Current state of art in predictability on index level

Possibilities of predictions on index level

Master Thesis Finance

Author: Jeroen Elst

Administration number: 895788

Program: MSc Finance

Faculty: Tilburg School of Economics and Management

Supervisor: Dr. L. Baele

Date: 3-12-2013

Defense: 10-12-2013

TABLE OF CONTENTS

Abstract

Evidence of stock index returns predictability by all type of explanatory variables is still controversial. For investors it can be very useful to increase information, to make the best investment decisions. In the past there are already done a lot of researches to come up with variables that are being tested for a predictive relation with the returns on the stock market index. The main research question of this paper that is tried to be answered is; are there explanatory variables available that have significant predictive power to forecast stock index returns. Contrary to earlier researches, this paper focuses not on just one predictive variable, but gives an overview of important well know (earlier) tested predictive variables. The focus lies on the dividend yield, price earnings ratio, short term interest rate, default spread, term spread, the Fed model and the Cay variable as possible predictive variables. The research is based on an extensive literature review where each variable is described, earlier findings on short and long term predictability are discussed and the expected negative or positive effect of each variable is described. In the empirical part of this research each variable is tested for explanatory power, by using simple linear OLS regression models. The empirical part is based on data from the SP 500 index returns for the period 1965-2012. The literature findings and the empirical results show huge difference between all the tested variables and some of them show definitely predictive power. For short run predictability the Cay variable shows the most predictive power and at longer horizons the dividend yield becomes more important. Overall this research gives an extensive overview off well known explanatory variables and provides insight in predictability on stock index returns, which can help investors to make better investment decisions.

1. INTRODUCTION

1.1. PROBLEM DEFINITION

Changes in stock index returns are characterized by a lot of variables that have influence. The most difficult part in doing predictions on what will happen with returns on the stock index is that explanatory variables have only small effects on the stock index. Another difficult point that occurs is that there are also non predictable events (for example a natural disaster) and they can also have huge effects on returns. Important to keep in mind is that predicting for the full hundred percent what will happen with the value of the stock index is not possible. In the past there are already done a lot of researches to come up with variables that are being tested for a relation with the returns on the stock market index. To find a one on one relation is not possible, but the idea was to use the information that came from an explanatory variable to get a better idea of what is likely to happen with returns on the stock market index. Because there are a lot of variables that can have effect and the value of the effect is very small, this paper will focus on the most important researches that are done in the past. An overview of the most important explanatory variables can maybe lead to better indications of what is likely to happen with stock index returns. In the best case scenario the revised overview of explanatory variables can help to get a better insight on what returns on the stock index drives and this information can be useful for all kind of investment decisions.

1.2. RESEARCH QUESTION

In this research the focus lies on the current state of the art in predictability on index level. There will be made an overview of the most well-known explanatory variables for doing predictions on index level. By combining a literature study of important studies about this subject and a small own empirical research, I will try to answer the main research question of this paper.

The main research question of this paper is:

Are there explanatory variables available that have significant predictive power to forecast stock index returns?

It is important to mention that I will concentrate only on variables that are earlier tested on stock return predictability. To keep the research workable, I choose to make only use of variables that are tested in earlier studies and not trying to find an explanatory variable by myself. There is already done a lot of research on all type of explanatory variables, so by picking the most well-known variables from the literature, the overview will contain the most important variables.

To answer the main research question, I will make use of the following sub-questions:

1) Can the dividend yield ratio be used as predictive variable for forecasting stock index returns?
2) Can the price-earnings ratio be used as predictive variable for forecasting stock index returns?
3) Can the short term interest rate be used as predictive variable for forecasting stock index returns?
4) Can the term/default spread be used as predictive variable for forecasting stock index returns?
5) Can the Cay ratio be used as predictive variable for forecasting stock index returns?
6) Can the Fed model be used as predictive variable for forecasting stock index returns?

1.3. COMPOSITION OF THE THESIS

This paper is composed out of five sections: Section 1contains the literature research, the theoretical framework. It contains the basic principle of asset pricing; this will help to understand why asset prices are what they are. Section 2 contains a general overview about the most well-known explanatory variables in the literature, where results of earlier studies on each variable will be discussed separately in an extensive way. The data and methodology for the empirical research are described in section 3. The practical findings (the empirical research) will be the focus of section 4. In this section I will do a small empirical research to check the discussed variables by running some regressions with the most recent data. The results will be presented and the most important results will be analyzed. Also there will be made a link between the findings of the literature review and the empirical research to answer the sub questions of this paper. The conclusions and recommendations will be the main part of section 5.

2. LITERATURE REVIEW

2.1. BACKGROUND ON ASSET PRICING; THE BASIC IDEA

As earlier mentioned there are a lot of researches done based on variables that can drive the value of the stock index. In this part the focus will be not already on the variables itself but on the asset pricing theory, by understanding this it would be clearer to understand the effect (positive/negative) of different variables on index prices.

According to John H. Cochrane (2000), the basic idea behind the asset pricing theory is that it tries to understand the prices/values of claims of uncertain future payments. In that case a low price indicates a high rate of return; this explains why some assets pay higher average returns than others. To value an asset, the delay of the payments and the risk of the payments must be taken into account. Corrections for time are not difficult to work out, but corrections for risk are much more important determinants of values of assets. Uncertainty or corrections for risk factors makes asset pricing very interesting and challenging. Asset pricing theory is based from one simple concept that the price equals the expected discounted payoff. The other part is elaboration, special cases, and tricks that make the equation useful for other applications. The most common approach in academic settings is the absolute approach where they use the asset pricing theory to give economic explanations for why the prices are what they are, or in order to predict why and how prices might change if there are changes in policy or economic structure etcetera. The central and unfinished task of the absolute asset pricing approach is to understand and to measure the sources of macroeconomic risks that drive asset prices. A lot of empirical work that is done has documented tantalizing stylized facts and links between macroeconomics and finance. For example, expected returns vary across time and across assets in ways that are linked to macroeconomic variables, or variables that can also forecast macroeconomic events.

According to Lucas, Robert E. Jr. (1987) standard macroeconomic models predict that people do not care much about business cycles. Cochrane (2000) states that asset prices are beginning to reveal that they forego substantial return premia, to avoid assets fall in recessions. This fact could tell us something about recessions. Furthermore he advocates a discount factor / generalized method of moments view of assets pricing theory and associated empirical procedures. He summarized asset pricing in two equations:

$$p_t = E(m_{t+1} x_{t+1})$$
$$m_{t+1} = f(\text{data, parameters})$$

where p_t *= asset price,* x_{t+1} *= asset payoff,* m_{t+1} *= stochastic discount factor.*

The major advantage of this approach is the simplicity and universality. Another advantage is that it allows us to conveniently separate the step of specifying economic assumptions of the model from the step of deciding which kind of empirical representation to understand. For a given model (choice of f), the first equation can lead to predictions stated in terms of returns, price dividend ratios, expected returns-beta representations, moments conditions and so forth. The idea behind asset pricing in general can help a lot to understand why stock market indices are moving in different directions. The above described approach can be also very helpful in digesting the results of empirical work, where a number of apparently distinct, but fundamentally connected representations are used. The organizing Principle that Cochrane (2000) used is that everything can be tracked back to specialization of the basic equation: $p = E(mx)$.

Another fundamental point in Cochrane's book is that risk corrections to asset prices should be driven by the covariance of asset payoffs with consumption or marginal utility. An asset that performs badly in states like a recession, in which an investor feels poor and is consuming little, is less desirable than an asset that performs badly during a boom. This theory is also captured in the CAPM principle to form portfolios to diversify the idiosyncratic risk away. In this paper we only discuss the basic principle of the CAPM. This will help to understand that prices of assets are what they are.

According to Fama and French (2004) the basic principle of the CAPM will be explained.
The CAPM is based on the idea that investors demand additional expected return (called the risk premium) if they are asked to accept additional risk. The risk that has to be taken has an important effect on the interest rate.

How riskier the asset, the higher the interest compensation will be. In the CAPM formula the Beta will be used to determine what the risk factor is. For individual security perspective the Security market line (SML) will be used and its relation to expected return and systematic risk (Beta), to show how to price an individual security in relation to their security risk class (This explains how securities can be discounted at the 'correct' interest rate) the graph is shown on the next page (figure 1).

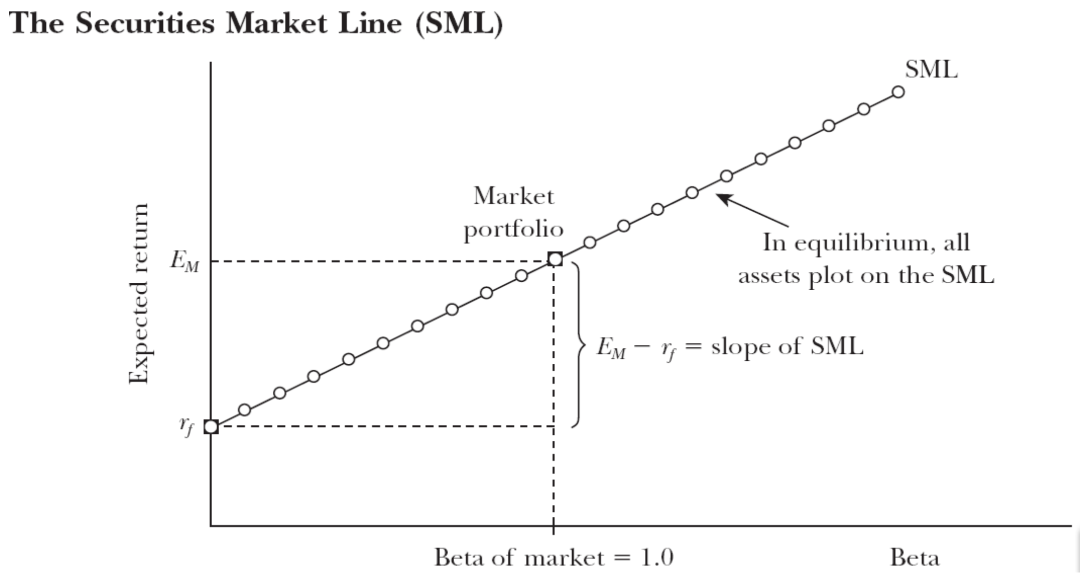


Figure 1

The CAPM model uses some assumptions for asset pricing and investors behaviour. They do this so it is possible to value stocks, securities, derivatives and/or assets by relating risk and expected return on a theoretical approach. These assumptions that will be discussed are based on Fama and French (2004), and Falahati (2009).

Investors are risk averse individuals, who maximize the expected return of their end of period wealth, so the assumption is that the basic CAPM is a one period model. This is a theoretical approach, but in a real dynamic world this assumption is not valid. They also approach that there are a definite number of assets and their quantities are fixed within one period. Investors have also homogeneous expectations about returns, because the assumption is made that everyone has the same information and all assets are perfectly priced in a perfectly competitive market. The assumption is that asset markets are frictionless and all information is costless and available for all investors and that there are no market imperfections, such as taxes. At least they assume that investors may borrow or lend unlimited amounts at the risk free rate. Although the assumptions mentioned above normally are not valid or met, CAPM remains one of the most used models to determine risk and return.

After the earlier discusses assumptions of the CAPM it is to imagine that there are some doubts about using the model in practical. According to Perold (2004) the Capital Asset Pricing Model is an elegant theory with profound implications for asset pricing and investor behaviour. In this article they ask how useful the CAPM model is, given the idealized world that underlies its derivation. They came with several answers.

- It is possible to examine whether real world asset prices and investor portfolios conform to the predictions of the model, if not always in a strict quantitative sense, and least in a strong qualitative sense.
- Even if the model does not describe our current world particularly well, it might predict future investor behaviour.
- The CAPM can serve as a benchmark for understanding the capital market phenomena that cause asset prices and investor behaviour to deviate from the prescriptions of the model.

This approach works very well in theory, but in practical we live in a dynamic world, and not all the assumptions of the basic CAPM are always valid, but the basic concept of the CAPM helps to understand that prices of assets are what they are.

According to Cochrane (2000) it is important to keep in mind that idiosyncratic risk does not affect prices. Only the component of a payoff perfectly correlated with the discount factor generates an extra return. *Idiosyncratic* risk, uncorrelated with the discount factor, generates not a premium. You might think that an asset with a high payoff variance is 'risky' and thus should have a large risk correction. However, if the payoff is uncorrelated with the discount factor m, the asset receives *no* risk-correction to its price, and pays an expected return equal to the risk-free rate! So to say it more in general, an investor gets no compensation for holding idiosyncratic risk. All investors are rewarded for bearing this systematic risk, but they are not rewarded for bearing idiosyncratic risk, because this can easily be eliminated without any cost.

According to Campbell R. Harvey and Stephen Gray (1997) diversification is a strategy to decrease the risk without decreasing the expected return, so most people will hold diversified portfolios. In the understanding graph is shown how diversification in the simplest form could be explained. The graph (figure 2) shows a set of risky assets and also the choice of two investors, namely investor A and investor B. By combining these two assets it is possible to diversify the risk of these two investment possibilities and create a 'new' choice to invest in.

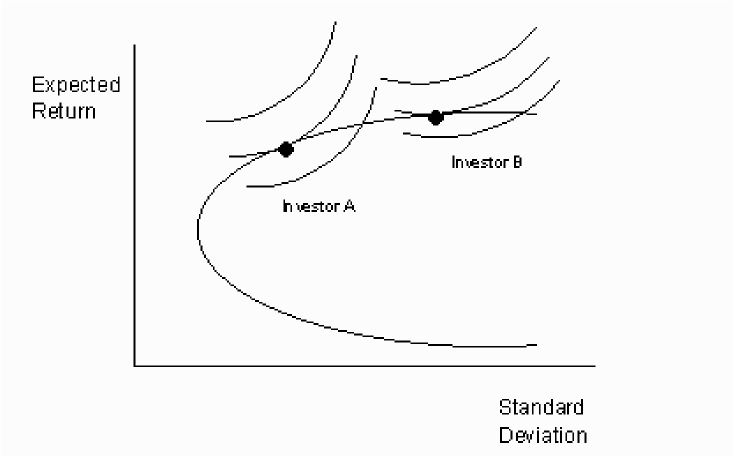


Figure 2

In the above figure (2) are only risky assets shown, but it is of course also possible to combine with a risk free asset.

An investment in the market portfolio and the riskless asset is an optimal strategy for all investors, there will be only a difference in the proportion invested in the market portfolio and the riskless asset. This is shown in figure 3: the straight line from *Rf* to *M*.

This holds only when the assets and portfolios returns are perfectly positive correlated with those of the market portfolio. In the real world this is not a fact, it is also possible that asset and portfolio returns are negative correlated with those of the market portfolio. This give us an extra opportunity, by combining two assets one that is positive correlated with those of the market and one that is negative correlated with those of the market. In that case the state of the economy makes not a very important sense, because one of the assets will give a high return in a recession and the other gives a high return during a boom. By combing these, there will be less risk and a relative high expected return. This is also shown in figure 3.

Capital Market Line (*CML*).

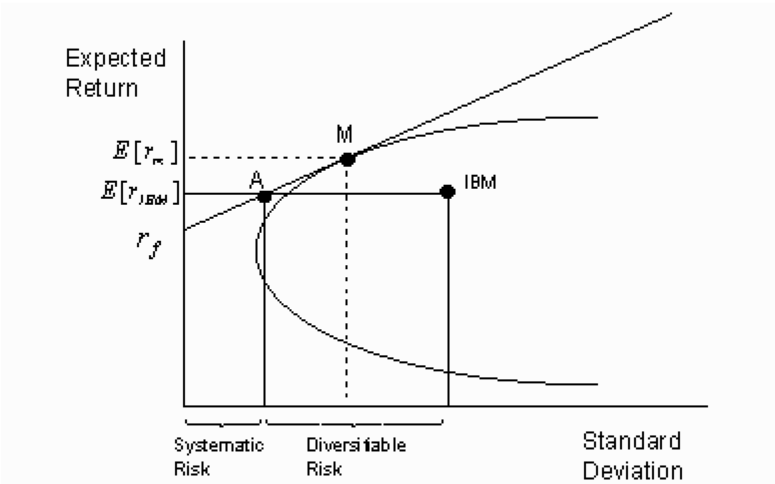


Figure 3

For example (This example is based on Campbell and Gray (1997)) in the above graph (figure 3) are shown an asset of IBM and an asset A, this is a combined portfolio of the risk free rate and the market portfolio. Every risk-averse investor will prefer A above only an asset IBM, because the expected return is in both cases the same, but the risk of asset A is smaller. It is not possible in this case to find an asset with the same expected return and less risk than asset A. So in this case the total risk of IBM can also be decomposed into systematic risk and idiosyncratic risk.

