Aaron Chaum
Bryan Price

What Affects Long Term Stock Prices?

Introduction

Numerous academic studies have attempted to explain movements in stock prices with well-known macroeconomic variables such as dividends, interest rates, market capitalization and beta. To varying degrees, these studies have found these variables to have ambiguous, if not statistically insignificant effects on stock prices. We propose to examine the relationship between dividends, interest rates and market capitalization and their predictive power on capital gains with regression analysis without adjusting for risk. Our hypothesis is that the recent tax environment has fundamentally changed the relationship between capital gains and our explanatory variables, and our regression will yield statistically significant results.

Literature Review

The theory of the impact of dividends on future stock prices goes back to the dividend irrelevance theorem of Miller and Modigliani (1961), which implies that dividends have no overall effect on stock returns because price appreciation has a compensating effect on the dividend distribution. Early studies of the dividend irrelevance theorem seemed to confirm Miller's and Modigliani's research. Using data from the period before the seventies, studies such as Friend and Puckett (1964) and Black and Scholes (1974) were unable to find any significant relationship between dividend policy and total returns on a risk-adjusted basis. However, some researchers have continued to look at the effect of the distribution of cash on investors' preferences, and recent research has focused on the effect of dividend yields on future stock returns. Rozeff (1984) finds that the ratio of the dividend yield to the short-term interest rate explains a significant fraction of movements in annual stock returns. Fama and French (1988) use a regression framework to show that the dividend yield predicts a significant portion of multiple year returns to the NYSE index. They further observe that the explanatory power of the dividend yield increases in the time horizon of the returns. Similar results are reported by Flood, Hodrick, and Kaplan (1987) and

Cambell and Shiller (1988). Lewellen (2001) has recently developed an alternative approach to assessing statistical significance in predictive regressions and finds strong evidence of predictability for the dividend yield.

Many of these forecasting studies use ordinary least squares (OLS) regression. Fama and French (1988) provide the strongest evidence in support of the dividend yield effect by using overlapping multiple-year horizon returns. They investigate the ability of dividend yields to predict compound returns on value-weighted and equal-weighted NYSE portfolios for intervals between one month and four years. Fama and French use CRSP monthly data, beginning in 1926, and they construct annualized dividend yields by summing the previous 12 months of dividends. The OLS specification of Fama and French (1988) is the following:

$$\ln(R_{t+k,k}) = \alpha_{k,1} + \beta_{k,1}(D_t/P_t) + \mu_{t+k,k}$$

where $\ln(R_{t+k,k}) + \ldots + \ln(R_{t+k})$ is the continuously compounded k-period rate of return. The error term $\mu_{t+k,k}$ is an element of the time t+k information set, and Fama and French define the one-period real return as $R_{t+1} = (P_{t+1} + d_{t+1})/P_t$ where P_t is the end of the month real stock price and d_t is real dividends paid during month t and the annualized dividend yield is D_t/P_t. Fama and French (1988) found that the coefficient of the dividend yield was positive and significant, and they also observed that the explanatory power of the dividend yield increases in the time horizon of the returns; over 4-year horizons, R^2's reach a high value of 64%. However, Fama and French (1988) point out many biases which will distort the findings of their research. Most notable is the fact that "dividend yields contain forecasts of future returns and dividend growth, which may bias downward the regression coefficient in the dividend yield regression."

Campbell and Shiller (1988) use a stochastic approximation to the dividend discount model and estimate the model in a VAR framework. Their method is to test a dividend-ratio model relating the dividend-price ratio D/P to the expected future values of the one-period rates of discount r and one-period growth rates of dividends g over succeeding periods. The model is based on the Gordon growth model, $D/P = r - g$, which was derived under the assumption that dividends will grow at a constant rate forever, and that the discount rate will never change. The benefit of this assumption is that it permits analysis of the variation through time in the dividend-price ratio in relation to predictable changes in discount rates and dividend growth rates. This is different from most previous studies of the dividend-price ratio that have been concerned with the cross-sectional relationship between dividend-price ratios and average returns [e.g. Black and Scholes (1974)], which relies for the most part on the assumption that discount rates are constant. Campbell and Shiller (1988) use a vector auto-regression to break down movements in the log dividend-price ratio into components attributable to expected future dividend growth, expected future discount rates, and other unexplained factors. Campbell and Shiller found some evidence that the log dividend-price ratio does move with rationally expected future growth in dividends. Their findings also parallel that of Fama and French (1988) and

