of the money supply. M2 figures were used. (John and Arthur Kraft write that "tests indicate[d] that the differences between M1 and other measures were negligible in terms of causality tests") [16].

4. Absence of risk. When discussing time-series studies that have been done in the past, Kraft and Kraft said that "It appears that further econometric studies of prices must seek out new determinants of stock prices, such as measures of risk, if those studies are to forecast variations in stock prices based on changes in significant causal variables" [17]. A measure of risk was much easier to proxy for in the cross-sectional study, because there we were required to assign some measure of it to individual companies, and we were able to compute standard deviations of individual company's share prices going back a reasonable number of periods. For the time-series, however, we do not discuss individual stocks but, rather, the market as a whole. Thus, we must find a
measure of risk valuations regarding the stock market in general (relative to the market for other investment vehicles), a much more difficult task.

Kraft and Kraft tried to proxy for risk in a later study [18] through the use of two measures – one was the ratio of the risk rate (which they proxied for with Moody's AAA corporate bond rate) to the riskless rate (which they proxied for with the U.S. government long term rate); the other was defined as the squared deviations between the risk and riskless rate of interest. This study does not quibble with their use of Moody's AAA corporate bond rate as reflecting the sum of the risk premium and the riskless rate of interest, or with their use of the U.S. government long term rates as the riskless rate. It does, however, contend their assertions that either AAA/(LT gov.) or (AAA - LT gov.) measure the risk that should be sought out in such studies as this one. The risk premium on AAA bonds is one that exists to compensate its purchasers for inflation risk, for the slight possibility of default of payment, and for market risk. The "riskless" rate of government securities is paid its purchasers to compensate them for inflation, as the probability for default on such bonds is, technically, zero. Thus, the difference between the two rates, measured as a difference or squared ratio, merely reflects how investors feel at a given time about the risk of AAA bonds resulting from the possibility of capital losses upon sale and from the possibility of default of payment relative to government bonds (cancelling out inflation risk, which both types of securities present). Thus, it says nothing directly about how investors feel about the stock market; however, the authors hoped, it would serve as a proxy for something which does (this paper will
use similar rationale in its choice of a risk variable)

Ben Branch used a mood indicator published by Barron's called the confidence index, based on interest rate differences between high and low risk bonds [19]. This index supposes that the "smart money" will move from speculative to quality bonds when the market outlook is more favorable. Thus, when the gap between high and low risk bonds is large, investors are requiring a high risk premium on the lower quality bonds indicating, supposedly, depressed conditions.

This study sought out some measure analogous to this one (though more readily available). Specifically, it sought a variable that would relate how investors feel about the risk premium they require for rate of return on stocks (R.O.R., including dividends and capital gains) to the riskless rate; in providing an indication of the compensation they require we would gain insight into how risky they feel stocks are. The desired measure would thus be (R.O.R - Gov. Securities)/R.O.R. This gives the spread between the two rates and makes the measure more general and useful for comparison by making it a per cent and correcting for different levels of return. That is, a spread of 2% - 1% would seem to indicate more risk than a spread of 15% - 14%, since in the first case investors require the rate of return on stocks to be twice that of government securities whereas in the second case they only require it to be fractionally higher. This standardization is important because the absolute level of interest rates has varied quite a bit over the time of this study.

However, quarterly data for rate of return on common stocks was not able
to be obtained (though such data was found on annual basis). Thus, a proxy had to be sought out. The only available data found that was believed to be fairly highly correlated with rate of return on common stocks was a middle grade bond rating, Moody's Baa. It was reasoned that such bonds would often move inversely with common stocks prices — when stocks were in favor, investors would move away from these bonds, and vice versa (similar to Branch's reasoning that low grade bonds would parallel stocks). Thus, I called the measure (Baa — Gov.)/Baa "absence of risk." What it really indicates is a level of attractiveness of Baa bonds relative to government bonds — the hope was that if these were favored at a given time, the spread between them and government bonds would be low (less of a premium would be required) and stocks would be more out of favor, or riskier. On the other hand, if the spread were high Baa bonds would be out of favor, and there would be an "absence of risk" (or a presence of safety from the risks cited) associated with stocks as an investment vehicle. The measure is not as desirable as one that would relate how investors feel about the risk premium they require for rate of return on stocks to the riskless rate, but seemed to be a reasonable proxy, as reasonable as Kraft and Krafts', anyway.

5. Weighted GNP deflator. This variable serves many purposes in the equation. First, it serves to hold the effects of changing prices constant by explicitly including them on the right hand side of the equation (thus, we can enter nominal values for some of the variables, such as the Dow Jones regressand). Second, it serves as a variable with an important announcement effect, as announcements regarding inflation often have a
powerful impact on the stock market. G.W. Schwert, in his study relating stock prices to inflation, writes that "it is interesting to note that the stock market seems to react at the time of the announcement of the CPI, approximately one month after the data are collected by the Bureau of Labor Statistics" [20]. He points out that "[an] important reason to expect a relationship between stock returns and unexpected inflation is that unexpected inflation contains new information about future levels of expected inflation" [21]. One reason for high expected inflation rates having a negative effect on stock prices would be that the government and/or the Federal Reserve might be expected to take actions to counteract the newly expected high levels of inflation. For example, the Fed might tighten monetary policy and send interest rates up, or the government might impose price controls, which could "distort optimal production-investment plans [and]...have a negative effect on the value of firms" [22].

For instance, it was announced on April 22, 1983, that consumer prices had risen by only 0.1 per cent in March; on the strength of this news, the Dow gained 8.03 points and closed at 1,196.30, an all-time high, as did the Nasdaq composite index of over-the-counter stocks and the market-value index for the American Stock Exchange. However, investors do not automatically revise their outlook on inflation based on the most recent government announcements. Rather, as with Friedman's permanent income hypothesis, they revise their expectations slowly over time. Because of this, and because it was reasoned that past levels of inflation would feed back and have an economic effect on the current situation as was postulated with the personal income variable, the GNP deflator was computed as a
weighted average of present and past levels of the variable. Specifically, the variable was modeled as follows:

\[
def = .45(\text{def}_t) + .35(\text{def}_{t-1}) + .20(\text{def}_{t-2})
\]

As the equation indicates, this paper is contending that revisions of the perceived current and future levels of inflation come about only slowly, with last period's level carrying almost as much weight as this period's. People get so used to a certain level of inflation when it persists that is usually the case that they are skeptical of long term improvement proclamations based on only one or a few periods. (Though it may be true that this is a relatively current phenomenon, and may not have been as operant twenty years ago).