Based on Bodie, Kane and Marcus (2011), investors are looking for chances of assets that lie above the SML. These assets are undervalued, because for a given amount of risk, they yield a higher return. In a perfectly efficient market all assets lie on the SML. When an asset lies above the SML, it is a very attractive asset for an investor and 'everyone' wants to buy, because of that there is an arbitrage opportunity. This is only for a short period; the market ensures that the asset will be priced 'correctly'. The problem is that each investor looks for these opportunities, so they are difficult to find. Predictions about what will happen with all kind of assets are important for investors and each investor has his own expectations. As earlier mentioned in the introduction there is already done a lot of research on all type of explanatory variables and that is where this research is about!

The question of whether stock returns are predictable has received an enormous amount of attention. This is not surprising because the existence of return predictability is interesting for investors, but can also have important implications for financial models of risk and return.

3. BACKGROUND; ON THE MOST IMPORTANT EXPLANATORY VARIABLES

According to Lettau and Van Nieuwerburgh (2007) one branch of the literature asses that expected returns contain a time-varying component that implies predictability of future returns. Due its persistence, the predictive component is stronger over longer horizons than over short horizons. Classic predictive variables are financial ratios, such as the dividend-price ratio, the earnings-price ratio and book-to-market ratio. The question arises or prices scaled by fundamentals contain useful information about future movements in the aggregate stock market.

Fama and French (1988) provide initial evidence that prices normalized by dividends or earnings can be used to capture time in expected returns. Campell and Shiller (1988A, B) used the definition of returns to show why the price dividend ratio should be expected to forecast long-run stock market movements. In the last decade of the 20th century, movements in aggregate stock prices and consequently returns were much different from what earnings and especially dividends would seem to have implied, raising doubts as to whether stock returns are at al predictable. This will be later discussed extensively when the variables will be discussed separately.

Other variables have also been found to be powerful predictors of long-horizon returns, some macroeconomic variables contain information about future returns in addition to that revealed by financial valuation ratios. Lettau and Ludvigson (2001A) use the budget constraint of a representative consumer to show why fluctuations in the aggregate consumption-wealth ratio should forecast movements in the aggregate stock market. They estimate a consumption-wealth ratio called the CAY ratio that predicts excess returns pretty well and captures a considerably larger fraction of the variation in real stock returns and excess stock returns over a Treasury bill rate. They find that this variable is a better forecaster for of future returns at short and intermediate horizons than the dividend yield, the dividend payout ratio and several other popular forecasting variables. Upon the ideas of Lettau and Ludvigson (2001A) Julliard (2007) combines the cay-ratio with future labor income growth to predict stock returns. The paper finds that one third of the variance of returns is predictable, over a horizon of one year, using expected future labor income growth rates and Cay together as forecasting variables.

Santos and veronesi (2006) developed a novel economic mechanism that generates stock return predictability in both the time series and the cross-section. Investors income has two sources, wages and dividend that grow stochastically over time. Fluctuations in the consumption-labor income ratio imply that the risk premium required by investors to hold financial assets will vary over time, and they show that a ratio of labor income to consumption is a good predictor of returns.

Bekaert and Ang (2006) found in their research ‘stock predictability is it there’ that the most robust predictive variable for future excess returns is the short rate. A caveat here is that it is only significant at short horizons.

Chen, Roll, and Ross (1986) investigates whether innovations in macroeconomic variables are risks that are rewarded in the stock market. They find that these sources of risks; spread between long and short interest rates, expected and unexpected inflation, industrial production and the spread between high and low grade bonds, were significantly priced.

Fama and French (1989) argued that the business cycle pattern of real output has direct influence on certain assets yields and movements in these yield may be used to explain the observed predictability in excess returns. If financial markets are different and future economic prospects are indeed reflected in the expected changes in prices in these markets, then this would suggest that excess returns are predictable, because risk itself is predictable. So in other words when these sources of risks are priced on the stock market it indicates that these variables also have forecasting power.

Campbell andYogo (2006) found also evidence for predictability with the short rate and the long-short yield spread and these results are based and valid under the conventional t-test.

Rangvid (2005) considers a set of standard macro variables; including interest rates, term spread, inflation rate, unemployment rate etc. Among the macro variables he states that, interest rates are the most consistent and reliable predictors of stock returns. In one of his other works Rangvid (2006) he scales share prices by fundamentals in return-predicting regressions as well as the newer findings on the importance of macro-economic variables for capturing variation in expected returns. In this paper he uses the Gross domestic product (GDP) ratio as a macro variable in return predicting regressions, i.e., replace aggregate dividends or earnings with GDP as a price normalizing variable. Some priori advantages of the price to GDP ratio to the earlier mentioned CAY ratio is that GDP data are available for a longer period.

Another interesting variable is the ‘Fed model’, the predictive role of the stock-bond yield gap, the difference between the stock market earnings (dividend) yield and the ten-year Treasury bond yield, also known as the "Fed model.". Maio P. (2012) shows that the yield gap forecasts positive excess market returns, both at short and long forecasting horizons, and for both value and equal-weighted indexes, and it also outperforms competing predictors commonly used in the literature.

Bekaert and Engstrom (2010) did also research on the ‘Fed’ model’ they find indeed a strikingly high times series correlation between the yield on nominal bonds ant the dividend yields on equities. Also this striking variable will be discussed later more extensively.
In the above section the most common variables that are interesting for predictability are mentioned and are all based on previous research. In this review, these are the most important variables that are taken into account and could have influence on stock return predictability. The variables are very broadly and far for from completely discussed. In the next paragraphs the variables will be discussed separately.

3.1. DIVIDEND YIELD

One of the most common predictive variables for stock returns is the dividend-yield, calculated as the ratio of the dividend and the price. As we discuss below the dividend yield had mainly predictive power over longer horizons.

Short-term predictive power

Ang and Bekaert (2001) and Valkanov (2003) show that the dividend price ratio has little power for predicting stock returns at short horizons, the main reason being that the dividend price ratio is a slowly moving variable. Lettau and Ludvigson (2001A) confirmed these findings and state that there are better variables available for predicting on the short horizon, which will be discussed later.

Fama and French (1988) used dividend price ratios to forecast stock returns on the value and equal weighted portfolios of the NYSE and made regressions of returns on yields. They conclude that returns on dividend yields indicate that time variation in expected returns accounts only for small fractions of short (monthly, quarterly) horizon returns, typically less than five percent.

Nevertheless Ang and Bekaert (2001) find that stock returns are predictable, mainly at a short horizon, not really on a long horizon. The predictive ability of the dividend yield can be best shown in a regression with short rates only at short horizons. High dividend yields predict high future interest rates and have good predictive power for future cash-flow growth rates and not for future excess returns. Their results imply that linear models of expected returns are unlikely to satisfactorily capture all the predictable components in returns.

Long-term predictive power

Letau and Ludvigson (2005) present a number of key stylized facts with respect to aggregate stock market behavior:

1. There are no large predictable movements in dividends in the U.S. data. In particular, the dividend price ratio does not have important forecasting power on a long horizon for the growth in dividend payments.
2. Returns on aggregate stock market indices in excess of a short-term interest rate are highly predictable over long horizons. The dividend price ratio is highly persistent and forecasts excess returns over horizons of many years
3. Variance decompositions of dividend price ratios show that changing forecasts of future excess returns include almost all of the variation in dividend price ratios. This forms the basis for the conclusion that expected dividend growth is approximately constant.

The empirical evidence on the behavior of the dividend price ratio made a change of how financial economists comprehend asset markets. The view that expected returns are approximately constant is changed into a view that time-variation in expected returns plays an important role of aggregate stock market variability.

The first key finding of Lettau and Ludvigson (2001A) is that they find important predictability in direct long-horizon regressions. The second important finding is that the dividend forecasts that they found positively co-variation with changing forecast of excess stock returns. This covariance is important because positively correlated fluctuations in expected dividend growth and expected returns have offsetting effects on the dividend price ratio. The market risk premium and expected dividend growth thus vary considerably more than is apparent using the dividend price ratio alone as a predictive variable.

Their results explain why the dividend price ratio has been found to be a relatively weak predictor. The expected dividend growth is not constant, but it varies over horizons from one to six years. Then it is measurable, but not on ultra-low frequencies and that dominates the sampling variability of the dividend price ratio. Another difficult point is that the common variation in expected returns and expected dividend growth makes it more difficult to display statistically significant power for future returns consistent with Ang and Bekaert (2001). Their results show that expected returns fluctuates at very high frequencies, and covary with changing forecast of dividend growth. Therefore the expected returns not only vary, but they vary much more than what can be discovered using the dividend price ratio alone as a predictive variable.

In contrast to the findings of predictability at short horizons Fama and French (1988) state that at a longer horizon (2-4 years). The dividend yield explains often more than 25% of the variances of returns. Their findings are not consistent with for example Ang and Bekaert (2001). The explanation that Fama and French gave for the striking results is that the growth of the variance of the regression residuals is attenuated by a discount-rate effect: shocks to expected returns are associated with opposite shocks to current prices.

The cumulative price effect of an expected return shock and the associated price shock is almost zero. On average, the expected future price increases implied by higher expected returns are offset by the immediate decline in the current price. Thus the time variation of expected returns gives rise to mean-reverting or temporary components of price.

Other remarkable research

A well-known research related to the dividend price ratio is done by Campbell and Shiller (1989) they derive the dividend-ratio model as a linear approximation to an exact relationship between stock prices, stock returns, and dividends. They propose a method for analyzing the historical movements of stock prices in relation to dividends. A linear approximation to the stock return implies that the dividend price ratio can be written as a discounted value of expected future dividend growth rates and discount rates. By comparing the dividend price ratio with the forecast of dividend growth and discount rates it is possible to test the adequacy of some measures of the ex post discount rate on stock return. This approach is different form the regression method that is commonly used.

The standard method would be to regress the one-period ex post stock return less the ex post discount rate, on a constant term and on some variables known at the start of the period. The advantage of the chosen approach is that it may have more power to detect long-lived deviations of stock prices from the fundamental value implied by the model. A second advantage is that their model compares movements in the dividend price ratio with the movements that are implied by the model, where regression results do not tell us whether the behavior of the dividend price ratio is quite different form that implied by the model or whether it is rather similar.

The conclusion of their research is that stock returns are somewhat predictable. The dividend price ratio has a positive effect (significant on a 5% level) on stock returns. The special feature of their approach is that they can use the dividend ratio model to compute the implications of this predictability for the behavior of the dividend price ratio; they find that the dividend price ratio moves about 50% too much. This result can be understood by considering what it means for the dividend price ratio to have a positive effect on subsequent stock returns. The dividend price ratio is high when prices are low, and the effect on returns implies that prices tend to rise subsequently. To eliminate the predictability of returns, the stock price would have to be less variable around the dividend. The main three results of their research are:

1 There is evidence that the dividend price ratio does move with rationally expected future growth in dividends

2 The various measures of short-term discount rates that are used are helpful in explaining stock price movements.

3 There is substantial unexplained variation in the dividend price ratio, it is clear that the long term expected return on stocks is highly variable. Also it does not move parallel with the short-term interest rates

In this research they took only a few variables into account, by adding more lagged variables the explanatory power of regressions explaining stock returns can only increase the results.

Campbell and Shiller (update 2001) did also another research on the effect of the dividend price ratio on prices on the stock market. They conclude that the ratio appears to be useful in forecasting future stock price changes, contrary to the simple efficient-market models. It seems reasonable to suspect that prices are not likely to drift too far from normal levels relative to indicators like dividends and earnings. When stock prices are very high relative to these indicators, prices will eventually fall in the future to bring the ratios back to more normal historical levels. In the first version they conclude that the dividend price ratio has a special high significance compared with many other statistics that might be used to forecast stock prices. The ratio was extraordinarily bearish for the US stock market and in with the more recent study the ratios are even more so. When we look back into time we can conclude that the extraordinarily bearish view for the US stock market was correct. This research was based on data till around the year 2000, here is also a challenge to check this with more recent data.

Expectation about the sign

I expect that the coefficient of the dividend yield would be positive because:

Lamont (1998) finds a positive coefficient on the dividend yield and argues that the predictive power of the dividend yield stems from the role of dividends in capturing permanent components of prices. High dividends predict high future returns. An interpretation is that dividends measure the permanent component of stock prices, due to managerial behavior in setting dividends. Dividend yields show that prices adjust to dividends, but dividends do not adjust to prices; this is because dividends reflect the permanent component of value.

Lamont states that high dividend yields signal high future earnings growth and dividend yields capture price effects. When dividend yields are high today, we predict low dividend growth in the future because payout ratios strongly decrease. This may be the result of dividend smoothing, or it may reflect prices anticipating higher growth opportunities that decrease the payout ratio. The positive relation between current high dividend yields and future earnings growth implies that these growth opportunities do not rapidly translate into higher future earnings. Lower prices today may reflect poor future earnings and future growth opportunities.

Fama and French (1988) state that the dividend yield forecast high returns when business conditions are persistently weak and low returns when conditions are strong. The intuition is that stock prices are low in relation to dividends when discount rates and expected returns are high (and vice versa), so *D/P* varies with expected returns. The fact that the dividend yield forecast stock returns suggests that the implied variation in expected returns is largely common across securities, and is negatively related to long- and short-term variation in business conditions. A story for these results is that when business conditions are poor, income is low and expected returns on stocks must be high to induce substitution from consumption to investment. When times are good and income is high the market clears at lower levels of expected returns. However, that variation in expected returns with business conditions is due to variation in the risks of stocks.

Considering the interpretation of the price and dividends tells that the negative coefficient on price is usually measured by the dividend yield. When required rates of return are high, stock prices are low, so low prices today forecast high returns in the future.

Second, the current level of dividends measures the discounted value of future dividends during this period, and so is a proxy for *price dividend ratio*.

Dividends contain information about future returns because they help measure the value of future dividends, while earnings contain information because they are correlated with business conditions. Both high current prices and high current earnings forecast low future returns. Thus using earnings yields alone to forecast returns is a bad idea, not because earnings are noisy, but because they are informative. In contrast, high dividends forecast high future returns, so using dividend yields alone to forecast returns is more successful.

3.2. PRICE-EARNINGS RATIO

The second variable that is taken into account is the price-earnings ratio; this variable is a ratio of a company's current share price compared to the earnings per share. The price-earnings ratio has some similarities (the method is quite the same) to the dividend yield and a lot of researchers combine these two predictive variables in their study. In some other researches they make use of the earnings yield instead of the price-earnings ratio, these variables have many similarities. The earnings yield can also been seen as the inverse of the price-earnings ratio and this shows the percentage of each dollar invested in the stock that was earned by the company.

Short term predictive power

Lamont (1998) state that earnings are a good measure of the current business conditions; the idea is that earnings contain information for future returns. Dividend and price contain information about future returns; however the information of these variables contains chiefly about short-run variation in expected returns. Both variables (dividend yield and price-earnings ratio) are variables that are unimportant for forecasting long-term returns.

Rangvid (2006) created some regressions for predicting the U.S. annual returns. These were some basic regressions with a constant and one of the predictor candidates, or a multivariate regression. Rangvid found that the price-earnings ratio captures around 15% in annual returns and 8 % in excess returns; these results are quite the same as for the dividend yield. There are also created some regressions at longer horizons, these results will be shown when the long term predictive power will be discussed. Nevertheless the final conclusion of Rangvid (2006) can be interpreted as that the predictive power is higher at a short horizon than a longer horizon, but the predictive power is rather low and depends a lot on the data that is used.

Lettau and Van Nieuwerburgh (2008) mentioned also the results of Lamont (1998) that the earnings-price ratio can be used to forecast returns. They did a research with more recent data and came to quite similar results and find very much the same patterns for price-earnings ratios as for the dividend price ratio. They also state that the macroeconomics literature has recently turned to models with persistent changes in fundamentals to explain the dramatic change in valuation ratios in the bull market of the 1990s. Most of such models imply a persistent decline in expected returns. In their paper, Lettau and van Nieuwerburgh (2008) argue that either of such changes leads to a persistent decline in the mean of financial ratios. Changes in the mean of the valuation ratios have important effects on estimation and inference of return forecasting regressions.