Flood, Hodrick, and Kaplan (1986), which all find that stock returns are more highly predictable when measured over a long period of time. They also find that the moderate predictability of one-year stock returns can have dramatic implications for the log dividend-price ratio, and in particular, the log dividend-price ratio has a standard deviation that is at least 50 percent higher than it would be if stock returns were unpredictable.

Blume (1980) uses a simple regression equation to examine the relationship between dividends and stock prices, which includes the anticipated dividend yields:

$$R_{it} = a_t + b_t\beta_{it} + c_t\delta_{it} + \varepsilon_{it},$$

where R_{it} is the total realized return on stock i over period t, β_{it} is the relevant beta coefficient, δ_{it} is the dividend yield anticipated over period t, and ε_{it} is a mean-zero disturbance uncorrelated with β_{it} and δ_{it}. The coefficients a_t, b_t, and c_t are allowed to vary randomly from one period to the next and the null hypothesis is that expected returns are unrelated to anticipated dividend yields (c=o). Blume uses the regression equation to estimate on a specific cross section and then again on another cross section and so on. Blume is different from previous studies because he calculates an alternative measure of the dividend yield, as the ratio of the dividends paid over the previous twelve months to the price at the beginning of these twelve months. This could be more accurate than previous studies' dividend yield, such as Black-Scholes, if companies tended to adjust dividend levels quickly to maintain a fixed payout ratio and the price-earnings ratio was relatively stable. Blume further adjusts the ratio for what he calls "market-wide changes (general market movements) in the level of dividend yields over the prior 12 months. He does this by assuming that the beginning-of-period price had increased or decreased the same percentage as the market average over these 12 months, and then he uses this adjusted price in the calculation of the dividend price ratio. The first cross-sectional regression is estimated using the quarterly total returns for the first calendar quarter of 1936 and he uses data prior to 1936 to estimate β_{it} and δ_{it}. Blume then performs the second cross-sectional regression using returns for the second calendar quarter of 1936 and this process is repeated for each and every calendar quarter through the end of 1976 for a total of 164 quarterly cross-sectional regressions.

Blume breaks the securities into groups in order to reduce the magnitude of measurement error and the impact of "non-stationarities" upon the estimates of β_t and δ_{it}. The simple averages of the r_{it}'s, β_t's and δ_{it}'s within each group are used as the basic data in estimating the cross-sectional regressions. For the grouping, all NYSE securities with complete data from 1926 through March of 1936 were ranked from low to high by their beta coefficients as estimated over the previous 5 years using the S&P Composite Index. These ranked securities are then partitioned into five groups of an equal number of securities, and thus the first group contains those securities with the smallest estimates of beta and the last group those with the largest estimates. Blume further divides each of the five groups of securities into five subgroups of roughly equal numbers of securities according to the adjusted 1935 dividend yield. Over the 1936-76 period, Blume finds that these cross-sectional regressions reveal a positive and significant relationship on average between the quarterly realized rates of return and both the beta coefficients and the anticipated quarterly dividend yields. For the portfolios grouped first on beta and then on dividend yield, the average coefficient on dividend yield for the 164 cross-sectional regressions estimated over these 41 years was 0.5232 with a t-value of 2.07, and the average coefficient on beta was 0.0215 with a t-value of 2.34. Taken at face value, these numbers imply that for a given beta, realized quarterly returns increased an average of 0.52% for each 1% increase in the anticipated quarterly dividend yield. Blume also found that the significance of the dividend yield variable varies over time. In the two decades from 1937 through 1946 and from 1957 through 1966, the average coefficients on dividend yield were positive but not significant at any usual level of significance. But in the decade from 1947 through 1956, the average coefficient on dividend yield was 0.8743 with a highly significant t-value of 4.27. In the last decade ending in 1976, the average coefficient on dividend yield was 1.1272 with a t-value of 1.93, indicating a level of significance between 5% and 10%.