Weighting the GNP deflator in this manner also turns it into a proxy for expected inflation, since "it is reasonable to suppose that much of the information about future inflation rates available to the market is contained in past rates of inflation." Thus, "it would seem worthwhile...to consider multiple regressions of stock returns on current and past inflation rates" [22]. Because we would expect that expected inflation would have a significant effect on stock market prices, this is another compelling reason to weight our inflation measure in this way.

The Consumer Price Index would not be as effective for the time-series here performed because in August 1971 the government imposed price controls which discouraged increases in quoted prices of goods; when price controls were phased out in 1973 and 1974, the changes in the measured price of goods in the CPI probably overstated the true inflation rate, as producers
were again able to use prices to ration their goods among customers. As a result, it is likely that "from August 1971 until sometime in late 1974 the CPI was not a good measure of the cost to consumers of obtaining goods and services" [23].

6. Real, Weighted, Market Interest Rate. Some measure of market interest rates was expected to be important for several reasons. Most obviously, they provide an indication of the rates of return that are always available to investors in alternative investment vehicles, such as Certificates of Deposit, bank money market funds, money market mutual funds, government bonds, corporate bonds, etc. High interest rates will tend to result in investors putting their money into these alternative vehicles and a fall in stock prices.

Interest rates are also important because of their effects on investment and consumption, and, as a result, their effects on the general health of the economy. When interest rates are high the user cost of capital is high and it becomes more expensive for firms to invest in capital, land, or labor. Similarly, it becomes more burdensome for consumers to finance purchases of automobiles, houses, and other durables. With investment and consumption down, national income and output suffer, corporate profits fall, dividends get cut, P/E's rise (with falling earnings), and the stock market is perceived to be a poor investment as stock prices respond negatively to all of these events.

Another reason that an interest measure might be useful in our equation stems from Irving Fisher's 1930 hypothesis which said that the nominal
interest rate fully reflects the available information concerning the possible future values of the rate of inflation. Thus, usage of such a measure could serve the above functions and at the same time act as a proxy for expected inflation (which we have already said should be an important stock price determinant) [24]. In a similar fashion, Fama argues that the short-term interest rate on treasury bills is a proxy for the market's expectation of inflation based on all information available at time t-1 [25].

There are many different rates that could justifiably be used based on the reasoning above, since there are so many different instruments on which interest is paid or on which some sort of rate of return is obtained.

The first thing that was decided regarding the measure to be used was that it should be made up of some combination of the most potentially significant ones available. The yield on U.S. government 3-month bills (new issues) was used as the primary rate because of Fama's reasoning mentioned above, and because this security serves as a popular alternative investment vehicle to stocks (which means that stock market participants are aware of its levels). In addition, because its maturity is short-term, it was thought that these bills would give a fairly accurate appraisal of current interest rates.

A measure of banks' prime lending rates were deemed to be the second most important rate, because these figures are so widely publicized by the media and in financial circles. It is very common to read that some major bank has changed its prime rate and that the stock market has reacted to
this, or that the stock market is "waiting" to see if other banks will follow suit before it reacts.

The final measure used was Moody's Aaa domestic corporate bond yield. This rate was intended to round out the interest rate measure and provide a component which gave an indication of the bond market's assessment of interest rates (also, like the T-bills already mentioned, these bonds provide an alternative investment to the stock market, making their rates important investment variables).

These components of the interest rate measure were weighted in the following manner, giving rise to an unlagged market rate of interest:

\[ i = .65(\text{Gov.}) + .20(\text{Prime}) + .15(\text{Aaa}) \]

However, it was thought that this variable should be lagged some number of periods for some of the same reasons that the GNP deflator variable was (because investors do not immediately revise their notions of expected future levels of the variable and because past levels of the variable are still effecting the economy). Thus, the values of the variable obtained above were lagged as follows:

\[ i = .7i_{t-1} + .2i_{t-2} + .1 \]

As the numbers indicate, it was postulated that investors revise their expectations regarding interest rates fairly rapidly (more rapidly than with inflation) and that previous periods' levels become obsolete rather quickly.

As mentioned, the dependent variable decided upon was the Dow Jones
Industrials Average, which is an index of thirty blue-chip industrial stocks listed on the New York Stock Exchange. The present Dow Jones Average began October 1, 1928 [26], when the list was expanded to 30 from 20 and several substitutions were made. The measure is no longer an average, as it was originally — adding up the prices of the thirty stocks comprising it and dividing by thirty would give an actual average of about $70 per share. The fact that the market closed recently at nearly 1200 (and has been as high as 1287) has been made possible by continuing decreases in the divisor, which are made to keep the measure, which has been more aptly described as a market movement indicator, undistorted by stock splits. If the divisor was not diminished to compensate for stock splits, the level of this indicator would fall for artificial reasons, and not because of a decline in the value of the holdings of investors [27].

The main reason it was chosen was because it is the most widely examined stock market barometer, and because data was readily available for it back to the first year of the study (1948). Some of the more broad-based stock market indices, on the other hand, have come into use only recently.

Eugene Fama notes in his article "The Behavior of Stock Market Prices" the fact that Dow companies, being the largest and most important in their fields, do not represent a random sample of stocks from the New York Stock Exchange; if the behavior of these stocks differs consistently from that of the majority of stocks traded, the empirical results obtained could only be strictly applied to similar company types. However, Fama goes on to point out that "the sample of stocks is conservative from the point of view of the Mandelbrot hypothesis, since blue chips are probably more stable than
other securities" [28].

In any event, for the most part, it probably makes little difference which of the many market indices are used. Ben Branch writes in his paper "the Predictive Power of Stock Market Indicators" that "the Dow Jones Industrial Average, Standard and Poor's 500 and the NYSE composite index are all reasonable choices for the index" [29]. He tried the first two and got almost identical results with both.