Long term predictive power

As earlier mentioned, Rangvid (2006) created also some regressions for predicting returns at a longer horizon. When they test on a longer horizon the price-earnings ratio captures 7% of the variation in six-year cumulative stock returns. The results of the price-earnings ratio are insignificant, which also implies that there is no evidence of the predictive power of the price-earnings ratio at a longer horizon. They also test for a sub sample during the years 1948-2003 and by doing that the price-earnings ratio capture a higher fraction on the variation in excess returns, this implies that there is some predictive power by the price-earnings ratio. There must be taken into account that results are very different during time slots and that the data that is used is very important. The data in this research was only till 2003, in this research it is interesting to check also for predictive power on the past 10 years, which will be done in the empirical part of the thesis.

Bekaert and Ang (2006) created a bivariate regression to compare the results with what Lamont (1998) find. Lamont (1998) finds a positive coefficient on the dividend yield and a negative coefficient on the earnings yield. The explanation is that the negative coefficient on the earnings yield is due to earnings being a good measure of business conditions.

Bekaert and Ang (2006) found that over the long US sample, while having the same sign found by Lamont are insignificant. Only when they exclude the year 1990 and when they pick the same time slot as Lamont they find significant coefficients for dividends and earnings yields. They conclude that there is little evidence that earnings yields predict excess returns. The earnings yields coefficients are not robust across different sample periods.

When they look at the earnings and dividend growth simultaneously they find the following pattern:

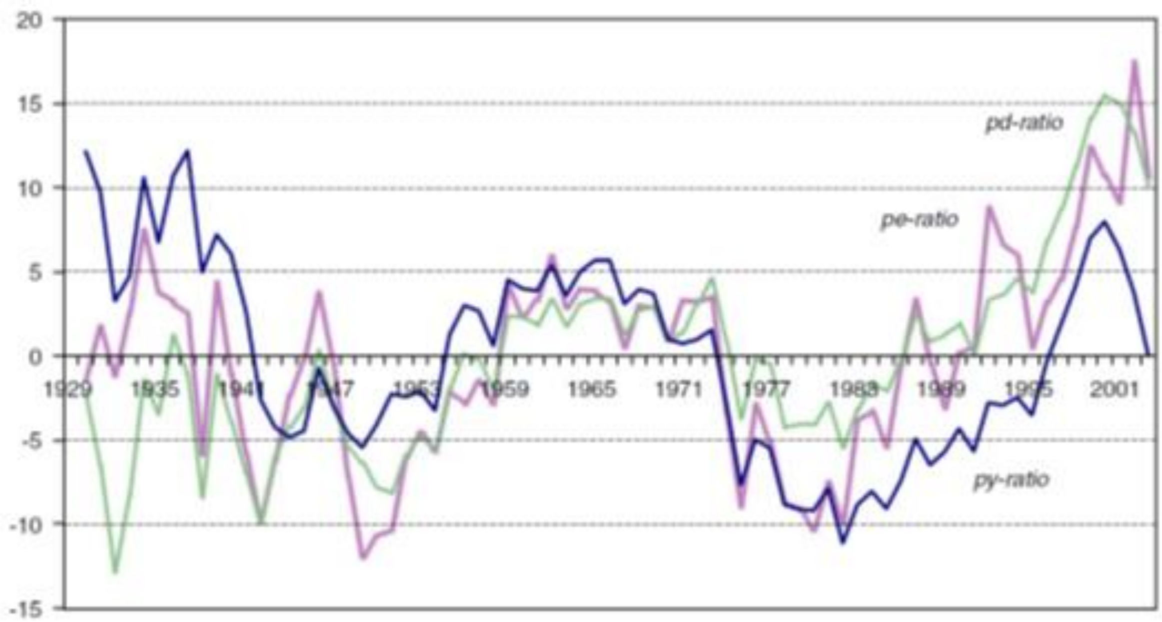
	Dividend growth	Earnings growth	Payout ratio
Dividend yield	−ve	+ve	−ve
Earnings yield	+ve	−ve	+ve

Figure 4

Their results imply that higher dividend yields and higher earnings yields strongly predict lower (higher) payout ratios tomorrow. The most important results are that predictability is mainly a short-horizon, not a long horizon phenomenon and dividend and earning yields have good predictive power for future cash-flow growth rates but not for excess returns. Hence, a potentially important source of variation in price-earnings ratios and dividend yields is the predictable component in cash flows. The linear models that are used are unlikely to capture all the predictable components in returns.

Other remarkable research

Rangvid (2006) did also research on output and expected returns. What earlier mentioned is that the price dividend ratio and the price-earnings ratio have similar characteristics. One of the results of Rangvid is shown very well in a graph, were you can see that the lines in the graph have a similar pattern.



The figure shows the de-meaned price–output ratio (*py*), the de-meaned price–earnings ratio (*pe*), and the de-meaned price–dividend ratio (*pd*) during the sample period 1929–2003.

Figure 5: P/E vs P/D (Rangvid 2006)

The correlation between the price-earnings ratio and the dividend yield is 0.80 which also implies that both ratios are quite similar. They also test for correlations between the financial variables and annual returns and these correlations are fairly low which implies that the predictive power is not very high.

Fama and French (1988) found that the price-earnings ratio has explanatory power, but the earlier described dividend-yield has more. The explanation that they give for this result is that prices/earnings are more variable than dividends. When this higher variability would be unrelated to the variation in expected returns, the earnings-price ratio would be a noisier measure for predicting the expected returns than the dividend yield. Another statement that they make is that stock prices are low relative to prices when discount rates and expected returns are high, and vice versa, so that yields capture variation in expected returns.

Fama and French (1988) estimated some regressions and their results are similar to the results of the dividend yield as a predictive variable. The regressions slopes suggest that the earnings-price ratio has reliable forecast, however they have less explanatory power than the dividend price ratio (because prices are more variable).

Bekaert and Ang (2006) report also a regression with earnings yield as the explanatory variable, and also state that the results are similar to the dividend yield regression. When they analyze their regressions they state that it implies that this result comes from the price in the dominator of both variables.

The above discussed findings on the price-earnings ratios of several researches have a lot of common results. Actually all of them state that the predictive power is higher at a short horizon than at a longer horizon, but the predictive power is rather low. This does not mean that this is negligible for predictability, because it has components that have influence on other variables (cash flow growth) that have effect on returns.

Expectation about the sign

I expect that the coefficient of the price-earnings ratio would show a negative sign.

Lamont finds a negative coefficient on the price-earnings ratio and states that earnings forecasts returns. There is a clear story for this fact: the level of earnings is a good measure of current business conditions. Risk premia on stocks covary negatively with current economic activity: investors require high expected returns in recessions, and lower expected returns in booms. Since earnings vary with economic activity, current earnings predict future returns.

A variety of evidence suggests that expected returns have a macroeconomic component (Fama and French (1989) and, Cochrane (1991)). Expected returns covary negatively with current macroeconomic activity: risk premia are high in recessions and low in expansions. Current corporate earnings covary positively with macroeconomic activity (Lucas (1977)) lists the cyclicality of profits as one of the seven main features of macroeconomic fluctuations. The story for earnings, then, is that they covary negatively with expected returns because earnings measure macroeconomic activity.

According to Bekaert and Ang (2007) the negative relation between the current earnings yield and future earnings may be consistent with either a price effect or mean reversion in earnings. The payout ratio reacts positively to an increase in the earnings yield. In the mean reversion story, this could be an artifact of dividend smoothing. In the price story, lower prices today may reflect poor future earnings and poor future growth opportunities. The poor growth opportunities may increase the payout ratio. Lamont (1998) argues that the dividend payout ratio should be a potentially potent predictor of excess returns, a result of the fact that high dividends typically forecast high returns whereas high earnings typically forecast low returns.

Previous research has generally regarded quarterly earnings as noise that should be discarded or smoothed. What has previously been classified as noise is actually useful information about short-term movements in expected returns. Both high current prices and high current earnings forecast low future returns. Thus using earnings yields alone to forecast returns is a bad idea, not because earnings are noisy, but because they are informative. In contrast, high dividends forecast high future returns, so using dividend yields alone to forecast returns is more successful.

3.3. SHORT TERM INTEREST RATE

The short term interest rate is another explanatory variable that is tested a lot in earlier researches. The short term interest rate is a variable that is easy to obtain, it are the rates at which short-term borrowings are effected between financial institutions or the rate at which short-term government bonds are issued or traded in the market. Typical used data for the short term interest rates is the 3-month T-bill rate and the relative 3-month T-bill rate (difference between the 3-month T-bill rate and a 12-month backward-looking moving average (Bekaert and Ang 2006, Rangvid 2005). Campbell and Yogo used the 1-month T-bill rate as variable for short interest rates. Torous, Valkanov and Yan make also use of the annual nominal 1-month T-Bill rate (average of bid and ask).

The idea to use the short term interest rate as a predicting variable for stock return is easy explicable, because interest rates can be a good indicator of the current state of the economy and expectations for the coming period. The expectation is that this information would be rewarded in returns.

According to Cochrane (2000) we can see some of these effects right away:

1. Interest rates are high when people are impatient, when the Beta is low. Everyone wants to consume now, it takes a high interest rate to convince them to save.
2. Interest rates are high when consumption growth is high. Investors consume less now and consume more in the future. Thus high interest rates lower the level of consumption today, while raising its growth rate from today to tomorrow.

Short term predictive power

Bekaert and Ang (2006) did research with the short rate as an additional regressor and found that a 1% increase in the annualized short rate decreases the equity premium by about 2.16%. The predictive power of the short rate dissipates quickly for longer horizons but remains significant at the 5% level at the one-year horizon. When they omit the 1990s the predictive power of the short term interest rate becomes even stronger. They find that the most predictive variable for future excess returns is the short rate, but hence that it is only significant at short horizons. The forecast from a one-quarter regression with the short rate and dividend yield have a correlation of 87% with true expected returns.

Long term predictive power

The above mentioned research of Bekaert and Ang (2006) did also a regression with the short rate and the dividend yield and found a correlation of 82% at a 5-year horizon. This implies strong predictability on expected returns.

Campbell and Yogo (2006) also find that the predictive power of the dividend yield is considerably weakened but that the predictive power of the short term interest rate is robust. This is in accordance with Ang and Bekaert (2006), but they are distinctive because they develop a new inference methodology (Pretest to determine when the conventional T-test leads to misleading inference) within the linear regression framework of Stambaugh (1999).

Using that pretest they find that the t-test was valid for the short term interest rates and this gives the results of both researches more evidence of predictive power of the short term interest rate. The hypothesis that stock returns are predictable at long horizons has been called a new fact in finance (Cochrane 1999). Campbell and Yogo have shown in their paper that there is indeed evidence for predictability. They also state that the most popular and economically sensible candidates for predictor variables like the Short term interest rates and dividend yield are highly persistent. Taken together these results suggest that there are predictable components in stock returns, but be careful by using efficient statistical tests. Campbell and Thomson (2007) compared predictive variables with the historical average return forecast and have shown that the 'Short term interest rate explanatory variable' performs better than the historical average return forecast.

Rangvid and Lioui (2007) did research on stock return predictability and build op on the idea that the short term interest rate determines the trade-off between money holdings and consumption. They build a theoretical model that shows the mechanism how the short rate finds its way to stock-return predictability regressions. They construct a co integration relation that links share prices and dividend to the short interest rate and this strongly predict stock/excess returns. In contrast to Ang and Bekaert (2006) they find that returns are predictable at long horizons, they differ from Ang and Bekaert, but hence they make here use of a combination of share prices with dividends and the short interest rate and then there is clear evidence of predictability, also at a longer horizon, so this result comes not only from the short term interest rate. Both variables already have been shown to be needed together in Ang and Bekaert (2006) and also by Torous, Valkanov and Yan (2004)

Other remarkable research

Rangvid, Rapach and Wohar (2005) examine the predictability of stock returns using macro variables in 12 industrialized countries. One of the variables was the short term interest rate and they also conclude that this variable stand out in terms of predictive ability, as well in sample and across countries. They also state that interest rates are typically not subject to revision and are available immediately, so their results pertaining to interest rates are likely to be relevant in real time.

Schotman, Tschernig and Budek (2008) study a multivariate model with five predictive variables for excess returns on stocks and bonds. One of the variables is the short term interest rate. The persistence in interest rates leads to a strong positive correlation of the cumulative returns of long term bonds and T-Bills. This strong correlation for long horizons is independent of predictability for excess returns as long as the real return on T-bills exhibits long memory. Long term bonds and short term T-bills become close substitutes in the minimum variance portfolio of a long term investor. The nominal short-term interest rate is an important predictor of stock returns; the innovations to the interest rate have low correlations with unexpected returns. This means that the nominal interest rate does not create a mean reversion in equity returns.

In all of the above described results of researches they came up with the fact that there is evidence for predictive power of the short interest rate variable, especially on the short run. They combine also several variables in a multivariate regression to find more evidence for predictive power of the variable. The combination of the short term interest rate and the dividend price ratio is a very popular one. Another closely related variable to the short term interest rate that is often used in the most common researches is the term and default spread and will be discussed as the next explanatory variable.

Expectation about the sign

I expect that the coefficient of the short term interest rate would be negative because:

Ang and Bekaert (2006) found that at short horizons, the short rate strongly negatively predicts returns. When the economy expands, also the corporate earnings will expand. This growth comes from the fact that the Federal Reserve having begun a rate-cutting program many months earlier. When the interest rates are decreasing the consumption will grow and this leads to an expanding economy. High rates typically lead to slower demand and this will slow down the economy. Cochrane (2000) states that interest rates are high when consumption growth is high. Investors consume less now and consume more in the future. Thus high interest rates lower the level of consumption today, while raising its growth rate from today to tomorrow.

Investors are concerned about rising and falling interest rates, not only because of how they affect the growth of the economy but also for the reason of present values. The higher the interest rate with which one discounts future cash flow streams, the lower the present value of these streams. The relationship between stock prices and interest rates has received considerable attention in the literature. Fama (1988) argues that expected inflation is negatively correlated with anticipated real activity, which in turn is positively related to returns on the stock market. Therefore, stock market returns should be negatively correlated with expected inflation, which is often proxied by the short-term interest rate.

The term spread can be defined as the difference between the yield on long-term government bonds and the short-term interest rate (3-month T-bill). The default spread is often measured with a bond rating, and the interest rate that corresponds to the rating that is estimated by adding a default spread to the risk-free rate. This can be estimated by finding a sampling of bonds within that rating class and obtain the current market interest rate on these bonds. Second step is to identify the interest rate of the bonds itself (yield to maturity) and compute an average of the interest rates. Then the average interest rate can be compared against investment-grade corporate bonds, treasury bonds or another benchmark bond measure. When the spread is wide between bonds of different quality ratings, investors can conclude that the market is factoring more risk of default on the lower graded bonds. This implies that the economy is slowing down and the market predicts a greater risk of default (Damodaran (2013)).

According to Ross, Roll and Chen (1986) the financial theory suggests that the spread between long and short interest rates and the spread between high-and low graded bonds should systematically affect stock returns and that this sources of risk is significantly priced. This can be implied as a logical indication that the term and default spread can have predictive power for stock returns and makes these variables interesting for further research. They argue that the term spread and the default spread is a measure for business conditions.

Short-term predictive power

Fama and French (1989) found that the term spread is more closely related to the shorter-term business cycles. In particular the term spread is low around measured business cycle peaks and high near troughs. The results of the regressions they made show that the term spread and also the default spread have information about expected returns on stocks and bonds.

One of the variables that Rangvid, Rapach and Wohar (2005) take into account in their research (to examine stock return predictability) is the term spread. They did research across 12 countries and they found sufficient differences per country. For France they found some predictive ability at the 24-month horizon and for Sweden at a 12-month horizon. Also for the U.S. they found some evidence of predictability at certain horizons. When they put the results of all the countries together they conclude that there is some evidence in some countries, but it is not the best predictive variable. The overall conclusion that they made was that interest rates are consistent and reliable predictors of stock returns.

Torous, Valkanov and Yan (2004) found that the term spread seen to forecast returns at the relatively short horizon of less than 12 months. For the default spread they conclude that it does not has the ability to have predictive power to forecast excess/stock returns.