Goetzmann and Jorion (1995) extend the analysis of dividend yield regressions on monthly U.S. data going back to 1871. They find little evidence of predictability over the whole sample period. When considering different sample periods, however, they find that the predictability results do not extend to pre-1926 U.S. data. This is because stock returns before 1926 were more stable, and in turn means that dividend yields were more stable and therefore with little variability there is little predictability. The Goetzmann and Jorion study also points out the survivorship bias inherent in this type of regression analysis because only those stocks that have long-term "complete" data are used, thus biasing the study in favor of finding that dividend yields help to predict long-horizon returns. As Fama and French note, the dividend yields also contain forecasts of future dividend growth, which means that they are a proxy for expected stock returns.

Chen, Grundy, and Stambaugh (1986) compare the average "risk-adjusted" returns for high- and low-yield stocks to determine whether there is a relation between cash dividends and required rates of return. Their research differs from previous studies because their main pricing model is a multifactor model that contains two risk measures, defined as the coefficients $\beta_{p(M)}$ and δ_p in the regression

$$R_{pt} = \alpha_p + \beta_{p(M)}RVW_t + \delta_p PREM_t + \varepsilon_{pt},$$

where R_{pt} is the return on portfolio p in excess of the Treasury-bill rate, RVW_t is the return in excess of the Treasury-bill rate on the value-weighted portfolio of stocks in the NYSE, and $PREM_t$ is the difference between the return on a portfolio of "junk" bonds and the return on a long-term U.S. government bond. They analyze monthly returns on portfolios that are formed at the end of each year using a simultaneous two-way classification based on dividend yield and firm size. The dividend yield for a given stock is computed as the sum of dividends per share paid during the previous year divided by the share price at the beginning of the previous year. At the end of each year, beginning in December 1942, each firm on the NYSE with a complete return data available for the previous 5 years and a positive dividend yield is classified into one of 20 portfolios. The 20 portfolios are defined by quintiles of market value and by quartiles of dividend yield. Each year, they estimate the risk measures $\beta_{p(M)}$ and δ_p for each of the 20 portfolios using monthly returns over the previous 5 years. In the second step, Chen, Grundy and Stambaugh regress the cross-sectional portfolio returns month by month on the estimated portfolio multiple risk measures and the dividend yields. This process is repeated for each year from 1943 to 1978, and the estimated risk measures and the dividend yields are updated every year. Their approach is different from previous studies because this method allows the risk premium to change every month, which may reduce the distorting effects of changing risk. In each case, Chen, Grundy, and Stambaugh (1986) find that if the value-weighted market beta is the only risk-adjustment, the estimated dividend coefficient is reliably positive. However, when they include a second risk measure "PREM" (measures changing investment opportunity set), the dividend coefficient is generally not statistically distinguishable from zero.

Researchers on the dividend yield give a number of reasons why these results should be regarded with caution. Stambaugh (1986) pointed out that the explanatory variable, the dividend yield, "is not properly exogenous because it contains a price level that also appears in the regressand." Fama and French (1988) also point out the fact that "yields contain forecasts of future returns and dividend growth, which may bias downward the regression coefficient in the dividend yield regression."

Theoretical Framework

The reason that dividends and interest rates have become increasingly important is that they have been more volatile in the modern era. Moreover, recent tax law changes have narrowed the gap in tax rates on capital gains and dividends, implying that, theoretically, investors should be indifferent between a dividend distribution and a capital gain. But the problem is that traditional stock valuation models such as the dividend-discount model and CAPM fail to distinguish between the dividend and capital gain components of a stock's return. Given the uncertainty of future monetary policy, investors who are more concerned with capital risk than income risk may prefer high-dividend stocks to low-dividend stocks. Especially in times of great economic uncertainty, stocks with low or high dividend-price ratios may exhibit greater price appreciation than can be explained by company financials alone. One possible explanation for this dividend preference is that, battered by the poor market of recent years, some investors have reached the conclusion that retained earnings bear only a tenuous relationship to subsequent capital gains and that high dividend-yielding stocks offer, on a risk-adjusted basis, greater expected returns than low dividend-yielding stocks. Also, dividend policy affects market risk because stocks that pay high dividends have a shorter duration than do stocks that pay low dividends. Theoretically, stocks with a greater duration should be priced low enough to compensate investors for the added risk, which implies that low-dividend yielding stocks should outperform high-dividend yielding stocks. We propose that in a market with rising interest rates and an increased emphasis the value of paying dividends goes beyond their nominal amount, especially in "defensive" market situations where growth prospects are low, and that the effect of dividend distributions may even contribute to future capital appreciation.