While some authors have supported using first differences and/or log transformations on the dependent variable [30], Kraft and Kraft have found that using the level of stock prices (i.e., the absolute level of the Dow Jones Average) or the percentage rate of change of stock prices (first differenced logs of Dow Jones values) led to negligible differences in the results [31]. The former measure (untransformed values of the variable) is used in this study.

The main results were expected to be as follows: 1. real personal income - positive sign, strong significance; 2. lagged Dow - positive sign, moderate significance; 3. money supply - negative sign, significant; 4. absence of risk - positive, uncertain significance; 5. weighted GNP deflator - negative, significant; and 6. real weighted market interest rate - negative, strongly significant.

A number of alternative results were deemed possible. Although it was postulated that the money supply coefficient would have a negative sign because of its announcement effect (a rise in the money supply this period would forebode monetary restriction in the next and lower stock prices), it
is conceivable that the previous period's effect would also be reflected by the measure. This is because the money supply variable is first differenced, and levels of MS - MS would correspond to levels of DOW t t-1 t in the regression. If last period was characterized by monetary growth, the fiscal stimulus would lead to beneficial economic effects this period and tend to raise prices. (It might also make it more likely that contractionary policy was undertaken in response leading to a positive announcement effect this period). Thus, the uncertain effects of first differencing make the expected sign of this variable ambiguous.

The expected sign of the weighted GNP deflator also seemed to be ambiguous. The announcement effect would seem to make it negative: if higher inflation is announced, prices on the stock market would be expected to fall in the short term in response to this negative economic news. However, the more long term effects of this variable would seem to make it positive. Over time, as the general level of prices in the economy rises, it would be expected that stocks, which are equity instruments and represent real assets which increase in nominal value when prices rise, should appreciate in value. Since the inflation measure has been weighted to some degree, it seems unclear which effect will dominate and the sign of this variable's coefficient becomes harder to predict. It would seem that the more weight that is placed on previous periods and the stronger is the long term effect of inflation on asset prices (as opposed to the short term capital loss effect), the more this variable will tend to be positive.
2.3 Methods and Estimation Procedure

As with the cross-section, ordinary least squares was the method of choice for use with the single equation model postulated. Quarterly data was used for all variables, from 1949 through the second quarter of 1983. Closing levels of the Dow Jones Industrial Average at the end of the middle month of each quarter were used for the dependent variable, while closing levels of the Dow at the end of first month of each quarter were used for the lagged Dow independent variable. Values for all other variables whose data was given on a monthly basis was taken from the middle of the quarter. The interest rate variable and the inflation rate variable were transformed as described in section 2.2. Real personal income was lagged one period, and the money supply measure (M2) was first differenced and logged as described in the preceding section; the modeling of the "absence of risk" variable was also described in that section.

2.4 Results

Four main runs were performed and the results have been tabulated in Appendix F. The only differences among them involved the modeling of the interest rate variable. In Run A, no "market" interest rate was calculated based upon a number of different rates; rather, the 3-month Treasury bill rate was used alone, lagged the same way over time that the interest rate variable of Run D was (Run D is the accepted and final run that has been described throughout this chapter). This rate was kept nominal. Run B used the same market rate as Run D and lagged it in the same fashion but, like Run A, kept the measure in nominal terms. Run C is the same as Run B.
except it uses a longer lag structure for the interest rate variable. Run D used the real, lagged, market interest rate described in section 2.2.

Most of the results conformed with what was expected. $R^2$'s were fairly low, but this was not surprising for many of the same reasons cited regarding the cross-section described in chapter one. However, they were still high in comparison to those found in many other studies. For example, Kraft and Kraft, in their time series study of common stock price $R^2$ determinants got $R^2$'s on the order of only .05, using a different risk rate, a different measure of the money supply, a simple corporate interest rate, and Standard and Poor's Composite Stock Index for 500 stocks [32].

The $F$'s were all significant at the 95% level ($\alpha=5\%$), indicating that in all runs significantly more variation was explained by the independent variables than was left unexplained. Further, the Durbin-Watson statistics were all close to two, indicating an absence of autocorrelation among the residuals. Theory did not indicate that heteroscedasticity would be a problem, and an examination of the residuals and $STP$ seemed to bear this out (though no plots were performed). In addition, multicollinearity was searched for using the methods alluded to in section 1.4, but no extreme cases of the problem were evident.

The final four runs described above were all very similarly specified and gave very similar results. Almost all of the variable's coefficients were significant at at least the 95% level, with most of them better. Only one variable was consistently lower, and it was significant at nearly the 90% level in one of the final runs.
A few of the results were somewhat unexpected, however. For instance, it would certainly have been expected that the personal income variable would have been positive. Indeed, stock prices and income are often positively related. In retrospect, however, the counterintuitive results obtained here are not really so surprising because the length of the lead-lag relationship between the two variables probably changes over time so that using a one period lag will sometimes show negative correlation. This was probably the case at the peak of bull rally beginning in August, 1982 (i.e., that a negative correlation had existed between the personal income of one quarter and the level of the Dow Jones Industrials Average of the next). The stock market rose very rapidly (over 50%) since the beginning of that rally, in anticipation of economic recovery. However, that positive change in the stock market had not been even close to one period adjacent to a significant positive change in personal income, which was being bullishly anticipated by stock market participants. In fact, a meaningful rise in personal income did not come until many quarters after the initial market upswing. This is just another example of how changeable and difficult to predict psychology makes modeling an equation bearing out the most important stock market variables difficult, if not impossible.

A result which was only slightly unexpected was the sign of the coefficient of the money supply variable. As was described in section 2.2, this variable was modeled with its announcement effect in mind, with the notion that an announcement of high (for example) money supply this period would lead to an expectation of tightening of monetary policy and a decline in stock prices in the next. Thus, it was thought that the variable would
be negative. However, it was also pointed out that the variable was first differenced and that some of the economic effects from last period may have set in this period, making for a positive sign (monetary growth last period will, it would seem, (ceteris paribus), lead to stock market appreciation this period). Since the coefficient of the money supply variable came out significantly positive, we assume that this is what happened.

The sign of the inflation variable was similarly slightly unexpected. As described earlier, a measure of the GNP deflator with a weighted lag of three periods was used; the announcement effect, it was thought, would make its sign negative. However, it was realized that in intentionally lagging the variable some of the economic effects would be picked up tending to make the variable positive. It was deemed possible that this effect might dominate and swing the sign to positive, and it seems that this is indeed what occurred, as this variable's coefficient came out significantly positive.