Chen (2008) investigates when macro variables can predict recessions in the stock market. One of the tested variables is the term spread and they did that on several horizons (1-month till 24 months). Term spreads produce consistently strong results across all horizons. The term spread is an important leading indicator for future recessions and contains information for stock returns. They also test on predictive power on stock returns and they found that the term-spread has definitely predictive power, especially at a 6-month horizon. Nevertheless they state that the term spread variable do a better job in predicting recessions by itself than predicting returns in the stock market. There is a logical explanation for this result, because there are more factors that have influence on the stock returns.

Long-term predictive power

Rangvid, Rapach and Wohar (2005) found also some in sample and out-of-sample predictive ability at longer horizons especially for Belgium. Off course they tested also at a longer horizon for all the countries in their research, but they did not found striking results that are significant. As already mentioned they state that there are better predictive variables, which are already mentioned in the previous described short-term results.

Fama and French (1989) demonstrate that the term spread and the default spread explain a substantial fraction of the long-term variation in bond and stock returns. They found that the default spread forecast high returns when business conditions are weak and low returns when conditions are strong.

According to Campbell (1986) term spreads reflects any predictable variation in excess returns and is therefore a powerful instrument. They make use of U.S. data for 1959-1983 and find that the state of the term structure of interest rates predicts excess stock returns. There are forecast able movements through time in excess returns on stocks and these movements are partially captured by a variety of term structure variables (term and default spread) which add considerably to the predictive power of short interest rates alone. The evidence they found for stock return predictability is very strong.

Campbell and Yogo (2004) made use of the pretest (which is also mentioned at the short interest rate variable) to test or the evidence of predictive power for the term spread variable is valid. They found that also for the term spread variable that the evidence for stock return predictability is valid.

Other remarkable research

Keim and Stambaugh (1986) show that the term spread can forecast stock and bond returns. They also state that it seems unlikely that combining the term spread with the dividend yield or the default spread captures all variation in expected returns. It follows that shocks to the term spread and default spread probably miss some of the adjustment of prices of expected returns because of shocks. It seems improbable that a single macro-variable can capture all variation in returns but it can give an indication what would be the most likely price change

Overall can be mentioned that also for the term/default spread there is evidence for stock return predictability, but it is important to look at the big picture. Term/default spreads have predictive power, but it is just part of the variation in expected returns. Especially Campbell (1986) is interesting because they found very strong evidence of predictive power from the term and default spread variables.

Another interesting research is done by Schotman, Tschernig and Budek (2008), which is earlier mentioned they study a multivariate model with five predictive variables for excess returns on stocks and bonds. The five different variables that they used are shown on the next page in figure (6):

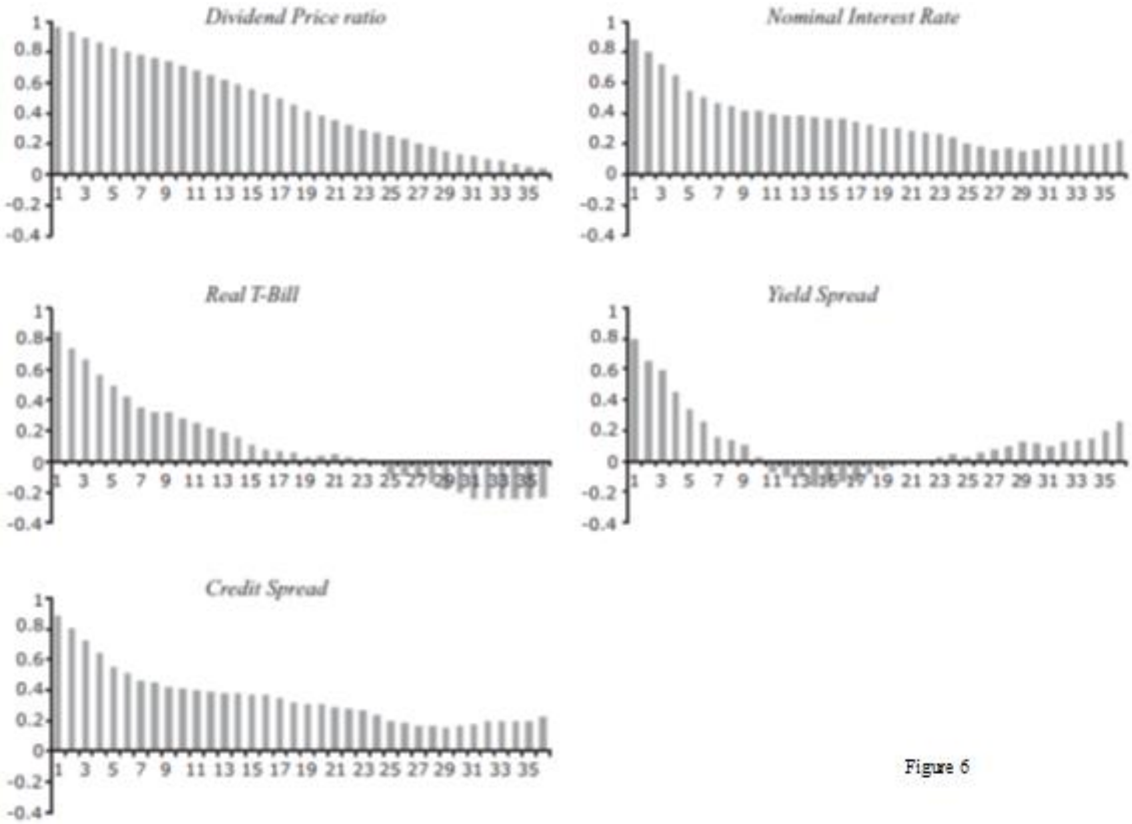


Figure 6

"This figure shows the autocorrelation functions of the five predictor series. The patterns for the nominal interest rate and the dividend-price ratio are typical for a long-memory process. Autocorrelations are large and decline very slowly. The real interest rate series looks somewhat less persistent: the autocorrelations decline faster and become even negative. The credit spread looks similar to the nominal interest rate, except that the first-order autocorrelation is much lower. The yield spread shows the least signs of persistence; its autocorrelations quickly decline as one would expect for a process with small d."

Important to see in the above figure is that there are huge differences per explanatory variable and nevertheless they are closely related to each other. This implies again that a sample of these explanatory variables can have more predictive power than just one explanatory variable. And this emphasizes the idea that for example Term/default spreads have predictive power, but it is just part of the variation in expected returns.

Expectation about the sign

I expect that the coefficient of the default spread would be positive because:

Fama and French (1989) found that the default spread captures similar variation in expected stock returns than de dividend yield which is earlier discussed. The major movements in these variables, and in the expected return components they track, seem to be related to long-term business episodes that span several measured business cycles. The dividend yield and the default spread forecast high returns when business conditions are persistently weak and low returns when conditions are strong.

Fama and French (1989) found that the slopes suggest that the default spread and the dividend yield track components of expected returns that vary with the level or price of some business-conditions risk. It is also appealing that a measure of business conditions like the default spread captures expected-return variation that increases from high-grade bonds to stocks in a way that corresponds to intuition about the business-conditions risks of assets.
One story for these results is that when business conditions are poor, income is low and expected returns on bonds and stocks must be high to induce substitution from consumption to investment. When times are good and income is high, the market clears at lower levels of expected returns. In one sentence the outcome of the sign from the default spread can be summarized; investors want to be compensated for taking more risk (because of a higher default spread) and expect a higher return on their investment.

I expect that the coefficient of the term spread would be also positive because:

Fama and French (1989) state that the term spread are more closely related to the shorter-term business cycles. In particular, the term spread is low around measured business-cycle peaks and high near troughs. In their regressions they found clear patterns across assets in the slopes from regressions of returns on the forecasting variables. The slopes for the term spread are positive and similar in magnitude for all the stock portfolios. This suggests that the spread tracks a term or maturity premium in expected returns that is similar for all long-term assets. A reasonable and old hypothesis is that the premium compensates for exposure to discount-rate shocks.

Chen (1989) documents the clear impression that the Term spread is positively related to future real activity. Since the term spread is low near business-cycle peaks and high near troughs. Chen’s results suggest that poor prospects for future real activity (and thus investments) near business peaks may help explain low expected returns around peaks. Likewise, good prospects for future activity and investment after business troughs may contribute to high expected returns around troughs.

3.5 CAY (CONSUMPTION-WEALTH RATIO)

The consumption-wealth ratio as explanatory variable comes from a study of Lettau and Ludvigson (2001). They make use of a new approach to investigate the link between macroeconomics and financial markets. An important note that they make is that aggregate consumption, asset holdings and labor income share a common long-term trend, but it may deviate substantially in the short run. They study the role of these transitory deviations from the common trend in consumptions for predicting stock market fluctuations. Their results show that the so called trend deviations are strongly univariate predictors of stock and excess returns over a T-bill rate and can account for a considerable fraction of the variation in future returns. This variable provides information about future stock returns that is not captured at the earlier discussed variables, and the cay variable has the greatest predictive power for returns from one to five quarters. In their study they also find that this variable would have improved out-of-sample forecasts of excess stock returns relative to alternative forecasting models, but it is important to keep in mind that they bear only little relationship to future stock returns.

To link the consumption-aggregate wealth ratio with future asset returns, there were two obstacles. The first obstacle was that aggregate wealth is unobservable. In their research they solve this by arguing that important predictive components of the aggregate wealth ratio may be expressed in terms of observable variables like consumption, asset holdings and current labor income. In the model they use, they show a common stochastic trend between the observable variables that are cointegrated. The parameters of this trend are the average shares of human capital and asset wealth in aggregate wealth. Deviations from the shared trend among the three variables (consumption, asset holdings and labor income) produce movements in the consumption-aggregate wealth ratio and predict future returns.

The second obstacle to use deviations in the common trend among consumption, labor income, and asset wealth is that the parameters of the shared trend are unobservable and also must be estimated. How Lettau and Ludvigson (2001) solve this, is not described in this thesis and can be written in Lettau and Ludvigson (2001)

The equations that Lettau and Ludvigson (2001) used to come up with the final equation for the aggregate consumption-wealth ratio are also not described in this thesis, because there are simply too many steps needed and it is hard to explain in a short way. The final equation and a detailed step plan how the equation is estimated can be written in Lettau and Ludvigson (2001).

The economic idea/interpretation behind the consumption aggregate wealth ratio is that investors who want to maintain a flat consumption path over time will try to “smooth out” transitory movements in their asset wealth arising from time variation in expected returns. When excess returns are expected to be higher in the future, forward looking investors will react by increasing consumption out of current asset wealth and labor income, allowing consumption to rise above its common trend with those variables. When excess returns are expected to be lower in the future, these investors will react by decreasing consumption out of current asset wealth and labor income, and consumption will fall below its shared trend with these variables. In this way, investors may insulate future consumption from fluctuations in expected returns, and stationary deviations from the shared trend among consumption, asset holdings, and labor income are likely to be a predictor of excess stock returns.

Lettau and Ludvigson (2001) used U.S. quarterly stock market data and they find that the fluctuations in the consumption-wealth ratio are strong predictors of stock returns and excess returns over a T-bill rate. They also state that this variable is a better predictor of future returns at short and intermediate horizons than for example the dividend yield.

One of their results is shown in figure 4, they plot the standardized trend deviation Cayt, and the excess return on the S&P over the period 1952-1998. This figure shows multiple episodes where positive trend deviations preceded large positive excess returns and negative episodes preceded large negative returns.

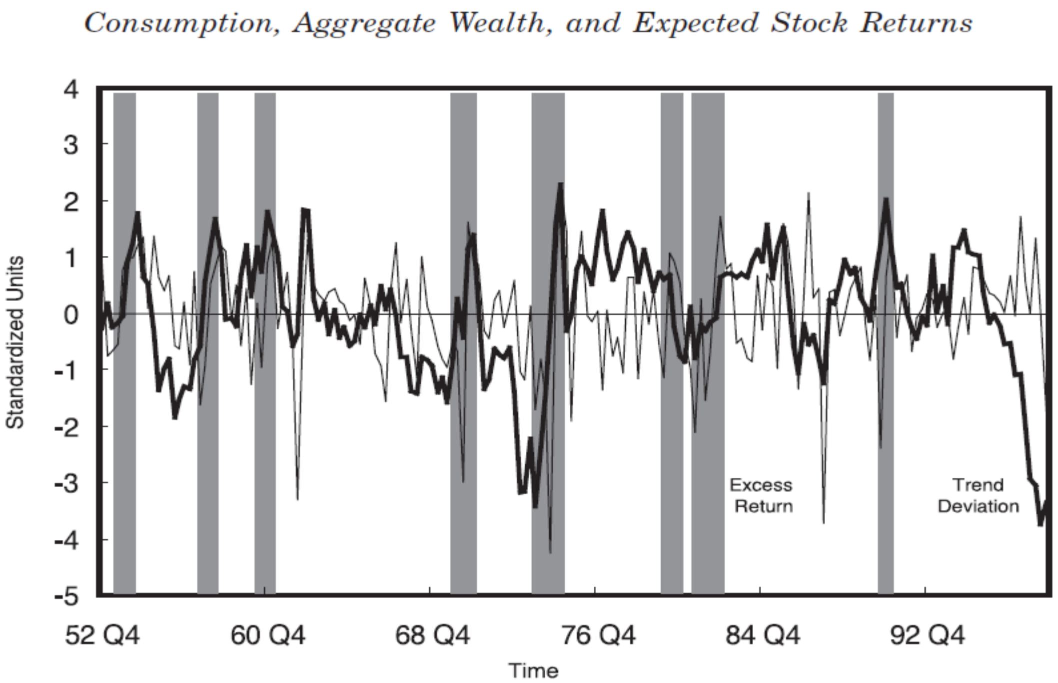


Figure 7

Most striking in the above figure (figure 7) is the steeply downward slope of the trend deviation around the year 2000. The values of Cay around 2000 were an indication of an extreme drop of excess returns. In the meantime it is obvious what happened with stock/excess returns and we can conclude that the Cay variable was a leading indicator to do predictions on stock returns.

The most important conclusion that Lettau and Ludvigson (2001) made is that the Cay variable is the best univariate predictor of stock returns for horizons up to one year. They also test for longer horizons, but these results are not significant and the r square was close to zero. Despite their findings they mention that the results do not imply forecastabillity in all periods.

Expectation about the sign

I expect that the coefficient of the Cay variable would show a positive sign, because:

First a negative cay value indicates that consumption is below that predicted given the current wealth of the investor. This measure would be consistent with individuals reducing their current consumption when they anticipate lower returns in the future. They reduce their current consumption and in the future they will increase their consumption again. Second a positive cay measure can be observed when they increase current consumption in anticipation of higher returns in the future.

Lettau and Ludvigson (2001) state that the Cay variable is positively correlated with excess stock returns. In figure 7 they show multitude of episodes during which positive trend deviations preceded large positive excess returns and negative ones preceded large negative returns. The trend deviation term also displays some notable cyclicality, typically rising during recessions and falling during booms. Investors own optimizing behavior suggests that deviations in the long-term trend among c, a, and y should be positively related to future stock returns, consistent with their findings. Not only does cay covary positively with expected future returns, the variation in cay is countercyclical: cay tends to decline during expansions and rise just prior to the onset of a recession.

Expansions are characterized by increasing consumption, but an even greater rate of increase in assets. Consumption booms are periods during which consumption increases above habit, leading to a decline in risk aversion. The decline in risk aversion leads, in turn, to a greater demand for risky assets and a decrease in expected excess returns. Thus, booms are times of rising consumption but declining ratios of consumption to wealth, consistent with the positive relation findings.

The last explanatory variable that will be discussed is the so called Fed-model. The Fed-model is based on the stock-bond yield gap. The difference between the stock market earnings yield (or dividend yield) and the 10-year Treasury bond yield (Maio 2013).The Fed model assumes that the dividend earnings yield on stocks should be equal to the yield of nominal treasury bonds, or at least that the two should be highly correlated (Bekaert and Engstrom 2010).

According to Asness (2003) the Fed model is a popular yardstick for judging whether the stock market is fairly valued. The Fed model compares the stock markets earnings yield on long term government bonds. This method is distinctive from the other explanatory variables because most of the other variables are on its own, without regard to the level of interest rates.