Interest rates have also become increasingly important as most stock valuation calculations include a discount rate that is relative to general interest rates. According the common present value calculation, as the discount rate rises the present value of future cash flows should fall. Typically, most people tend to associate increases in interest rates with decreases in bond prices, but as our theory would predict a rise in interest rates should decrease stock prices as well.

INCLUDEPICTURE "http://www.agecon.purdue.edu/academic/agec217/deboer/graphs/intratels.JPG" * MERGEFORMATINET

The economic theory would also predict that small capitalization stocks should garner a risk premium and that this will lead to above-market average returns over time. Market capitalization is found by multiplying a company's outstanding common shares by the company's price per share, and this calculation gives an impression of how large the company is. Using market capitalization as an explanatory variable for capital gains allows our regression to test whether a readily available piece of data can have predictive power.

Market capitalization is important because large companies are easier to follow in the marketplace, and more information is readily available about them relative to small companies. Therefore, it is often thought that because of the lack of information about small companies, and their relatively riskier nature, investors should be compensated for the added risk in investing in small capitalization stocks. According to this theory, we should find a statistically significant relationship between market capitalization and stock prices. When determining how best to incorporate market capitalization into the model we determined that market capitalization is only relevant in relation to the capitalization of the other companies in the data set.

Data Set

For our data set, we used the Center for Research in Security Prices (CRSP) which maintains the most comprehensive collection of security price, return, and volume data for the NYSE, AMEX and NASDAQ stock markets spanning from 1925 to the present. CRSP is a research center at the University of Chicago Graduate School of Business, and access to the database is located on the Wharton Research Data Services (WRDS) site. To narrow the range of our data set, we used the Standard & Poor's 500 stock index, which is often cited as a good measure of the overall stock market. We examined the period from 1965 to 2005. In the data set, past research has tended to focus on a broad index of companies that have various characteristics, ranging from small to large market capitalizations and growth "start up" companies to well-established "blue chips." Examining only the S&P 500 does have its limitations, considering the fact that the companies in the S&P 500 are closely followed and there is probably not a risk premium with those S&P 500 stocks with the smallest market capitalization but we would argue that in looking at the effects of dividends we will be better able to see the impact on large, more well-established companies such as those in the S&P 500. For our regression we use monthly stock market data from 1965 – 2005. Although the primary reason we are using this time period is due to the availability of price and dividend data, we also believe that the modern period of stock valuation is significantly different than the previous period, as to render the previous period irrelevant for our analysis. This is based on the fact that the industrial composition of the index has changed substantially over time, from mostly industrialized utilities such as railroads to a more heterogeneous mix of companies. In the past these companies were much more stable investments, as reflected in their lower standard deviation, and they paid much higher dividends and their annual total return was lower because of their lower risk.

Model Specification

The specification of our model was fairly straightforward, as we wanted to measure the percentage change in stock prices as a result of the percentage change in our variables. Therefore, we used the double-log functional form with the natural log on both the left-hand side and right-hand side of the equation. We chose capital gains (or price appreciation or depreciation) as our dependent variable, and interest rates, dividends and market capitalization as our independent variables.

It is likely that in our regression equation we have excluded some important variables that are known to have a significant impact on stock prices. Numerous studies have shown that the most important indicator of future market returns is the level of risk, and that investors who are willing to bear greater risk are usually compensated with higher stock returns. GDP and earnings per share are also other variables that have been shown to have some impact on share prices, but there are several reasons why not to include all of these variables in our regression equation. The first problem is that many of these variables are significantly correlated with one another, resulting in the violation of Classical Assumption VI, which specifies that no explanatory variable is a perfect linear function of any other explanatory variables. While we would always expect some level of correlation between independent variables, multi-collinearity is a greater problem in a regression such as ours because of the time-series nature of the data to constantly increase over time. The second problem is that, while certain variables have a well-known impact on stock prices, we are unable to predict with any certainty the future course of movements in these variables and so using them as a barometer to invest in the stock market would be useless.