Regarding the announcement effect of inflation, Schwert writes that if the stock market is efficient, inflation changes should be transmitted to stock price changes when they occur [33]. On the other hand, though, if the Bureau of Labor Statistics "provides incremental information about inflation by collating and assimilating a large sample of prices from different locations, the stock price announcement should occur in the announcement period" [34], which is about three weeks after the start of each month (for the CPI). Based on his regression results, Schwert writes that "it seems that the stock market reacts negatively to the announcement of unexpected inflation in the CPI. Apparently, the data collection process
carried out...provides the market with information which is not available from other sources, and the stock market reacts to that information" [35]. Thus, this paper's basic theory here is supported by his results. However, Schwert goes on to say that "unexpected inflation seems to have a significant impact on stock market returns for the five trading days on either side of the CPI announcement" [36]. Because the inflation variable in this paper was lagged for a period much longer than five days, it must have been picking up levels of inflation that operated primarily via their economic effects. In addition, the possibility that there could be "both leakage of information prior to the formal announcement, and an inefficient, slow response by the stock market subsequent to the announcement" [37] would serve to make the variable less positive.

The fact that the magnitude of the coefficient of the GNP deflator is so small is also relatively surprising, as the relationship between it and stock prices was postulated to be a strong one. Fama suggests that such relatively weak correlation might exist because "unexpected inflation is contemporaneously correlated with unexpected movements in important 'real' variables, such as capital expenditures or real GNP, so that the correlation between stock returns and expected inflation is spurious" [38]. Or, perhaps it is just difficult to find the true magnitudes of the relationships because of undetected multicollinearity.

The only other surprising result was the relatively low significance of all of the interest rate variables, as interest rates were considered to be a key stock market index determinant over time. It was also surprising that the short-term government rate alone performed better than the lagged
and weighted market rates, which were thought to be better and more general indicators. Perhaps the government rate worked well because of Fama's reasoning that such a rate would be indicative of future inflation (if this is indeed true, we might expect that there was some multicollinearity with the GNP deflator variable and a diminution of its true magnitude), and perhaps also because the T-bills they represent are somewhat competitive as investment instruments with stocks (making their interest rate an important one in the eyes of investors). The relative insignificance in general of these measures (only the government rate was close to 90% significance) is probably again attributable to uneven and unpredictable stock market psychology. Indeed, at the peak of the '82-'83 bull rally, it was evident that investors had become somewhat accustomed to and more tolerant of higher and higher interest rates and that, because of this, the relationship was not as clear-cut (at least at that time) as might have been supposed. Stock market prices then had been rising dramatically with nominal interest rates hovering at the 10% level, as they did during the brief but powerful rally of August, 1984 (when real rates were much higher than in 1982). A cut in the prime from the current 11.25 to 10.5% would undoubtedly be perceived as great news for the market, while at any other time in history previous to five or six years ago a 10.5% rate would have been disastrous for the market. Evidently, over time the relationship between stock prices and interest rates has changed (not in direction but in magnitude).

The other results — strong significance in all the other variables and correct signs for lagged DOW (positive) and absence of risk (positive) —
were all expected according to the theory described in section 2.2.

Regarding the dependent variable, Branch writes (as mentioned) that he obtained almost identical results in his regressions utilizing the Dow, S&P's 500, and the NYSE composite index. Thus it seemed unnecessary to do any testing with alternative market indices. (His $R^2$'s, which would be very similar to his $R^2$'s since he only uses a few variables, ranged from 0.050 to 0.399) [39].

2.5 - Conclusions and Directions for Future Research

Umstead writes the following in his article "Forecasting Stock Market Prices":

Theory tells us that stock prices are determined by expectations, which, unfortunately, are not directly measurable. However, these expectations, if rational, must be derivable from existing measures of the economy. It appears that expectations are formed in a systematic relationship to the leading elements of economic activity.... Stock prices, to some degree at least, appear to respond in a predictable manner to this...flow of information [40] (the emphasis is his).

As with the cross-sectional study of chapter one, an attempt has been made here to find the most potent proxies for the variables which form the basis of investors' expectations and decisions and which thus cause stock price movements. The results seem to confirm those of Chapter One, namely, that stock market prices may be loosely predictable using, or are at least loosely correlated with, a number of explanatory variables such as the ones herein described. However, it would be wise to keep in mind some of the
same cautions outlined at the end of the first chapter regarding scope of study and the dangers of extrapolations outside of that scope.

Also, as with the cross-sectional study, various manipulations of the equations used could be suggested, including the use of log transformations, polynomial lags, percentage changes instead of absolute levels for variables, different lag lengths and weights, additional variables (though they would likely overlap the ones already used), etc.

One specific such upgrading that might be useful would be to try different lag lengths for the GNP deflator, in order to get a better idea of the relative potencies of the announcement effect and the economic effect. The theory here was somewhat unclear, and the results only told us that with the specific lag length used the economic effect dominated. It would be interesting to know how short of a lag it would take to get this variable to swing to negative (five days would be suggested by Schwert, though he performed a different sort of analysis to obtain this result).

Similarly it would be interesting to try varying the lag lengths of DOW and the interest rate variable, and to experiment with different market rates for the interest rate variable.

Finally, it would be useful to try different period lengths (though using a period of less than one quarter for a time-series of this nature is probably not feasible, because of data unavailability), and to try regression runs for different sub-periods and going back different lengths of time to see how much the magnitudes of corresponding coefficients differ
in different times (which are subject to different groups of investors, psychologies, and macroeconomic environments).

Because the results obtained in this chapter deal with macroeconomic variables, a number of which can be affected by monetary or fiscal intervention by the Government and the Fed, some possible policy implications become evident. For example, we have found that changes in the money supply will affect stock market prices. Because the stock market affects economic activity (it acts as a source of funds, provides a breeding ground for mergers and takeovers, provides an investment vehicle for many institutions and individuals, and as a leading economic indicator affects investors' expectations regarding the economy in general), these influences of the money supply on the stock market should not be ignored. (Indeed, they are probably not, though it is doubtful that the Fed is guided by the stock market to any major degree in setting its mone supply or interest rate targets. It can have a more direct, though probably less potent, effect on the stock market by changing margin requirements).