Maio (2013) takes another approach than most of the literature on the Fed model. In earlier research the focus often on the correlation between stock and bond yields and whether the two yields should be approximately equal or strongly correlated, Maio 2013 focusses on the forecasting ability of the yield gap for the aggregate equity premium. By using the definition of stock return, Maio derives a dynamic accounting decomposition for the yield gap based on the earnings yield, as a function of future equity premiums. In his study he split it up in short-run and longer horizon predictions.

Short-term predictive power

The results of Maio (2013) at the short horizon are striking; at a 1-month horizon the yield gap has significantly more forecasting power for the equity premium than the earlier discussed earnings-price ratio and dividend price ratios and term/default spread.

Asness (2003) did also research to the Fed-model as predictive variable and created also a couple regressions to test the evidence of the Fed-model as explanatory variable. When he looked at a short horizon (1 year) the R-squared values fall dramatically. The reason that he gave is because of the predictable component of stock returns is small but slowly changing, which leads to reasonable reliable long-term forecasts, but poor short term forecasts. This research will be described more in detail when the long term predictive power of this variable will be discussed.

Long-term predictive power

Maio (2013) found not only evidence on the short-term, but also when he tested for a longer horizon. The results from long-horizon regressions also show that the yield gap has significant forecasting power. He found significant forecasting power at horizons between 3 months and 5 years.

Overall the results show that the yield gap forecasts positive excess market returns at the short and long horizons for value-and equal weighted stock indexes and it outperforms the earlier discussed predictors like for example the P/E ratio, dividend yield, term and default spread. This makes the results even more valuable. By performing an out-of-sample analysis the results show that compared to historical averages the yield gap has reasonable out-of-sample predictability for the equity premium. Like earlier mentioned Maio (2013) state that the yield gap has greater forecasting power than the alternative variables and an investment strategy based on the forecasting ability of the yield gap produces significant gains in sharp ratios.

There is also research done by Asness (2003) with very different results. Asness (2003) found empirical evidence that the Fed model has no power to forecast long-term stock returns, contrary to some other traditional methods like the E/P ratio which is very effective. To prove this, Asness (2003) did a test where the data decides which method is historically seen a better tool to forecast stock returns; the earnings-price ratio or the relative valuation of the Fed Model (E/P-Y). The way to test this was by running three regressions over different horizons; a one year, ten year and 20 year horizon. On the left hand side of the regressions the S&P 500 was used at different horizons. Asness (2003) run three different regressions; the first one was a regression with on the right the E/P ratio of the S&P 500. The second regression that he did was based on the E/P ratio of the S&P 500 minus the 10 year Treasury bond yield (Fed Model). In the last regression he combined the E/P ratio and the 10 year Treasury bond yield separately in a two-variable regression. The idea behind these three different regressions is simple, when the E/P ratio and/or the Fed model have forecasting power it should show up in the single variable regressions. The bivariate regression is useful because E/P –Y can also have statistical power even if only E/P itself has actual efficacy, this is possible because E/P-Y can be a noisy measure of E/P. Another strong point at this research is that there are used different time periods to run the above regressions. He runs the regressions over 181-2001, 1926-2001 and 1955-2001.

Analyzing the results Asness (2003) conclude that at a long horizon (10 year and 20 year) the results are quite the same. The earnings-price ratio has strong forecasting power for stock market returns and for the Fed model there is no evidence of forecasting power. The Fed model seems to have some power, but this is only because the earnings-price ratio is part of the Fed Model (E/P-Y). When E/P and Y are tested together in a two variable regression, only E/P matters and the Y part of the Fed model can be ignored.

Overall the conclusion of Asness (2003) is that the Fed Model has no power to forecast stock returns. This result is in contrast to the results of Maio (2013), the most common explanation is that Maio (2013) made only use of a dataset of one time period and only tested at shorter horizons. It would be interesting to check it again in the empirical part of this paper, by using the most recent data and test for different horizons.

Expectation about the sign

I expect that the coefficient of the Fed model would show a positive sign, because:

Maio (2012) founds that the yield gap forecasts positive excess market returns, both at short and long forecasting horizons Conditional on other forecasting variables, the yield gap forecasts positive stock excess returns, at several horizons ahead. The shocks in the yield gap are highly positively correlated with innovations in both future discount-rate and cash flow news, confirming that the spread conveys information about future earnings and returns. An investment strategy based on the forecasting ability of the Yield gap produces higher Sharpe ratios than passive strategies in the market index.

Hasseltoft (2009) argued that the Fed model attributes a large part of changes in realized correlations between stock returns and bond returns to changes in macroeconomic risk. High volatility of consumption growth and inflation caused stock returns to commove strongly. Risk premiums on equity react in a similar way to changes in macroeconomic volatility, making their returns positively correlated.

Maio (2008) says that high values of the Yield gap (earnings yield is high relative to the bond yield), are associated with expected combinations of higher future stock returns, lower dividend to earnings payout ratios, lower growth rate on future equity earnings. Thus the Fed model forecasts higher expected stock market returns. The fact that the yield gap is positively correlated with discount rate news arises from the positive correlation of the earnings yield to future returns. The fact that the yield gap is also positively correlated with cash flow news might be attributable to a negative correlation between current shocks in bond yields and future cash flows.

4. DATA

In this chapter the several data sources that are used will be described. The development of the several hypotheses will be shown and the definitions of the variables will be explained. From all the used variables the descriptive statistics will be shown and finally there will be given an explanation of the methodology that is used to test the hypotheses.

4.1 SOURCES

The first part of the research – the literature review – comes from current existing literature in the finance world, to provide a theoretical framework on which to base the analyses. The literature will be found in the several databases, for which access is provided by the University of Tilburg Library, like JSTOR and Wiley Online Library. The most articles that are used in this research are published in top financial journals. The most important journals in finance that are used in this research are the journal of finance, Journal of financial economics and the Review of financial studies. In this research also some important journals in economics are used, like Journal of monetary economics, Quarterly journal of economics, Journal of political economy and the American economic review.

The empirical data that is used in this research comes from different sources and is merged to one dataset, where the regressions are based on. The biggest part of the data comes from the online available data of Professor Robbert J. Shiller (Yale University). The index file of the S&P 500 comes from Robbert. J. Shiller and is also compared with the CRSP dataset which is available on the Wharton research data services. By doing that the exact date of the monthly data of Robbert J. Shiller is found and the interest rates could be accurately linked to this data. The interest rates come from the Federal Reserve Bank of St. Louis and also the term and default spread are based on this data. The data for the Cay variable is online provided by Proffesor Martin Lettau (California University).With the available data from the above mentioned sources, all the variables to test the hypothesis are available or could be calculated. In the paragraph variable description will be described how each variable is calculated.

4.2 HYPOTHESIS

In this paragraph the hypotheses will be formulated, which will help to answer the main research question. The idea of the hypotheses is very simple, because in this research 6 variables are taken into account. For each variable there is made a sub question and by putting all the results together, the main research question can be answered. The hypotheses may not be formulated like the sub questions, but must contain an expected outcome that can be tested. The factors mentioned in the research questions are divided into six hypotheses which are used as proxies to answer the main research question.

Hypothesis 1: the dividend yield has significant positive predictive power to forecast index returns.

Hypothesis 2: the price-earnings ratio has negative significant predictive power to forecast index returns.

Hypothesis 3: the short term interest rate has negative significant predictive power to forecast index returns

Hypothesis 4: the Term/default spread has positive significant predictive power to forecast index returns.

Hypothesis 5: the Cay ratio has positive significant predictive power to forecast index returns.

Hypothesis 6: the Fed model has positive significant predictive power to forecast index returns.

By testing the above mentioned hypotheses, each variable will be tested separately; it is interesting to combine some variables and check what the effect will be in contrast to the effect of a variable on a standalone bases.

Hypothesis 7: the dividend yield, price-earnings ratio, Term/default spread and the short term interest rate together have significant predictive power to forecast index returns.

Hypothesis 8: the Cay ratio, Fed model and the short term interest rate have significant predictive power to forecast index returns.

In this research the most common predictive variables are taken into account, the last hypothesis is a combination of all the earlier described variables. Putting the results of the separately tested variables, the sub combined variables and a total test of all the variables together must lead to a reliable answer to the main research question.

Hypothesis 9: : the dividend yield, price-earnings ratio, Term/default spread, Cay ratio, Fed model and the short term interest rate together have significant predictive power to forecast index returns.

4.3 VARIABLE DESRCIPTION

In this paragraph the definitions of the different variables are given and when it is applicable, also the way how the variables are calculated will be described. This is necessary to continue with the empirical analysis of the data. First the definition of the dependent variable will be provided, second the definition of the independent variables will be provided and finally the control variable that is used will be explained.

Dependent variable

The dependent variable in this research is based on the returns on the S&P 500 index on a monthly basis.

Independent variables

The independent variables in this research are quite obviously; these are the dividend yield, price-earnings ratio, short term interest rate, term/default spread, Cay variable and the Fed model.

The dividend yield and the price-earnings ratio are based on the dividends and earnings of the US stock market. The short term interest rate is based on the three month Treasury bill rate. Based on the data of the Federal Reserve Bank of St. Louis the term spread and the default spread are conducted. The Term spread is calculated by calculating the difference between the short term interest rate and the 10 year Treasury bill rate and the default spread is based on the BAA corporate bond yield and the triple A corporate bond yield. For calculating the data for the Fed-model the earnings-price ratio is used minus the 10 year Treasury bill rate. In this research is chosen not to calculate the Cay variable by itself, but this data comes from Proffesor Martin Lettau (California University, because of that the data is only available at a quarterly basis.

4.4 MULTICOLLINEARITY

Before continuing with the descriptive statistics, also correlations between the independent variables must be considered to control for any linear dependency. In the basic regressions with only one independent variable, the multicollinearity problem does not exist, but when there are more independent variables in a regression multicollinearity can occur. Multicollinearity can occurs when there is a linear relationship among one or more of the independent variables. Intuitively the problem arises because the inclusion of an extra independent variable adds not more information to the model than the inclusion of just one independent variable. Another way to interpret this is that the regression model is asked for an additional parameter, without supplying additional information (Golder University of Pennsylvania).

To check already for correlations between the dependent and independent variables in table 1 is shown a correlation matrix. The expectation was that the correlations of all the variables in the correlation matrix would be quite high; because the idea is that the independent variables are closely related to the state of the economy. Therefore these variables could be good predictors for the stock market index. Especially the P/E ratio, the dividend yield and the short term interest rate show high correlations. When these variables are one by one closely related to the state of the economy it is also reliable that those variables are closely related to each other.

Table 1

	S&P returns	P/E ratio	Dividend yield	Short-term Interest Rate	Fed model	Default spread	Term spread	CAY
S&P	1							
P/E ratio	-0.0021	1						
Dividend yield	0.1001	-0.4235	1					
Short term IR	-0.0095	-0.4911	0.717	1				
Fed model	0.1741	-0.3313	0.1766	-0.1724	1			
Defaultspread	0.0575	0.151	0.454	0.2168	-0.0765	1		
Termspread	-0.0035	0.242	-0.093	-0.4848	-0.2513	0.2536	1	
CAY	0.1069	-0.055	0.0712	0.2288	-0.4894	-0.1822	0.1147	1

The most common way to solve the multicollinearity problem is to drop variables out of the model. The only hypotheses where there is a potential multicollinearity problem are hypotheses 4, 7,8 and 9, because there will be used multiple regression models. The multicollinearity problem can be detected by estimating a variance inflation factor. The Variance inflation factor tells us the extension to which the standard error of the coefficient of interest, the variance (that is the square of the standard error) has been inflated upwards. A rule of thumb is that the standard error will not have been inflated more than twice its basic size. When the values of the VIF are equal to 4 our greater there is a reliable chance of multicollinearity. By estimating the VIF for hypotheses 4, 7, 8, and 9, only at hypothesis 9 there is a problem, which is shown in the left hand side of table 2 where all the variables are taken into account. The VIF of the short term interest rate, the dividend yield and the fed model is problematic high because they have values above 4. In the middle of the table the results are shown where the short term interest variable is dropped out of the model, the VIF of all the variables is below 4, which implies that the multicollinearity problem is canceled out. To be as complete as possible for testing on predictive power, on the right hand side of the table not the short term interest rate is dropped out of the model, but the dividend yield. Also these results have a value below 4, so also in this case the multicollinearity problem is cancelled out. This will be taken into account by testing the regression model with as much as possible independent variables into the model (Hamrick University of San Francisco).

Table 2

	All variables		**Short term IR dropped**		**dividend yield dropped**	
Variable	**VIF**	**1/VIF**	**VIF**	**1/VIF**	**VIF**	**1/VIF**
Short term IR	12.28	0.08141	-	-	3.92	0.254977
Dividend yield	6.45	0.154971	2.06	0.485372	-	-
Fed model	4.72	0.211917	1.71	0.585764	2.22	0.451084
Term spread	4.31	0.23178	1.21	0.828862	2.5	0.399646
Default spread	1.99	0.501285	1.92	0.521369	1.92	0.519971
P/E ratio	2.64	0.378994	1.59	0.628791	2.37	0.421162
CAY	1.65	0.606238	1.65	0.606238	1.6	0.62385
Mean VIF	4.86		1.69		2.42	

4.5 DESCRIPTIVE STATISTICS

The following descriptive statistics provide an overview of the dependent and independent variables that are being used in the regressions. The numbers of observations for all variables are 576, except for the Cay variable. There are only 192 observations available for the Cay variable, because this data is only available on a quarterly basis and all the other variables provide data on a monthly basis.

Table 3

Variables	**Obs**	**Mean**	**Median**	**Std. Dev.**	**Min**	**Max**
Dependent						
S&P returns	576	0.0059	0.0086	0.0440	-0.2176	0.1630
Independent						
P/E ratio	576	18.9018	17.0691	12.8964	6.7876	123.6575
Dividendyied	576	0.0304	0.0302	0.0119	0.0111	0.0624
Short term IR	576	0.0526	0.0506	0.0307	0.0001	0.1630
Defaultspread	576	0.0106	0.0094	0.0046	0.0032	0.0338
Termspread	576	0.0155	0.0161	0.0130	-0.0265	0.0442
Fed model	576	-0.0021	-0.0066	0.0213	-0.0444	0.0583
CAY	192	-0.0004	0.0117	0.0187	-0.0387	0.0397

4.6 METHODOLOGY

The above formulated hypotheses will be tested with the basic linear regression models. The first six hypotheses can be tested with a simple linear regression with one dependent variable: x_i and two parameters, β_0 and β_1 which leads to the following basic model:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \ldots, n.$$

For testing the hypotheses with more than one explanatory variable there will be made use of simple multiple regression models with one dependent variable: x_i and more predictor variables, which leads to the following basic multiple regression model

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i + \varepsilon_i, \; i = 1, \ldots, n.$$

The above formula is based on two predictive variables, it is obvious that there can be added more predictive variables to the model. As mentioned in the paragraph variable description, in this research also a control variable will be taken into account.

The above described basic regression models can be applied to the whole dataset. The idea is to make a distinction between short term and long term predictive power by creating regressions on different horizons. There is chosen to create the above models at the following time horizons, 1 month, 6 months, 1 year, 3 years and 6 years. This is very easy applicable by shifting up the data from the independent variables and compare them with the returns on the S&P 500.

For analyzing the regressions, the marginal effect (dy/dx) will be presented to interpret the variable coefficients and to show the economic significance.

4.6.1 ORDINARY LEAST SQUARES (OLS)

Ordinary Least Squares (OLS) regressions make several assumptions. These assumptions and checks are necessary to create a good and unbiased estimation model for the actual population. Whether or not the OLS estimator b provides a good approximation to the unknown parameter vector β depends crucially upon the assumptions that are made about the distribution of εi and its relation to *xi*. The conditions that are important are also called the Gauss-Markov conditions. The theory is based on Verbeek (2008) and the assumptions are:

- The first assumption is that there is no linear dependency. This means that none of the xi's (for i=1,..,n) is an exact linear combination of the other xi's. When the independent variable would be correlated with another, the coefficient would not be very accurate to the actual population parameter because of the high variability in the coefficients. The actual strength of the relationships can be underestimated, which may lead to faulty inferences about the relationship between the independent and dependent variables Multicollinearity makes the coefficients difficult to interpret. To encounter the (possible) linear dependency of the independent variables, correlation matrices are created and consequently in some models it was necessary to exclude a variable from the regression.
- The second assumption is that εi…εn) and x1,…,xn are independent and that there are no outliers. Extreme cases can have a large impact on the regression line and change the results significantly. In this research extreme outliers are not really there, some values are striking, but they are all explanatory because of the economic situation. In this research it is important not to realize to test for predicting the index not only in good times but also in turbulent times.