In our regression analysis, we limited our research to information that is readily to known at the time of the investment decision, such as the interest rate level, dividend yield (dividends paid last year over the average stock price last year) and market capitalization. .

Another problem with our time-series data set is that the independent variables can appear to be more significant than they actually are if they have the same underlying trend as the dependent variable. This is an especially difficult problem with stock prices, dividends, and market capitalization because they tend to increase over time and this may not be attributable to an underlying causal relationship between them. Therefore, a regression analysis of these two variables might result in a *spurious regression*, probably caused by the fact that our independent variable is non-stationary and has a mean and variance that change over time. Often times, because financial variables tend to increase over time the regression estimators will falsely attribute to the wrong independent variable the impact on the dependent variable.

Regression Results

```
Number of obs = 213,566
R-squared     =  0.3787
```

lncapgain	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lndividend	1.129363	.239779	4.71	0.000	.6593961	1.599329
lnrate	-.7880864	.1921628	-4.10	0.000	-1.164725	-.4114473
lnmktcap	.8309234	.0958389	8.67	0.000	.6430791	1.018767
_cons	6.463215	1.071843	6.03	0.000	4.324027	8.564027

As expected, our coefficient on lndividend is positive, indicating that an increase in dividends will tend to increase stock prices. The coefficient 1.129 means that a 1% increase in dividends will increase the price by 1.129%, meaning that an increase in dividends has a more than one-for-one impact on capital gains. This is an interesting result, but it seems quite normal considering a dividend represents a positive outflow of cash from the company. Accordingly, a cash outflow should decrease the value of the company as assets fall in order to pay the dividend, so we would expect that the stock price should fall by the per share value of the dividend. Given our theory, we would have expected to see a larger decline in share price, and so there is the possibility that the company will benefit from paying dividends because the cash outflow signals to investors that the company is strong. Also, as a proxy for interest rates we used the yield on long-term corporate bonds. We predicted the coefficient on lnrate to be negative, meaning that an increase in interest rates will decrease stock prices. This is nicely correlated with our economic theory of how stocks are valued by using a discount rate. As interest rates rise, the discount rate used to value future cash flows should also rise, which will have the effect of decreasing the present value of a stock. The meaning of the coefficient on lnrate is that a 1% increase in interest rates will decrease stock prices by .788%. Our third variable, lnmktcap, had a positive coefficient as expected, and a large t-statistic. As expected, the R^2 for our regression equation is somewhat low, at 0.3787, meaning that our dividend, interest, and market capitalization variables only explain 37.87% of the variation in stock prices. This is significant, however, when we consider the importance of other measures of company performance such as EPS that are heavily relied on to value a company's shares.

We tested for serial correlation by first plotting our data set on a time-series graph, with the residuals against time:

```
. plot err1 year

 .975074 +
         |                                                         *
         |                                                          *
         |                                                       *    *
         |
         |                                                      *      *
*
    R    |                                                    *
*
    e    |                                                   *
    s    |                                                **
    i    |
    d    |                                              *
    u    |  * **                                       *
    a    | *                                     *  *
    l    |       **                                  *
    s    |          **                         *  *
         |             ** *                   *
         |                 *                *
         |                    * *
         |                   *   *
         |                         *       *
-.753042 +                          * ** *

+----------------------------------------------------------------+
             1965                      year
2005
```

It seems from the graph that the expected value of our error term observations does not equal zero, and it appears that this is evidence of serial correlation in our data. On our graph, the error term tends to have the same sign from one period to the next, implying positive serial correlation. This is probably due to the lingering effect of some major economy-wide shocks, such as the rise and fall in GDP growth in the 1960s and 1970s and the dynamic change of in the interest rate during the 1970s through the 1980s. Examining our Durbin-Watson statistic, with $D_L = 1.44$ and $D_U = 1.54$, we would reject the null hypothesis that there is no serial correlation because d = .0498.

```
. estat dwatson
Durbin-Watson d-statistic(  2,    213,566) =  .0498492
```

In our regression analysis, we would expect positive serial correlation because we have omitted several variables that have strong effects on stock prices—such as GDP—that will cause our equation to have specification bias. However, as we have noted before, including such variables will have negative consequences on the rest of our regression and this is why we have chosen to leave them out. As the best remedy for serial correlation is to find the omitted variables, then there is nothing further we can do.