In any event, it seems to be the case that some inefficiencies exist in the stock market over time just as they do over different companies on the same day, and that these inefficiencies may allow us to derive relationships between certain variables and prices, and, though more doubtfully, may provide for the possible exploitation of relevant information in predicting stock market prices to some small degree. While this study has shown only a loose connection between some independent variables and stock market prices, it seems to support the notion that,
eventually, more accurate results for some specific markets may be attainable.
Bibliography

General References for the Cross-Section

ARTICLES:


Griffin, Paul A., "Competitive Information in the Stock Market: An


BOOKS:


OTHER:


General References for the Time-Series

ARTICLES:


**OTHER:**


**Statistical Sources for the Cross-Section**

*The New York Times*, "Business Day" section, Tuesday through Saturday issues from Wednesday, February 17 to Saturday, April 24, 1984.]

**Statistical Sources for the Time-Series**

*Business Conditions Digest*, Series 227, Department of Commerce/Bureau of Economic Analysis and Bureau of the Census, *(Handbook of Cyclical Indicators, 1977).*


*Survey of Current Business*, volumes 30 (1950), 40 (1960), 48 (1968), 55
Notes


[7] Ibid.


[12] Ibid., p. 58.


[23] Fama and Schwert, p. 118


[29] Branch, p. 276.

- 89 -
[33] Schwert, p. 20.
[34] Ibid.
[35] Ibid., p. 25.
[36] Ibid., p. 27.
[37] Ibid.
[38] Fama and Schwert, p. 115.
## APPENDIX A

### Cross-Section Regression Results

<table>
<thead>
<tr>
<th>RUN</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5 original variables</td>
<td>Run A, with logs of v, y, R</td>
<td>Run B plus the 3 dummies</td>
<td>Run C plus the 2 interaction terms</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.3577</td>
<td>.4260</td>
<td>.5124</td>
<td>.5684</td>
</tr>
<tr>
<td>S.E.E.</td>
<td>1.9007</td>
<td>1.18506</td>
<td>1.9303</td>
<td>7.71301</td>
</tr>
<tr>
<td>Durbin-Watson stat.</td>
<td>1.907</td>
<td>2.067</td>
<td>2.005</td>
<td>1.972</td>
</tr>
<tr>
<td>F</td>
<td>5.118</td>
<td>6.797</td>
<td>6.605</td>
<td>5.399</td>
</tr>
<tr>
<td>degrees of freedom</td>
<td>4.47</td>
<td>4.47</td>
<td>6.42</td>
<td>9.42</td>
</tr>
<tr>
<td>significance level</td>
<td>99%</td>
<td>99.4%</td>
<td>99.9%</td>
<td>99.9%</td>
</tr>
<tr>
<td>$\hat{c}$</td>
<td>5.345</td>
<td>3.5227</td>
<td>3.753</td>
<td>1.2476</td>
</tr>
<tr>
<td>$t$ for $\hat{c}$</td>
<td>1.0522</td>
<td>1.1525</td>
<td>1.4940</td>
<td>2.2566</td>
</tr>
<tr>
<td>1st significance level, $\alpha$</td>
<td>.05%</td>
<td>90%</td>
<td>90%</td>
<td>97.5%</td>
</tr>
<tr>
<td>$t$ for $\hat{c}/\sigma(\hat{c})$</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>$\hat{c}$: P/E ratio</td>
<td>-1.497</td>
<td>-1.567</td>
<td>-3.635</td>
<td>-3.217</td>
</tr>
<tr>
<td>$t$ for $\hat{c}$</td>
<td>-1.497</td>
<td>-1.154</td>
<td>-1.1250</td>
<td>-2.134</td>
</tr>
<tr>
<td>1 - $\alpha$</td>
<td>99%</td>
<td>97.0%</td>
<td>99.9%</td>
<td>97.5%</td>
</tr>
<tr>
<td>$\hat{c}_i$: P/E</td>
<td>5.5619</td>
<td>6.2970</td>
<td>7.1477</td>
<td>9.054</td>
</tr>
<tr>
<td>$t$ for $\hat{c}_i$</td>
<td>4.5269</td>
<td>4.750</td>
<td>5.530</td>
<td>4.958</td>
</tr>
<tr>
<td>1 - $\alpha$</td>
<td>99.95%</td>
<td>99.95%</td>
<td>99.95%</td>
<td>99.95%</td>
</tr>
<tr>
<td>$\hat{c}_i$: P/E on (d/2)</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>$t$ for $\hat{c}_i/(\hat{d}/2)$</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>1 - $\alpha$</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>$\hat{c}_b$: volume</td>
<td>0.849</td>
<td>2.190</td>
<td>2.829</td>
<td>0.004</td>
</tr>
<tr>
<td>$t$ for $\hat{c}_b$</td>
<td>2.5564</td>
<td>3.7114</td>
<td>3.9714</td>
<td>2.500</td>
</tr>
<tr>
<td>1 - $\alpha$</td>
<td>99%</td>
<td>99.95%</td>
<td>99.95%</td>
<td>97.5%</td>
</tr>
<tr>
<td>$\hat{c}_b$: risk</td>
<td>-4.980</td>
<td>-5.1536</td>
<td>-5.094</td>
<td>-2.443</td>
</tr>
<tr>
<td>$t$ for $\hat{c}_b$</td>
<td>-3.2118</td>
<td>-3.043</td>
<td>-3.358</td>
<td>-1.601</td>
</tr>
<tr>
<td>1 - $\alpha$</td>
<td>97.5%</td>
<td>99.9%</td>
<td>99.9%</td>
<td>99.9%</td>
</tr>
<tr>
<td>$\hat{c}_b$: P/E on (d/2)</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>$t$ for $\hat{c}_b/(\hat{d}/2)$</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>1 - $\alpha$</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>$\hat{c}_b$: chemicals dummy</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>$t$ for $\hat{c}_b$</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>1 - $\alpha$</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>$\hat{c}_b$: petroleum dummy</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>$t$ for $\hat{c}_b$</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>1 - $\alpha$</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
</tbody>
</table>