- The third assumption about the error terms is that its variance is a constant, $\sigma 2$, in every observation ($V(\varepsilon i) = \sigma 2$ for $i = 1,..,n$). This is called homoscedasticity and this is taken into account by using the robustness commands while running the regression. This will make sure that the model performs properly even when the OLS assumptions would be somewhat violated.
- The fourth assumption for OLS regressions is that of the absence of autocorrelation of the error terms ($cov\ ,\varepsilon i,\varepsilon j) = 0$ for $i,j = 1,...,n$ and $i \neq j$). This occurs most often when the data are time series, since the error terms will then have a high correlation to each other. In this research in some models there is a variable omitted and any noise that may be caused by possibly omitted variables is picked up by the error term, it may cause autocorrelation. A further assumption regarding the error terms is that they are on average zero ($E(\varepsilon i) = 0$ for $i = 1,...,n$). If the average error term would not be zero, it would mean that the estimation could be improved by adding or subtracting the same constant for each observation. However, when the expected error term is zero, it means that the regression line is correct on average.

4.6.2 MODELS

This paragraph contains the specific models that will be used to test the earlier mentioned hypotheses. For each hypothesis the regression model(s) will be described, also the dependent, independent and the control variable will be shown.

Hypothesis 1: $= \beta 0 + \text{Dividend yield} * \beta 1 + \varepsilon i$

Hypothesis 2: $= \beta 0 + \text{Price-earnings ratio} * \beta 1 + \varepsilon i$

Hypothesis 3: $= \beta 0 + \text{Short term interest rate} * \beta 1 + \varepsilon i$

Hypothesis 4: $= \beta 0 + \text{Term spread} * \beta 1 + \varepsilon i$
Hypothesis 4: $= \beta 0 + \text{Default spread} * \beta 1 + \varepsilon i$
Hypothesis 4: $= \beta 0 + \text{Term spread} * \beta 1 + \text{Default spread} * \beta 2 + \varepsilon i$

Hypothesis 5: $= \beta 0 + \text{Fed model} * \beta 1 + \varepsilon i$

Hypothesis 6: $= \beta 0 + \text{Cay variable} * \beta 1 + \varepsilon i$

Hypothesis 7: = β0 + Dividend yield * β1 + Price-earnings ratio * β2 + Term spread * β3 + Default spread * β4 + Short term interest rate * β5 + εi

Hypothesis 8: = β0 + Fed model * β1 + Cay variable * β2 + Short term interest rate * β3 + Control variable Crisis + εi

Hypothesis 9: = β0 + Dividend yield * β1 + Price-earnings ratio * β2 + Term spread * β3 + Default spread * β4 + Fed model * β5 + Cay variable β6 + Short term interest rate * β7 + εi

Hypothesis 9: = β0 + Dividend yield * β1 + Price-earnings ratio * β2 + Term spread * β3 + Default spread * β4 + Fed model * β5 + Cay variable β6 + εi

Hypothesis 9: = β0 + Short term interest rate * β1 + Price-earnings ratio * β2 + Term spread * β3 + Default spread * β4 + Fed model * β5 + Cay variable β6 + εi

Hypothesis 9: = β0 + Dividend yield * β1 + Price-earnings ratio * β2 + Term spread * β3 + Default spread * β4 + Fed model * β5+ εi

Hypothesis 9: = β0 + Short term interest rate * β1 + Price-earnings ratio * β2 + Term spread * β3 + Default spread * β4 + Fed model * β5 + εi

The above models (for each hypothesis) will be tested at different time horizons as already mentioned in paragraph 4.6 methodology.

5. RESULTS

In this chapter the regression results from the basic OLS regressions which were described in the previous chapter will be discussed. Also the tables with the results are shown. The structure of this chapter is based on the hypotheses of this research; the results will be shown in the same order as the hypotheses are constructed. First the results of the simple linear regressions with one explanatory variable will be described and finally also the results of the multiple regressions will be shown and discussed.

5.1 RESULTS DIVIDEND YIELD

The results of the regression with only the dividend yield as independent variable shows positive coefficients at all the tested horizons. This is in line with the expectation that is already described in paragraph 3.1. By comparing the different time horizons, there are no big differences between them, where the literature findings imply more predictive power at longer horizons.

The coefficients show a rather high positive value at all horizons which imply a strong effect of the dividend yield as explanatory variable. By testing the results on economic significance (Coefficient * Standard deviation independent / Standard deviation dependent) the strong effect can be interpreted as an existing strong relation. There is no specific rule of thumb to state at which value the economic significance can be interpreted as high, this depends on the type of research. In this case the economic magnitude is rather high, because the effect on returns is in common not that high, otherwise the returns would be so predictable that everyone would earn money by doing investments based on a predictive variable like this.

Table 4

Dependent: S&P returns	1 month	6 months	1 year	3 year	5 year
Dividend yield	0.2329	0.3057	0.2524	0.1862	0.2951
	(1.41)	(1.94)*	(1.62)	(1.17)	(1.78)*
Economic significance	0.063	0.083	0.068	0.049	0.078
Constant	-0.0012	-0.0034	-0.0018	0.0001	-0.0034
	(-0.23)	(-0.66)	(-0.35)	(0.02)	(-0.62)
Number of obs	576	576	576	576	576
F	1.98	3.78	2.62	1.38	3.18
Prob >F	0.1597	0.0525	0.1062	0.2412	0.0752
R-squared	0.004	0.0069	0.0046	0.0024	0.006
Root MSE	0.0439	0.04384	0.4389	0.04393	0.4386

The results must be interpreted carefully, because the values of the R-squares are very low. This implies that the goodness of fit of the used model is not very high and therefore the coefficients are not really valuable. Also the t values cannot be interpreted as statistically significant at the 5% level. Only the outcome of the 6 month and 5 year horizon regression is statistically significant at the 10% level. Nevertheless the outcomes give the indication that the effect (of the dividend yield as predictive variable) is positive on returns, which is in line with the literature findings.

5.2 RESULTS PRICE-EARNINGS RATIO

The results of the regression with only the price-earnings ratio as independent variable shows a small positive coefficient at a 1 month horizon and for all the tested horizons longer than one month a small negative coefficient. The expectation from the literature findings was a negative relation and except the 1 month horizon regression the findings are in line with the expectation of a negative coefficient which is described in paragraph 3.2.

Table 5

Dependent: S&P returns	1 month	6 months	1 year	3 year	5 year
Price-earnings ratio	0.000023	-0.000012	-0.000106	-0.000033	-0.000263
	(0,13)	(-0,09)	(-0,61)	(-0,28)	(-1,14)
Economic significance	0.006690	-0.003521	-0.031114	-0.009785	-0.041333
Constant	0.005457	0.006115	0.007899	0.006524	0.010485
	(1,49)	(1,98)	(2,15)	(2,11)	(2,29)
Number of obs	576	576	576	576	576
F	0.02	0.01	0.37	0.08	1.31
Prob >F	0.89890	0.92970	0.54480	0.78120	0.25370
R-squared	0.00000	0.00000	0.00100	0.00010	0.00170
Root MSE	0.04399	0.04399	0.04397	0.04399	0.04395

Nevertheless the coefficients are at all horizons so small that, there is almost no (positive or negative) effect. Therefore the calculated economic significance is also very low and negligible. Based on this simple linear regression test it is not possible do to some valuable statements on the price-earnings ratio as predictive variable. The results are not statistically significant even not on the 10% level and the R-squares are (at all horizons) close to zero.

The literature findings, that the price-earnings ratio has more explanatory power at short horizons than at longer horizons cannot be confirmed by using these test results. The question arises or there is predictive power, because this results show no relation. The only statement that can be confirmed is that the dividend yield has more predictive power than the price-earnings ratio and this is in line with the literature findings.

5.3 RESULTS SHORT TERM INTEREST RATE

The results of the regression with only the short term interest rate as independent variable in the model shows a negative coefficient at a 1 month horizon and for all the tested horizons longer than one month a positive coefficient. These results are striking, because the expectation was that the sign of the coefficients would be negative and this is only the fact at a 1 month horizon. The only interpretation for the positive sign that I can give is that a rising interest rate, still expects to rise (in this data sample) and therefore investors wants to be compensated with a higher return.

By analyzing the coefficients the 0.1457 from the regression at a 5 year horizon is most striking. The coefficient is rather high, especially compared to the regression results at shorter horizons. Also the economic significance at the 5 year horizon can be called high with a value of 0.0904. For the other horizons the results are less powerful. The results are striking because of two reasons. First the sign of the coefficients is not what I expected and second the predictive power at a longer horizon is bigger than over short horizons.

Table 6

Dependent: S&P returns	1 month	6 months	1 year	3 year	5 year
Short term interest rate	-0.0052	0.0433	0.0391	0.0376	0.1457
	(-0,08)	(0,7)	(0,63)	(0,63)	(2,06)**
Economic significance	-0.0036	0.0299	0.0267	0.0243	0.0904
Constant	0.0062	0.0036	0.0038	0.0038	-0.0022
	(1,65)	(0,98)	(1,05)	(1,06)	(-0,5)
Number of obs	576	576	576	576	576
F	0.01	0.48	0.39	0.39	4.26
Prob >F	0.9342	0.4866	0.5306	0.5318	0.0395
R-squared	0	0.0009	0.0007	0.0006	0.0082
Root MSE	0.0440	0.0440	0.0440	0.0440	0.0438

By checking the results on statistically significance only the result of the regression at a 5 year horizon is significant at the 10% level. Nevertheless the results must be interpreted carefully, because the R-squares are at all horizons close to zero and this states that the goodness of fit from the model is low.

In this paragraph the results of two variables will be combined. First the results of both regressions with one independent variable will be shown. Finally both variables are combined in a multiple regression model and also these results will be presented in this paragraph.

The results of the regression with only the term spread as independent variable in the model shows a positive coefficient at all horizons, except for the coefficient at a 5 year horizon. There can be stated that this is for the majority in line with the expectation based on the literature findings. The economic significance is especially high for the regressions with a maximum horizon of three years and at a 1 year and three year horizon the results are statistically significant at the 10% level.

Table 7

Dependent: S&P returns	1 month	6 months	1 year	3 year	5 year
Term spread	0.1837	0.1629	0.2619	0.2085	-0.0713
	(1,29)	(1,15)	(1,88) *	(1,65)*	(-0,46)
Economic significance	0.054	0.048	0.078	0.060	-0.020
Constant	0.0030	0.0034	0.0019	0.0029	0.0069
	(1,06)	(1,14)	(0,6)	(1,05)	(2,58)
Number of obs	576	576	576	576	576
F	1.67	1.32	3.53	2.56	0.21
Prob >F	0.1972	0.2514	0.0608	0.1099	0.6471
R-squared	0.003	0.0023	0.0061	0.0037	0.0004
Root MSE	0.0439	0.0439	0.0439	0.0439	0.0440

The results of the regression model with the default spread as independent variable shows kind of the same pattern as for the term spread. The coefficients are positive at all horizons and this is in line with the expectation that is described in paragraph 3.4. Also the economic significance is comparable to the results of the term spread. Striking is that at a 1 year horizon the economic significance is the highest with the term spread as predictive variable and with the default spread as predictive variable the economic significance is the lowest at a 1 year horizon.

Another important thing to mention is that the results of the default spread are not statistically significant, even not on the 10% level. For both of the models (term and default spread), the goodness of fit is low, thus again the results must be interpreted carefully.

Table 8

Dependent: S&P returns	1 month	6 months	1 year	3 year	5 year
Default spread	0.5940	0.5918	0.2155	0.4937	0.4907
	(1,2)	(1,54)	(0,61)	(1,45)	(0,84)
Economic significance	0.063	0.063	0.023	0.054	0.046
Constant	-0.00043	-0.00036	0.00363	0.00080	0.00107
	(-0,08)	(-0,08)	(0,88)	(0,2)	(0,18)
Number of obs	576	576	576	576	576
F	1.43	2.37	0.37	2.09	0.71
Prob >F	0.2324	0.1245	0.5407	0.1485	0.4004
R-squared	0.0039	0.004	0.0005	0.0029	0.0021
Root MSE	0.0439	0.0439	0.04398	0.04393	0.04394

By combining the default spread and the term spread in a multiple regression model the outcomes (table 9) are not much better. The R-squares are still very low, so by pulling those variables together in one model the model does still not fit very well. The results from the multiple regression models show similar coefficients than those of the single regression models.

Table 9

Dependent: S&P returns	1 month	6 months	1 year	3 year	5 year
Default spread	0.4940	0.5070	0.021319	0.3631	0.5776
	(0,99)	(1,32)	(0,06)	(1,04)	(0,98)
Economic significance	0.0522	0.0539	0.0023	0.0394	0.0546
Term spread	0.1383	0.1150	0.2598	0.1693	-0.1186
	(0,98)	(0,81)	(1,84) *	(1,27)	(-0,77)
Economic significance	0.0410	0.0341	0.0773	0.0491	-0.0335
Constant	-0.00151	-0.00124	0.00169	-0.00031	0.00187
	(-0,28)	(-0,27)	(0,38)	(-0,08)	(0,32)
Number of obs	576	576	576	576	576
F	1.14	1.38	1.76	1.74	0.69
Prob >F	0.3197	0.2522	0.1722	0.177	0.5037
R-squared	0.0055	0.005	0.0061	0.0051	0.0032
Root MSE	0.0439	0.0439	0.0439	0.0439	0.0440

Compared to the literature findings the sign of almost all the coefficients is in line with the expectations. There cannot be done some reliable statements on differences between short and longer horizons, because the differences are not that big. The difference between the both tested variables is also very small, but based on the results I would state that the term spread is a somewhat better predictive variable because of a little bit higher economic magnitude and the statistically significance at a 1 and 3 year horizon on the 10% level. Nevertheless again the low value of the R-squares must be taken into account and therefore the results are not really valuable to do some reliable statements on the predictive power of those variables.

5.5 RESULTS CAY VARIABLE

In this paragraph the results of the Cay variable as explanatory variable are shown. The first thing that is important to mention is that the data that is used for this research was based on quarterly data. This is the reason why there are fewer observations available compared to the other explanatory variables. There are only 192 observations available instead of the 576 observations of the other single regression models. This is not a problem, because 192 observations is still much more than 30 observations (rule of thumb) and so this must be enough to do a reliable single linear regression test.

The results show a positive coefficient at all horizons, except the outcome of the regression model based on a 5 year horizon. The expectation was a positive sign of the outcomes as described in paragraph 3.5, so the results are in line with the expectation. Most striking is the outcome of the regression at a 1 month horizon. The coefficient is rather high and the economic magnitude is 0.121, which is the highest outcome compared to all the other single regression results.

Table 10

Dependent: S&P returns	1 month	6 months	1 year	3 year	5 year
Cay variable	0.2846	0.1581	0.1887	0.0325	-0.0341
	(2,07) **	(1,06)	(1,16)	(0,18)	(-0,2)
Economic significance	0.121	0.067	0.079	0.013	-0.013
Constant	0.0096	0.0053	0.0053	0.0053	0.0054
	(2,7)	(1,94)	(1,93)	(1,89)	(1,94)
Number of obs	192	192	192	192	192
F	4.28	1.13	1.35	0.03	0.04
Prob >F	0.04	0.2895	0.2464	0.8535	0.8434
R-squared	0.0114	0.0059	0.0082	0.0002	0.0002
Root MSE	0.0495	0.0382	0.0381	0.0383	0.0383

In line with the findings of Lettau and Ludvigson (2001) described in chapter 3, the results are statistically significant (5% level) at a 1 month horizon. The statement that the Cay variable has more predictive power at short horizons looks correct. Nevertheless also these results must be interpreted carefully because the goodness of fit of the models is still very low.