Besides the formal regression biases that plagued our regression analysis, we believe that there is a natural bias in using market capitalization as an independent variable. Because market capitalization reflects the current market price of company shares, there is going to exist a direct relationship between market capitalization and capital gains. As prices rise, market capitalizations will rise as well and so it becomes difficult to predict whether one variable causes the other to change. We believe then that this caused a spurious correlation in our regression, which cripples any explanatory power that we could have derived from the coefficient on market capitalization. Another reason that market capitalization was a poorly conceived independent variable on our part has to do with the selection of our data set. Because we limited our data to the S&P 500, our observations consisted mostly of large-capitalization companies. This is a problem because the OLS estimator tries to attribute the change in capital gains as a result of differences in market capitalizations, but the market capitalizations of the S&P 500 are all relatively large.

Conclusion

Despite the fact that our data set has generated statistically significant coefficients for our independent variables, it is likely that our regression equation suffers from omitted variable bias and does not explain enough of the variation in stock prices to warrant investing according to our explanatory variables. This is the inevitable result of most financial time-series regressions, but it does not mean that certain factors do not affect stock prices, but rather it means that it is difficult to prove the exact relationship between them. We still believe that in certain time periods, especially when there are many accounting errors and earnings estimates cannot be trusted, some investors will prefer high-dividend paying stocks and such stocks will deliver above market average returns over that time period. But the serial correlation in our time-series data set has rendered our OLS estimator to bias our standard errors and make unreliable any subsequent hypothesis tests. Theoretically, dividends should play an important role in an investment strategy, whether they are a signal to investors of a company's strength or a telling sign that the company cannot reinvest its cash at a higher rate than the investors' required rate of return. But we could also make the argument that dividends may very well be influenced just as much by past prices as current prices are influenced by past dividends (Granger causality).

Another reason why our regression analysis will not pick up some of the variation in the capital gains as an effect of dividends is because, while companies are willing to increase dividends to signal company profitability to investors, they often do not cut dividends. Therefore, dividends will tend to increase over time and as stock prices increase as well the OLS estimator will falsely attribute to dividends the increase in stock prices that is probably caused by some underlying macroeconomic variable such as inflation. While our market capitalization variable could be considered a failure, we think that there is a strong relationship between interest rates and stock appreciation. This has serious implications in our current market environment in which the Federal Reserve uses interest rates to stabilize the economy, and this has the side effect of causing greater fluctuations in interest rates during our modern era than in the past. If one believes that interest rates will rise in the future, our regression results will certainly indicate that any positive company earnings will be tempered by declines in the present value calculation as an effect of the rising discount rate.

Fama, Eugene F., and Kenneth R. French. "Multifactor Explanations of Asset Pricing Anomalies." The Journal of Finance 51.1 (1996): 55-84.

Campbell and Shiller. "The Dividend-Price Ratio and Expectations of Future Dividends." *Review of Financial Studies*, 195-228.

Hodrick, R., 1992, "Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement," *Review of Financial Studies* 5, 357-386.

Blume, Marshall E. "Stock Returns and Dividend Yields: Some More Evidence." The Journal of Finance 62.4 (1980): 567-577.

Goetzmann, William N., and Philippe Jorion. "A Longer Look at Dividend Yields." The Journal of Business 68.4 (1995): 438-508.

Chen, Grundy, and Stambaugh. "Changing Risk, Changing Risk Premiums, and Dividend Yield Effects"

Table 1 in (fill in info)

Given that the company's rate of return on capital is equal to the investor's rate of return on other investments