(continued on next page)
<table>
<thead>
<tr>
<th>St. Nature</th>
<th>G91</th>
<th>G92</th>
<th>1-α</th>
<th>1.674 * *</th>
<th>1.962 * *</th>
<th>7.2157</th>
</tr>
</thead>
<tbody>
<tr>
<td>For G92</td>
<td>N.A.</td>
<td>N.A.</td>
<td>1.674 * *</td>
<td>1.962 * *</td>
<td>7.2157</td>
<td></td>
</tr>
<tr>
<td>1-α</td>
<td>N.A.</td>
<td>N.A.</td>
<td>1.674 * *</td>
<td>1.962 * *</td>
<td>7.2157</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>G91 yield &amp; risk</th>
<th>N.A.</th>
<th>N.A.</th>
<th>1.674 * *</th>
<th>1.962 * *</th>
<th>7.2157</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-α</td>
<td>N.A.</td>
<td>N.A.</td>
<td>1.674 * *</td>
<td>1.962 * *</td>
<td>7.2157</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>G90 1.674 * * &amp; risk</th>
<th>N.A.</th>
<th>N.A.</th>
<th>N.A.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-α</td>
<td>N.A.</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
</tbody>
</table>

* don't really care about - superseded by unconstrained data
** importance of these numbers is ambiguous (see text)
N.A. not applicable

[NOTE: 1-α really means 100(1-α)x ]
APPENDIX B - Interpolation

For t-ratios the null hypothesis was $H_0: b_i = 0$; I used 40 d.f. (degrees of freedom), to be conservative. For *1, *2, *3, *4 calculations I interpolated to the $t_{.05}$ and $t_{.10}$ values associated with 47 d.f. (since these were less precise than the acceptable level of 1- = .90, I wanted to find out by precisely how much. So after finding out the precise level of 47 d.f. associated with the t-boundaries of $t_{.05}$ and $t_{.10}$, I interpolated to find the precise t-value at 47 d.f. associated with $b_i$ for Run A, $b_k$ for Run B, $b_5$ for Run C, and $b_{10}$ for Run D).

For *1, *2:

\[ d.f. = n-k = 52-5 = 47 \]

\[ (n-k = \text{number of constants on right-hand side of equation less 1, or number of independent variables [for RSS]} \]

<table>
<thead>
<tr>
<th>d.f.</th>
<th>$t_{.05}$</th>
<th>$t_{.10}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>.681</td>
<td>1.303</td>
</tr>
<tr>
<td>47</td>
<td>.679</td>
<td>1.303</td>
</tr>
<tr>
<td>60</td>
<td>.679</td>
<td>1.296</td>
</tr>
</tbody>
</table>

\[ .679 + (.681-.679)(\frac{4.7-60}{40-60}) = .680 \]

\[ 1.296 + (1.303-1.296)(\frac{.03-60}{40-60}) = 1.301 \]

*1: \[ \frac{t}{.680} \]

\[ \begin{array}{c} \alpha \\frac{1.0322}{.10} \end{array} \]

\[ = .10 + (.25-.10)(\frac{1.0322-1.301}{.680-1.301}) = .165 = \alpha. \]

Thus, we can reject $H_0$ with a prob value of .167, or at a level of significance of 83.5% (which is 1-$\alpha$).

*2: \[ \frac{t}{.680} \]

\[ \begin{array}{c} \alpha \\frac{1.1754}{.10} \end{array} \]

\[ = .130 \Rightarrow 1-\alpha = 87.0\% \]

(calculating as above)

[NOTE: 1-$\alpha$ is actually 100(1-$\alpha$); $\alpha$ = level of significance of test, and .10 is my cut-off for acceptability, or at least the point where I decide to interpolate to gain further precision; $\alpha$ = .05 oftentimes gives the level of acceptability for similar studies.]
For **3, 4, 5:**

\[
\text{d.f.} = 52 - 10 = 42
\]

<table>
<thead>
<tr>
<th>d.f.</th>
<th>( t_{.25} )</th>
<th>( t_{.00} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>1.681</td>
<td>1.303</td>
</tr>
<tr>
<td>42</td>
<td>1.681</td>
<td>1.303</td>
</tr>
<tr>
<td>60</td>
<td>1.681</td>
<td>1.303</td>
</tr>
</tbody>
</table>

(Using same method of calculation as on previous page).

\[
\text{d.f.} = 40
\]

\[
\text{d.f.} = 60
\]

\[
\text{d.f.} = 42
\]

\[
\geq 1,302 \rightarrow = 1.302
\]

\[
\geq 1.302 \rightarrow = 77.12\%
\]

(as said, though, this figure is unimportant because it only represents part of the unconstrained risk coefficient, \( \alpha \). The calculation for the term as a whole can be found in the text with the other unconstrained coefficient results.)

\[
\text{d.f.} = 1.302
\]

\[
\text{d.f.} = 1.302
\]

\[
\text{d.f.} = 1.302
\]

\[
\geq 1.302 \rightarrow = 87.60\%
\]

\[
\geq 1.302 \rightarrow = 82.8\%
\]

\[
\geq 1.302 \rightarrow = 82.8\% \quad (\alpha = .172)
\]
## APPENDIX C - Names of Companies and their Industry Classifications

<table>
<thead>
<tr>
<th>No.</th>
<th>Company</th>
<th>1. AAR Corp.</th>
<th>2. ACF Indus.</th>
<th>3. AMF, Inc.</th>
<th>4. ARA Services</th>
<th>5. Abbott Labs</th>
<th>6. AMP-CIEVEC</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.</td>
<td>Ace Electric</td>
<td>Fuel, cars, eq. transformers</td>
<td>Food and drug, food super markets</td>
<td>Adams- mills</td>
<td>Diversified energy &amp; mfg. co., nat. gas, oil, petro, agric. indus., air, railroad, house, boats, etc.</td>
<td>10. Astra Life-Care</td>
<td>11. Air Prod. inc.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Oil + gas, Chemicals, fibers, plastics, etc.</td>
<td>Oil + gas, Chemicals, fibers, plastics, etc.</td>
<td>21. Alcoa</td>
<td>Electric, utility &amp; public utility system</td>
<td>23. Alcoa-Pul</td>
<td>Auto accessories, testing eq.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

continued on next page
Industry Classification Assignments
(The numbers refer to those next to the companies on the preceding pages)