5.6 RESULTS FED MODEL

The results of the last single regression model with the Fed model as independent variable are presented in this paragraph. Striking is that the results show a positive coefficient at a 1 month and at a 6 month horizon regression and at regressions with an horizon from 1 year and longer a negative coefficient. The expectation was a positive coefficient, so at horizons till 6 months this is in line with the expectation described in paragraph 3.6. Maio (2008) found also a downward slope of the coefficient the longer the horizon, so this is in line with the results, but Maio (2008) did not found a negative coefficient. Another finding of Maio (2012) was that the predictive power is especially there at short horizons. The results in table 11 show no statistically significant results and show very low R-squares at all horizons. Nevertheless the results of the shortest horizon (1 month) show the most predictive power compared to the other tested horizons.

Table 11

Dependent: S&P returns	1 month	6 months	1 year	3 year	5 year
Fed model	0.0948	0.0272	-0.0363	-0.0695	-0.0563
	(0,95)	(0,3)	(-0,4)	(-0,72)	(-0,65)
Economic significance	0.046	0.013	-0.017	-0.030	-0.025
Constant	0.0061	0.0060	0.0058	0.0056	0.0057
	(3,32)	(3,25)	(3,15)	(3,03)	(3,11)
Number of obs	576	576	576	576	576
F	0.91	0.09	0.16	0.51	0.43
Prob >F	0.341	0.7619	0.6877	0.4738	0.5134
R-squared	0.0021	0.0002	0.0003	0.0009	0.0006
Root MSE	0.0439	0.0440	0.0440	0.0440	0.0440

Also Asness (2003) found predictive power in the short run, which are already described in paragraph 3.6. The results of the simple linear regression do not show statistically significant predictive power, but again the statement can be done that there is more predictive power at the very short run than over longer horizons. This is also confirmed by the economic magnitude that has only a noteworthy value at the 1 month horizon regression results.

5.7 RESULTS MULTIPLE REGRESSION MODELS

In this paragraph the results of the several multiple regressions will be presented. The structure of this paragraph is also based on the sequence of the hypotheses. First the results of the regression for hypotheses 7 are presented in table 12

Table 12

Dependent: S&P returns	1 month	6 months	1 year	3 year	5 year
Dividend yield	0.4264	0.3922	0.253251	0.0828	0.4718
	(1,43)	(1,47)	(0,91)	(0,33)	(1,4)
Economic significance	0.12	0.11	0.07	0.02	0.12
Price-earnings ratio	0.0000486	0.0000970	0.0000274	0.0000071	0.0007427
	(0,26)	(0,61)	(0,11)	(0,04)	(1,61)
Economic significance	0.01	0.03	0.01	0.00	0.12
Short term interest rate	-0.1053	-0.0140	0.0812	0.0599	0.1702
	(-0,81)	(-0,12)	(0,7)	(0,47)	(1,32)
Economic significance	-0.07	-0.01	0.06	0.04	0.11
Default spread	0.1728	0.0019	-0.5268	0.0979	-0.3998
	(0,29)	(0)	(-1,1)	(0,19)	(-0,48)
Economic significance	0.02	0.00	-0.06	0.01	-0.04
Term spread	0.0698	0.1548	0.4151	0.2555	0.0308
	(0,34)	(0,9)	(2,35) **	(1,32)	(0,16)
Economic significance	0.02	0.05	0.12	0.07	0.01
Constant	-0.0054	-0.0096	-0.0075	-0.0048	-0.0278
	(-0,69)	(-1,37)	(-0,82)	(-0,52)	(-1,62)
Number of obs	576	576	576	576	576
F	1.02	1.46	1.79	0.79	1.15
Prob >F	0.406	0.2028	0.1139	0.5586	0.3306
R-squared	0.01	0.0107	0.0142	0.0071	0.0124
Root MSE	0.0439	0.0439	0.0438	0.0440	0.0439

By combining more independent variables into a multiple regression model, the idea was to find a model with a higher goodness of fit. The above results (table 12) still show very low R-squares and because of that the statement can be done that the combination of the set of independent variables in hypothesis 7 does not lead to a model with striking explanatory power. Almost all the results are not statistically significant, except the coefficient of the term spread at a one year horizon. By analyzing the economic significance the results of the term spread at a 1 year horizon is also rather high which imply that the economic magnitude of the term spread on the S&P returns is there. However the results of the dividend yield are not statistically significant I want to mention these results (especially at a 1 month, 6 months and 5 year horizon), because the economic significances of the dividend yield coefficients are the highest from all the independent variables that are taken into account in the model of hypothesis 7.

Second the results of the regression for hypotheses 8 will be shown in table 13. The first thing that is important to mention is that the number of observations in the tested model are lower, because of the Cay variable in the model (only available at quarterly basis).

Especially the results of the Cay variable and the Fed model at a 1 month horizon are striking. Both are statistically significant at a 99% confidence level and in both cases the economic magnitude is high. In both cases the explanatory power will decrease, the longer the time horizon that is used in the regression model. This is also in line with the literature findings that also state more predictive power on short horizons.

Table 13

Dependent: S&P returns	1 month	6 months	1 year	3 year	5 year
Fed model	0.6842	0.1157	0.0592956	0.0491	0.0176
	(2,75)***	(0,77)	(0,39)	(0,35)	(0,15)
Economic significance	0.33	0.05	0.03	0.02	0.01
Cay variable	0.6811	0.2526	0.2535	0.0482	-0.0269
	(3,67)***	(1,55)	(1,38)	(0,26)	(-0,15)
Economic significance	0.29	0.11	0.11	0.02	-0.01
Short term interest rate	-0.0272	-0.0854	-0.1048	0.0398	-0.0088
	(-0,23)	(-0,93)	(-1,15)	(0,59)	(-0,08)
Economic significance	-0.02	-0.06	-0.07	0.03	-0.01
Constant	0.0126	0.0101	0.0110	0.0032	0.0059
	(1,82)	(1,75)	(1,97)	(0,65)	(0,88)
Number of obs	192	192	192	192	192
F	4.73	1.28	1.03	0.17	0.02
Prob >F	0.0033	0.2827	0.3785	0.9157	0.9953
R-squared	0.0791	0.014	0.0158	0.0016	0.0003
Root MSE	0.04804	0.0382	0.03817	0.03844	0.03847

The R-squared of the model at a 1 month horizon is around 0.08, which can be defined as high, especially compared to the earlier tested models. It is also important to taken into account, that at this type of research, the expectation of the R-squared is not very high. When the goodness of fit would be very high, it would imply that stock returns would be very strong predictable and that is not possible. Another important thing to mention is that the longer the tested horizon, the lower the R-square of the model, so the most predictive power is found on a short horizon.

The next step in the research is running a regression with as much as possible independent variables into the model. As already mentioned in paragraph 4.4 there can be arise a multicollinearity problem and this is the fact when all the tested variables are pulled together in one model. Nevertheless the results of the regression are attached in the appendix and can be found at appendix A.

The solution that is chosen to solve the multicollinearity problem is to drop out some variables from the model. Which variable to drop out of the model is based on the Variance Inflation Factor, which is already discussed in paragraph 4.4. For the first regression the short term interest rate is dropped out of the model and the results are shown in table 14.

Table 14

Dependent: S&P returns	1 month	6 months	1 year	3 year	5 year
Dividend yield	0.0252	0.1810	0.00258	0.0319	0.5089
	(0,05)	(0,56)	(0,01)	(0,09)	(0,85)
Economic significance	0.01	0.05	0.00	0.01	0.13
Price-earnings ratio	0.00047	0.00019	0.00004	0.00013	0.00044
	(1,37)	(1,01)	(0,12)	(0,74)	(0,64)
Economic significance	0.14	0.06	0.01	0.04	0.07
Short term interest rate	-	-	-	-	-
	-	-	-	-	-
Economic significance	-	-	-	-	-
Default spread	1.3331	-0.2293	0.5671	0.5390	-0.6546
	(0,87)	(-0,31)	(0,8)	(0,79)	(-0,62)
Economic significance	0.14	-0.02	0.06	0.06	-0.06
Term spread	-0.0256	0.3444	0.3784	-0.1023	0.0259
	(-0,1)	(1,51)	(1,7)	(-0,49)	(0,11)
Economic significance	-0.01	0.10	0.11	-0.03	0.01
Fed model	0.8810	0.1959	0.1764704	0.0918	-0.0533
	(3,39)***	(1,31)	(1,2)	(0,52)	(-0,26)
Economic significance	0.43	0.09	0.08	0.04	-0.02
Cay variable	0.8623	0.2278	0.2744	0.1230	-0.1035
	(3,78)***	(1,21)	(1,42)	(0,61)	(-0,5)
Economic significance	0.37	0.10	0.11	0.05	-0.04
Constant	-0.0119	-0.0062	-0.0069	-0.0022	-0.0121
	(-0,87)	(-0,67)	(-0,5)	(-0,19)	(-0,46)
Number of obs	192	192	192	192	192
F	3.45	1.09	1.04	0.53	0.17
Prob >F	0.003	0.3678	0.3996	0.7857	0.9837
R-squared	0.1086	0.0262	0.0348	0.0068	0.0057
Root MSE	0.0477	0.0383	0.0381	0.0387	0.0387

The results of table 14 show again only statistically significant results for the Fed model and the Cay variable at a 1 month horizon. Again the economic significance of those two variables (at a 1 month horizon) are very high, respectively 0.43 and 0.37. When the other coefficients will be analyzed, the only statements that can be done are based on the sign of the coefficients. The sign of the coefficients is in line with the outcomes of the single linear regression models, discussed earlier in this chapter. Another striking result is that the value of the dividend coefficient is much lower than at the single linear regression model. This implies that compared to the Cay variable and the Fed model the predictive power of the dividend yield is lower.

By analyzing the R-squares, the result at a 1 month horizon (0.1086) is rather high and also till the 1 year horizon there can be spoken of kind of some goodness of fit from the model. Compared to the regression model in hypothesis 8, the R- squared at the 1 month horizon is only increased with around 0.03. This shows also that by adding the other independent variables to the model the goodness of fit is not striking increasing, which implies again that the Fed model and the Cay variable are the most important predictive variables in the short run.

To be as complete as possible there is also created a regression with only the dividend yield that is dropped out of the model and these results are shown in table 15.

Table 15

Dependent: S&P returns	1 month	6 months	1 year	3 year	5 year
Dividend yield	-	-	-	-	-
	-	-	-	-	-
Economic significance	-	-	-	-	-
Price-earnings ratio	0.0006	0.0002	-0.0001	0.0002	-0.0004
	(1,29)	(0,78)	(-0,21)	(0,93)	(-0,58)
Economic significance	0.16	0.06	-0.03	0.06	-0.06
Short term interest rate	0.0743	0.0692	-0.1168	0.0713	-0.0873
	(0,23)	(0,42)	(-0,57)	(0,43)	(-0,33)
Economic significance	0.05	0.05	-0.08	0.05	-0.05
Default spread	1.1547	-0.1491	0.9213	0.3701	0.0497
	(0,66)	(-0,2)	(1,18)	(0,56)	(0,04)
Economic significance	0.12	-0.02	0.10	0.04	0.00
Term spread	0.0729	0.4168	0.2211	-0.0136	-0.0613
	(0,14)	(1,42)	(0,63)	(-0,04)	(-0,16)
Economic significance	0.02	0.12	0.07	0.00	-0.02
Fed model	0.9245	0.2532	0.1181	0.1312	-0.0607
	(2,98)***	(1,38)	(0,57)	(0,73)	(-0,33)
Economic significance	0.45	0.12	0.05	0.06	-0.03
Cay variable	0.8471	0.2405	0.3043	0.1135	-0.0281
	(3,68)***	(1,31)	(1,58)	(0,58)	(-0,15)
Economic significance	0.36	0.10	0.13	0.04	-0.01
Constant	-0.0161	-0.0065	0.0006	-0.0059	0.0167
	(-0,92)	(-0,54)	(0,04)	(-0,47)	(0,76)
Number of obs	192	192	192	192	192
F	3.52	1.02	1.11	0.67	0.08
Prob >F	0.0025	0.4156	0.3567	0.6752	0.9977
R-squared	0.1091	0.0255	0.0369	0.0075	0.0013
Root MSE	0.0476	0.0383	0.0381	0.0386	0.0388

Analyzing the results of table 15 it shows kind of the same results as table 14. Again there can only be spoken of statistically significant results at the 1 month horizon. These results are in line with the expectation, because (as already stated) the value added to the model by the other variables than the fed model and the Cay variable is very low. The statement can be done that especially the Fed model and the Cay variable have explanatory power in the short run. For longer horizons and other explanatory variables the model does not show statistically significant explanatory power.

Regression models with the Cay variable into the model have fewer observations, because the data of the Cay variable is only available at quarterly basis. To check also for explanatory power based on the maximum number of observations there are also created two regression models without the Cay variable. A model with only the Cay variable dropped out of the model is not possible, because when the Cay variable is dropped out of the model there is still a multicollinearity problem (see VIF in appendix B). Therefore there are created two regressions models, one without the Cay variable and the short term interest rate and one without the Cay variable and the dividend yield. By subtracting in each regression two variables the multicollinearity problem is solved (also shown in appendix B). The results are respectively shown in table 16 and table 17.

Table 16

Dependent: S&P returns	1 month	6 months	1 year	3 year	5 year
Dividend yield	0.2619	0.3665	0.3705947	0.1832	0.6647
	(1,19)	(1,76)*	(1,62)	(0,73)	(1,86)*
Economic significance	0.07	0.10	0.10	0.05	0.17
Price-earnings ratio	0.000146	0.000119	-0.000026	-0.000037	0.000498
	(0,76)	(0,79)	(-0,11)	(-0,23)	(1,15)
Economic significance	0.04	0.03	-0.01	-0.01	0.08
Short term interest rate	-	-	-	-	-
	-	-	-	-	-
Economic significance	-	-	-	-	-
Default spread	0.1208	0.0009	-0.4679	0.1258	-0.2048
	(0,21)	(0)	(-1,03)	(0,26)	(-0,28)
Economic significance	0.01	0.00	-0.05	0.01	-0.02
Term spread	0.2116	0.1799	0.3219	0.1749	-0.1849
	(1,49)	(1,25)	(2,19)**	(1,23)	(-1,06)
Economic significance	0.06	0.05	0.10	0.05	-0.05
Fed model	0.1328	0.0408	- 0.0455151	-0.0631	-0.1580
	(1,27)	(0,43)	(-0,47)	(-0,54)	(-1,29)
Economic significance	0.06	0.02	-0.02	-0.03	-0.07
Cay variable	-	-	-	-	-
	-	-	-	-	-
Economic significance	-	-	-	-	-
Constant	-0.00914	-0.01023	-0.00511	-0.00316	-0.01954
	(-1,14)	(-1,52)	(-0,58)	(-0,39)	(-1,2)
Number of obs	576	576	576	576	576
F	1.2	1.45	1.7	0.78	1.02
Prob >F	0.3067	0.205	0.1335	0.5657	0.4052
R-squared	0.012	0.0109	0.0137	0.0071	0.0122
Root MSE	0.04388	0.0439	0.04384	0.04398	0.04387

Analyzing the results of table 16 shows that by dropping out the Cay variable of the regression model the R- squared at the 1 month horizon decreased a lot, which again implies the importance of the Cay variable for predictions on the short run. Another striking result is that in the models without Cay the coefficients of the Fed model are also not statistically significant anymore at the 1 month horizon and also the economic significance decreased a lot. In contrast to the Fed model, the dividend yield shows some statistically significant results at the 10% level. Also the term spread show a statistically significant coefficient at the 5% level based on the 1 year horizon. In all those cases of statistically significant results also the economic significance is rather high. Nevertheless, these results must be interpreted carefully because the goodness of fit of the model is not very well.

Table 17 shows some striking results for the term spread which are statistically significant at the 10% level based on a horizon till 6 months and at a horizon of 1 year at the 5% level. Also in this case the economic significance is rather high, but because of the low goodness of fit from the model, the results must be interpreted carefully.