1. Agricultural Implements: 15, 22
2. Air Conditioning: 37
3. Aircraft-Manufacturing: 1, 15
4. Air Freight and Freight Forwarding: 18, 21, 12
5. Aluminum: 18, 30, 32
6. Amusements and Amusement Supplies: 5, 24, 30, 36
7. Apparel and Clothing Manufacturers: 3
8. Apparel and Hosiery: 9
10. Automobiles: 3
11. Automotive Parts and Accessories: 2, 19, 24, 25
12. Building Materials and Equipment: 25, 30, 37, 41
13. Can Manufacturing: 39
14. Candy-Confectionery-Gum: 49
15. Chemicals: 5, 11, 13, 19, 25, 30, 32, 33, 37, 39, 41
16. Coal and Coke: 19, 25, 30, 32
17. Construction (Heavy) and Supplies: 15, 41, 48
18. Distillers: 19
19. Drugs-Medicines-Cosmetics: 5, 7, 41, 49
20. Electrical Equipment and Supplies: 13, 30, 42, 3, 9
21. Electronics: 1, 15, 19, 24, 42, 3, 7
22. Engineering Services: 15
23. Fertilizer and Fertilizer Materials: 25, 41
24. Food – Miscellaneous Products: 4, 13, 49
25. Furniture and Fixtures: 50
26. Grocery and Food Chains: 17
27. Hardware and Supplies: 33
28. Hospital, Medical and Dental Equipment and Supplies: 5, 41, 50, 51
29. Household Appliances and Utensils: 30, 49
30. Leather: 13
31. Lumber and Wood Products: 39
32. Machine Tools: 6
33. Machinery: 19, 22, 48, 3
34. Maintenance Service (Business): 26, 37
35. Management Services: 21
36. Marketing Research: 39
37. Metal Products – Miscellaneous: 2, 18, 19, 22, 30, 33
38. Mining and Processing: 32, 13, 18, 30
39. Motion Pictures and Theaters: 36
40. Office Equipment and Supplies: 38
41. Oil Royalty: 11
42. Oil Service: 30
43. Paints, Paint Materials and Lacquers: 49
44. Paper – Other Paper and Board: 39
45. Paper – Containers, Boxes, Cartons: 39
46. Paper—Miscellaneous: 39
47. Parking Lots: 29
48. Petroleum Producing, Refining, Transporting and Distributing: 25, 30, 32, 34
49. Plant Equipment: 2, 11, 15, 48
24. Food - Miscellaneous Products: 4, 13, 49
25. Furniture and Fixtures: 50
26. Grocery and Food Chains: 17
27. Hardware and Supplies: 33
28. Hospital, Medical and Dental Equipment and Supplies: 5, 41, 50, 51
29. Household Appliances and Utensils: 30, 49
30. Leather: 13
31. Lumber and Wood Products: 39
32. Machine Tools: 6
33. Machinery: 19, 22, 48, 3
34. Maintenance Service (Business): 26, 37
35. Management Services: 21
36. Marketing Research: 39
37. Metal Products - Miscellaneous: 2, 18, 19, 22, 30, 33
38. Mining and Processing: 32, 13, 18, 30
39. Motion Pictures and Theaters: 36
40. Office Equipment and Supplies: 38
41. Oil Royalty: 11
42. Oil Service: 30
43. Paints, Paint Materials and Lacquers: 49
44. Paper - Other Paper and Board: 39
45. Paper - Containers, Boxes, Cartons: 39
46. Paper - Miscellaneous: 39
47. Parking Lots: 29
48. Petroleum Producing, Refining, Transporting and Distributing: 25, 30, 32, 34
49. Plant Equipment: 2, 11, 15, 48
50. Plastics and Plastic Materials: 5, 2, 11, 13, 14, 25, 33, 3, 9, 41
51. Pollution Control Equipment: 11
52. Printing and Engraving: 38, 39
53. Publishing-Magazines: 39
54. Publishing-Newspapers: 36
55. Radio and Television - Broadcasting: 36
56. Railway Equipment: 2, 30
57. Real Estate: 15, 30, 35
58. Restaurants and Confectioneries: 35, 4
59. Retail Stores - Apparel and Clothing: 27
60. Shipbuilding: 15, 3
61. Steel and Iron: 21, 22, 30
62. Sugar Producing - U.S. Continental: 31
63. Sulphur: 15
64. Textiles - Synthetic Yarns: 13, 25
65. Textiles - Synthetic and Silk Fabrics: 41, 16
66. Tires, Rubber and Rubber Goods: 5, 19, 33
67. Tobacco Products: 3
68. Visual Aids: 51
69. Wholesale, Distributors and Jobbers: 17, 18, 50, 4
70. Vending Machines: 4
71. Retail Stores - Drug: 8
72. Insurance: 10, 22, 44, 45, 46, 47
73. Electric Utility: 23, 43
74. Gas Utility - Natural Gas: 14, 25
75. Broadcasting: 36, 45
76. Investment Service: 40, 44, 45

77. Travel: 44, 45

[NOTE: Many of these assignments were found already made for the companies in Moody's Industrial Manual, Vol. 1.]
PROCESS_GENERATING_APPENDIX_C:

I used industry classifications found in Moody's Industrial Manual (vol. 1) and the Dache Stock Guide to generate the first two pages of this appendix. Then, using a conglomerate of the industry classifications found in the above volumes, and the information written down on pages 95 and 96, I made the industry classification assignments that appear on pages 97, 98, 99, and 100.

The ultimate narrowing of this list down to three useful categories is described in the text (these three classifications are set off in the foregoing list by asterisks).
### Appendix D - Covariance Matrix, Run D (Cross-Section)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.27E+17</td>
<td>1.38E-16</td>
<td>-1.37E+19</td>
<td>1.85E+10</td>
<td>1.10E+17</td>
<td>-1.27E+16</td>
<td>-1.26E-16</td>
<td>1.10E+01</td>
<td>5.20E-17</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>6.00E+00</td>
<td>5.77E-10</td>
<td>1.71E+00</td>
<td>1.00E+00</td>
<td>1.13E+17</td>
<td>-1.87E+16</td>
<td>-1.00E+01</td>
<td>5.20E-17</td>
<td>1.00E+01</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>5.00E+00</td>
<td>5.00E+00</td>
<td>1.00E+00</td>
<td>1.00E+00</td>
<td>1.00E+00</td>
<td>1.00E-15</td>
<td>1.00E-15</td>
<td>1.00E+00</td>
<td>1.00E+00</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1.00E+17</td>
<td>1.00E+17</td>
<td>1.00E+17</td>
<td>1.00E+17</td>
<td>1.00E+17</td>
<td>1.00E+17</td>
<td>1.00E+17</td>
<td>1.00E+17</td>
<td>1.00E+17</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1.00E+00</td>
<td>1.00E+00</td>
<td>1.00E+00</td>
<td>1.00E+00</td>
<td>1.00E+00</td>
<td>1.00E+00</td>
<td>1.00E+00</td>
<td>1.00E+00</td>
<td>1.00E+00</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>5.20E-17</td>
<td>5.20E-17</td>
<td>5.20E-17</td>
<td>5.20E-17</td>
<td>5.20E-17</td>
<td>5.20E-17</td>
<td>5.20E-17</td>
<td>5.20E-17</td>
<td>5.20E-17</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1.00E+01</td>
<td>1.00E+01</td>
<td>1.00E+01</td>
<td>1.00E+01</td>
<td>1.00E+01</td>
<td>1.00E+01</td>
<td>1.00E+01</td>
<td>1.00E+01</td>
<td>1.00E+01</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>5.20E-17</td>
<td>5.20E-17</td>
<td>5.20E-17</td>
<td>5.20E-17</td>
<td>5.20E-17</td>
<td>5.20E-17</td>
<td>5.20E-17</td>
<td>5.20E-17</td>
<td>5.20E-17</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1.00E+01</td>
<td>1.00E+01</td>
<td>1.00E+01</td>
<td>1.00E+01</td>
<td>1.00E+01</td>
<td>1.00E+01</td>
<td>1.00E+01</td>
<td>1.00E+01</td>
<td>1.00E+01</td>
<td></td>
</tr>
</tbody>
</table>

### Variances and Covariances of B-Coefficients

- **On the Diagonal:**
  - Terms: Covariances of row-numbered B(i) and column-numbered B(j) terms is the highest-numbered coefficient, row, and column.

<table>
<thead>
<tr>
<th>i</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.25E+00</td>
<td>.52E-01</td>
<td>.19E-02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.52E-02</td>
<td>.84E-01</td>
<td>.12E-03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.66E-01</td>
<td>.12E-03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>.22E-02</td>
<td>.22E-02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>.11E-01</td>
<td>.18E-04</td>
<td>.21E-02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>.21E-04</td>
<td>.21E-01</td>
<td>.13E-03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>.49E-02</td>
<td>.93E-01</td>
<td>.95E-03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>.11E-01</td>
<td>.11E-01</td>
<td>.11E-01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>.11E-01</td>
<td>.11E-01</td>
<td>.11E-01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>.11E-01</td>
<td>.11E-01</td>
<td>.11E-01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Residuals

<table>
<thead>
<tr>
<th>Y(1)</th>
<th>ESTIMATED Y(1)</th>
<th>RESIDUALS</th>
</tr>
</thead>
<tbody>
<tr>
<td>2020</td>
<td>96478.0</td>
<td>-17422.2</td>
</tr>
<tr>
<td></td>
<td>3.81034</td>
<td>2.173666</td>
</tr>
<tr>
<td></td>
<td>-2.78454</td>
<td>4.83954</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.24027</td>
</tr>
</tbody>
</table>
For the cross-section (see text for description):

\[ \text{STP} = \left( \begin{array}{c} P \end{array} \right) \left( \begin{array}{c} -P/s \end{array} \right) = b(i) + b(y) + b(P/E) + b(d/P) + b(V) + b(R) + b(YR) + b(dR) + b(C) + b(P) + b(NG) + b(e) \left( E^{-1} \right) i, t-1 + b(e) \left( E^{-1} \right) i, t-1 \]

For the time-series:

The only differences among the alternative time-series formulations were the modelling of the interest rate variables. These differences are described in the text and are given in Appendix F.
## APPENDIX F

### Time-Series Regression Results

<table>
<thead>
<tr>
<th>RUN</th>
<th>MODEL</th>
<th>( R^2 )</th>
<th>S.E.E.</th>
<th>D.W.</th>
<th>F</th>
<th>Degrees of Freedom</th>
<th>Significance Level</th>
<th>( T ) for ( b_j )</th>
<th>1-( \alpha )</th>
<th>( T ) for ( b_m )</th>
<th>1-( \alpha )</th>
<th>( T ) for ( b_{\text{rate}} )</th>
<th>1-( \alpha )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>5 independent variables plus the 3-month government T-bill rate</td>
<td>.3103</td>
<td>.0548</td>
<td>1.588</td>
<td>5.94</td>
<td>6.131</td>
<td>95%</td>
<td>95%</td>
<td>95%</td>
<td>95%</td>
<td>95%</td>
<td>95%</td>
<td></td>
</tr>
</tbody>
</table>

\( b_j \): Real PPI
\( d.f. \): 13
\( T \) for \( b_j \): .13
1-\( \alpha \): 97.5%

\( b_m \): Dow Down
\( d.f. \): 131
\( T \) for \( b_m \): 1.892
1-\( \alpha \): 95%

\( b_{\text{rate}} \): interest rate
\( d.f. \): 131
\( T \) for \( b_{\text{rate}} \): 2.304
1-\( \alpha \): 99.5%

\( b_{\text{asset of risk}} \): asset of risk
\( d.f. \): 131
\( T \) for \( b_{\text{asset of risk}} \): 2.054
1-\( \alpha \): 99.5%

\( b_{\text{GNT deflator}} \): GNT deflator
\( d.f. \): 131
\( T \) for \( b_{\text{GNT deflator}} \): 1.900
1-\( \alpha \): 99.95%

\( b_{\text{rate variable}} \): rate variable
\( d.f. \): 131
\( T \) for \( b_{\text{rate variable}} \): 1.559
1-\( \alpha \): 95%

\( 1-\alpha \) really means \( 100(1-\alpha)\% \)
Appendix G - "No heteroscedasticity plot"