Table 17

Dependent: S&P returns	1 month	6 months	1 year	3 year	5 year
Dividend yield	-	-	-	-	-
	-	-	-	-	-
Economic significance	-	-	-	-	-
Price-earnings ratio	0.000235	0.000220	0.000089	-0.000004	0.000381
	(1.04)	(1.08)	(0.3)	(-0.02)	(0.87)
Economic significance	0.07	0.06	0.03	0.00	0.06
Short term interest rate	0.1508	0.1942	0.2045	0.0819	0.2630
	(1.14)	(1.58)	(1.52)	(0.61)	(1.48)
Economic significance	0.11	0.13	0.14	0.05	0.16
Default spread	0.0888	-0.0059	-0.5051	0.1484	-0.2757
	(0.14)	(-0.01)	(-0.99)	(0.28)	(-0.32)
Economic significance	0.01	0.00	-0.05	0.01	-0.03
Term spread	0.3882	0.4037	0.5588	0.2759	0.1623
	(1.81)	(1.98)*	(2.51)**	(1.24)	(0.62)
Economic significance	0.12	0.12	0.17	0.08	0.05
Fed model	0.2406	0.1804	0.1027958	0.0092	0.0647
	(1.67)*	(1.39)	(0.76)	(0.07)	(0.41)
Economic significance	0.12	0.09	0.05	0.00	0.03
Cay variable	-	-	-	-	-
	-	-	-	-	-
Economic significance	-	-	-	-	-
Constant	-0.01293	-0.01427	-0.00966	-0.00399	-0.01468
	(-1.35)	(-1.59)	(-0.84)	(-0.4)	(-0.95)
Number of obs	576	576	576	576	576
F	1.33	1.38	1.76	0.79	1.01
Prob >F	0.2501	0.2315	0.12	0.558	0.4091
R-squared	0.0123	0.0106	0.0138	0.0069	0.0094
Root MSE	0.04387	0.04391	0.04384	0.04399	0.04393

By dropping out the Cay variable out of the models, the explanatory power of the Fed model also decreased. In the single linear regression models of the Fed model and the Cay variable were no striking high R-squares shown, but a combination of the Cay variable and the Fed model looks much more successful in the short run. The added value of the other independent variables to increase the goodness of fit from the model is rather low. To check this, there is also created a multiple regression model with only the Cay variable and the Fed model as independent variables. Those are compared with the results of the single regression models and the results confirm that on the short run (1 month) a combination of the Cay variable and the Fed model is quite successful. The results show an R-squared of almost 0.08 and are statistically significant at a 99% confidence level. For longer horizons than 1 month the goodness of fit from the model is decreasing rapidly, the results are shown in appendix C.

6. CONCLUSION

In this final chapter, I summarize the main findings of the predictive regressions discussed in chapter 5, describe the main limitations of the analysis and provide some suggestions for further research.

6.1CONCLUSION

Existing research shows that the dividend yield predominantly predicts stock returns in the long run, and has little predictive power in the short run. The main reason being that is that the dividend yield is a slowly moving variable. My analysis shows that there cannot be made a reliable distinction between predictability in the short and longer run, but the results show a positive effect at all tested horizons. The effect is high, especially compared to other explanatory variables, but because of insignificant results the results must be interpreted carefully

Contrary to the dividend yield, the price-earnings rate is said to have more short-term than long-term predictability, as it is more strongly linked to current business conditions. I find that the effect of the price-earnings rate on the expected stock index returns is very small, almost negligible. Combined with other variables the effect increases, but based on the own research there cannot be done some reliable statements from predictive power of the price-earnings ratio. The sign of the price-earnings-ratio is almost in all cases negative, which is in line with the literature. A reason for the negative effect is that the level of earnings is a good measure of current business conditions. Risk premia on stocks covary negatively with current economic activity: investors require high expected returns in recessions, and lower expected returns in booms.

Based on the literature the high correlation between the dividend yield and the price-earnings ratio is striking, the price-earnings ratio has more predictive power on the short run and the dividend yield at longer horizons. In line with the literature the results of the empirical research show more predictive power from the dividend yield than the price-earnings ratio.

Expected could be that the short term interest rate would show kind of the same findings as the Price-earnings ratio, because this can be seen also like a good measure for the current business conditions. The same patterns and the corresponding magnitude as the price-earnings ratio are confirmed by the literature findings. Nevertheless this cannot be concluded from the own empirical research. Striking is that the results of the own research show a positive coefficient at horizons longer than 1 month, which is not in line with the literature and is difficult to explain. The only interpretation for the positive sign that I can give is that a rising interest rate, still expects to raise in many cases (in this data sample) and therefore investors wants to be compensated with higher returns. Overall the findings of the short term interest rate as explanatory variable shows kind of predictive power, but not in the same order of magnitude as the dividend yield.

There is a close relation of the term and default spread to the short term interest rate and this leads to the expectation of more predictive power in the short run, which is in line with the literature findings. Tested these variables separately in the empirical research shows that the economic magnitude of the variables are quite similar to each other. Nevertheless the results of the term spread at the 1 year and 3 year horizon show statistically significant results at the 10% level and therefore the term spread is a more reliable predictor of stock index returns.
The sign of the default spread is positive and this is in line with the literature. This confirms that investors want to be compensated for taking more risk (because of a higher default spread) and expect a higher return on their investment. Also the effect of the term spread is positive and can be related to the literature findings, good prospects for future activity and investments, after business troughs may contribute to high expected returns around troughs.
Combining the Term and default spread in one model does not lead to tremendous predictive power and the goodness of fit from the model is not strikingly increasing. The statement can be done that the combination of those two variables is not a successful predictor for stock index returns.

Lettau and Ludvigson (2001) have shown that their Cay variable that is discussed in chapter 3, paragraph 3.5 has much better short-term predictability than the dividend yield and the term spread. My own empirical research confirms the findings of Lettau and Ludvigson that the Cay variable is the best predictive variable in the short run. Especially at a 1 month horizon, the longer the horizon the predictive power is diminishing. The effect of the Cay variable is positive and this is in line with the explanation of Lettau and Ludvigson; booms are times of rising consumption but declining ratios of consumption to wealth, consistent with the positive relation findings.

The last explanatory variable is calculated as the earnings-price rate minus the 10 year Treasury bill rate, the so called Fed model has as component the earnings-price rate in the model and because of that the expectation of a close relationship is obvious. The literature findings shows that in the model E/P matters and the Y part of the fed model can be ignored. The specific test is not being done in this research, but by comparing the results of the price-earnings ratio and the Fed model the results are quite similar. The close relationship with the price-earnings ratio explains the negative effect at horizons longer than six months. This is also an indication that Y did not have additional effect to the model

In the last part of the research, the variables are tested in multiple regression models and the results show especially predictive power in the short run. Statistically significant results at horizons longer than 1 year are more difficult to find and are only shown by the dividend yield.

The most important variables that has to be mentioned for short run predictability are the Cay variable and the Fed model. A combination of those two variables into a model leads to the most striking results at a 1 month horizon. Compared to the single regression models the results show that they stimulate each other and this leads to the effect that the predictive power is increasing. In the short run, a single regression with the Cay variable is the best predictive variable, in a multiple regression model the Cay variable and the Fed model together have the most predictive power.

The most important variable for longer horizon predictability is the dividend yield, but the results are only significant at the 10% level. The findings are in line with the literature that the dividend yield is in the long run the best predictive variable. There is a small difference between the single and multiple regression model outcomes and this indicates that the dividend on its own leads to predictive power at longer horizons.

Based on the previous findings and my own results, I conclude that there are definitely possibilities of predictive power on index level. In the short run the Cay variable and the Fed model are the most important and at longer horizons the dividend yield has the most predictive power.

6.2LIMITATIONS AND FURTHER RESEARCH

The most severe limitations of this research will be mentioned in this paragraph. The first one is the fact that there is made use of quarterly data of the Cay variable. With as result that there are fewer observations available to do reliable statements of predictability from the Cay variable. When the Cay variable is in a multiple regression model it has also effect on the number of observations and therefore it affects also the results of the other (monthly available) independent variables.

Another limitation is that there is only made use of the basic regression models to test for predictability on index level returns. This choice is made from the point of view that this research is an overview of earlier tested variables and to keep it workable the focus lies on the literature research and the basic regression methods.
Next to the above mentioned methodological limitations it is obvious that not all possible explanatory variables are taken into account, but only the most well know predictive variables.

This research gives an overview of the most well know predictive variables and these are tested at several horizons, but for further research it would be an opportunity to add more variables to this research.

Another opportunity would be to make use of monthly Cay data to make it possible to do more reliable statements on the predictability of the Cay variable separately. This can be also useful to create more reliable predictions on the multiple models with the Cay variable in the model.

The used regression methods are the basic regression models, in further research there would be a good chance to add some more extensive methods to check for predictive power of all the tested variables

By expanding this research with more explanatory variables, with more extensive methods in the empirical part of this research this would lead to a complete overview. Nevertheless this research already gives a very extensive overview of the most well know predictive variables to do predictions on index level returns. The differences between predictability in the short and long run are already (on a basic level) taken into account and also the sign (effect) of each explanatory variable is taken into account.

7. REFERENCES

- Asness C. (2003) Fight the Fed Model
- Asness C. (2003) the relationship between stock market yields, bond market yields and future returns.
- Bekaert G, Ang A., 2006. Stock return predictability: is it there?
- Bekaert G. and Engstrom E. (2010) Inflation and the stock market: understanding the ' FED Model'.
- Bodie, Kane, Marcus Investments and Portfolio Management 9th edition
- Campbel R. Harvey and Stephan Gray (1997) Global Financial Management
- Campbell and Shiller(2001), Valuation Ratios and the Long-Run Stock Market Outlook: An Update
- Campbell J. Y., Thomson S. B. (2007) Predicting Excess stock returns out of sample: can anything beat the historical average?
- Campbell J., Yogo M. (2006) Efficient tests of stock return predictability
- Campbell J.Y. (1986) Stock returns and the term structure
- Campbell, J.Y., Shiller, R.J., 1988a. Stock prices, earnings, and expected dividends. Journal of Finance 43,661–676
- Campbell, J.Y., Shiller, R.J., 1988b. The dividend-price ratio and expectations of future dividends and discount factors. Review of Financial Studies 1, 195-228
- Chen S. (2008) predicting the bear stock market: macroeconomic variables as leading indicators
- Cochrane John H. (2007) the dog that did not bark: A defense of Return Predictability. Published by Oxford University
- Cochrane John H. 2000 Asset pricing
- Damodaran A. (2013) investment valuation third edition
- Eugene F. Fama and Kenneth R. French (2004) The Capital Asset Pricing Model: Theory and Evidence
- Falahati Kazem (2009) The Capital Asset Pricing Model
- Fama, E. F. and French, K. R. (1989). "Business Conditions and Expected Returns on Stocks and Bonds", *Journal of Financial Economics,* Vol. *25,* No. 1, pp. 23-48.
- Fama, E.F., French, K.F., 1988. Dividend yields and expected stock returns. Journal of Financial Economics 22, 3–25.
- Golder M. (2006) Multiple regression multicollinearity, University of Pennsylvania

- Hasseltoft Henrik (2009) the “FED-Model and the changing correlation of stock and bond returns: An Equilibrium approach
- Julliard, C., 2007. Labor income risk and asset returns.
- Keim D.B., Stambaugh R.F. (1986) Predicting returns in the stock and bond markets
- Lamont O. (1998) Earnings and Expected returns
- Lettau and Ludvigson (2001B) Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia Are Time-Varying
- Lettau and Ludvigson (2005), expected returns and expected dividend growth
- Lettau M. and Van Nieuwerburgh S. (2007) Reconciling the return predictability evidence
- Lettau M., Ludvigson, S., 2001a. Consumption, aggregate wealth and expected stock returns. Journal of Finance56, 815
- Lucas, Robert E. Jr., 1987, *Models of Business Cycles*, London and New York: Blackwell)
- Maio Paulo F. (2008) The FED model and Expected Asset Returns. Hanken School of Economics
- Maio Paulo F. (2012), The Fed Model and the Predictability of stock returns. Hanken School of Economics
- Nai-Fu Chen, Richard roll and Stephen A. Ross 1986 Economics Forces and the Stock-Market Page 383-403
- Perold Andre F. (2004) The Capital Asset Pricing Model
- Rangvid J, Rapach D, Wohar M., 2005. Macro variables and international stock return predictability
- Rangvid J. 2006 Output and Expected returns
- Rangvid J., Lioui A. (2007) Stock return predictability in a monetary economy
- Rangvid J., Rapach D.E., Wohar M.E. (2005) Macro variables and international stock return predictability
- Ross S.A., Roll R., Chen N. (1986) Economic forces and the stock market
- Santos, T., Veronesi, P., 2006. Labor income and predictable stock returns. Review of Financial Studies 19, 1–44.
- Schotman P.C., Tschernig R., Budek J. (2008) Long memory and the term structure of risk.
- Stambaugh R.F. (1999) Predictive regressions

- Torous W., Valkanov R., Yan S. (2004) On predicting stock returns with nearly integrated explanatory variables
- Verbeek M. (2008) Modern Econometrics, 3rd edition, Wiley.

APPENDIX

APPENDIX A

Dependent: S&P returns	1 month	6 months	1 year	3 year	5 year
Dividend yield	-0.2777	0.2315	0.5117649	-0.1865	0.8434
	(-0,4)	(0,39)	(0,93)	(-0,31)	(1,18)
Economic significance	-0.08	0.06	0.14	-0.05	0.22
Price earnings ratio	0.00060	0.00017	-0.00019	0.00023	0.00004
	(1,3)	(0,54)	(-0,4)	(0,95)	(0,05)
Economic significance	0.18	0.05	-0.06	0.07	0.01
Short term interest rate	0.1977	-0.0333	-0.3403	0.1453	-0.3154
	(0,4)	(-0,11)	(-1,12)	(0,53)	(-1,01)
Economic significance	0.14	-0.02	-0.23	0.09	-0.20
Default spread	1.2300	-0.2104	0.7855	0.4312	-0.1795
	(0,74)	(-0,28)	(1,01)	(0,62)	(-0,14)
Economic significance	0.13	-0.02	0.08	0.05	-0.02
Term spread	0.2075	0.3052	-0.0218	0.0737	-0.3624
	(0,3)	(0,74)	(-0,05)	(0,19)	(-0,78)
Economic significance	0.06	0.09	-0.01	0.02	-0.10
Fed model	1.0211	0.1724	-0.0626	0.2038	-0.2900
	(2,55)**	(0,63)	(-0,24)	(0,71)	(-1,05)
Economic significance	0.49	0.08	-0.03	0.09	-0.13
Cay variable	0.8623	0.2278	0.2758	0.1251	-0.0916
	(3,83)***	(1,21)	(1,42)	(0,62)	(-0,45)
Economic significance	0.37	0.10	0.11	0.05	-0.03
Constant	-0.0178	-0.0052	0.0032	-0.0063	0.0019
	(-0,96)	(-0,42)	(0,19)	(-0,51)	(0,07)
Number of obs	192	192	192	192	192
F	3.05	0.94	1.01	0.58	0.3
Prob >F	0.0046	0.4765	0.4262	0.7734	0.9531
R-squared	0.1098	0.0263	0.0411	0.0081	0.011
Root MSE	0.0477	0.0384	0.0381	0.0387	0.0387

	Cay dropped		Without Cay and STIR		Without Cay and dividend yield	
Variable	**VIF**	**1/VIF**	**VIF**	**1/VIF**	**VIF**	**1/VIF**
Short term IR	12.34	0.081028	-	-	3.81	0.262127
Dividend yield	6.31	0.158539	1.95	0.512878	-	-
Fed model	4.16	0.24027	1.16	0.861804	2.04	0.490028
Term spread	4.33	0.23088	1.18	0.849603	2.39	0.418095
Default spread	1.76	0.567351	1.69	0.592309	1.72	0.581435
P/E ratio	2.66	0.375954	1.65	0.607326	2.46	0.407068
Mean VIF	5.26		1.52		2.48	

Dependent: S&P returns	1 month	6 months	1 year	3 year	5 year
Fed model	0.6873	0.1244	0.0664053	0.0491	0.0191
	(2.81)***	(0.84)	(0.44)	(0.35)	(0.16)
Economic significance	0.33	0.06	0.03	0.02	0.01
Cay variable	0.6726	0.2267	0.2235	0.0529	-0.0257
	(3.52)***	(1.36)	(1.21)	(0.29)	(-0.14)
Economic significance	0.29	0.10	0.09	0.02	-0.01
Constant	0.0112	0.0056	0.0055	0.0054	0.0054
	(3.19)	(2.08)	(2.01)	(1.94)	(1.95)
Number of obs	192	192	192	192	192
F	6.33	0.96	0.74	0.08	0.03
Prob >F	0.0022	0.383	0.4764	0.925	0.9666
R-squared	0.0788	0.0096	0.0092	0.0008	0.0003
Root MSE	0.04792	0.03819	0.0382	0.03836	0.03837