

BI Norwegian Business School - campus Oslo

GRA 19502

Master Thesis

Component of continuous assessment: Thesis Master of Science

CAPE vs the Fed Model: Comparative analysis of the out-of-sample performance in predicting future stock returns

ID number: 0939950, 0986821

Start: 02.03.2017 09.00

Finish: 01.09.2017 12.00

GEORGII MESHCHERIAKOV: **0939950**

MARINA POLUIANOVA: **0986821**

Hand-in date:

31.08.2017

Programme:

Master of Science in Business with Major in Finance

Master of Science in Finance

“This thesis is a part of the MSc programme at BI Norwegian Business School. The school takes no responsibility for the methods used, results found and conclusions drawn.”

Abstract

The goal of this thesis is to compare the out-of-sample performance in predicting future stock returns of the Fed model and Shiller's cyclically adjusted price-to-earnings ratio (CAPE). The two models are also augmented with 10-year Treasury bond yield. Additionally, a version of CAPE that uses after-tax corporate profits and a Fed model adjusted for perceived risk are analyzed. The four models are tested using US market time series in a period ranging from 1871 to 2016 using regression analysis and compared using visual inspection of plots, forecast statistics, and forecast equivalence tests. We find that in-sample and out-of-sample tests show low R^2 , which may still have economic significance for risk-averse investors. Versions of CAPE model dominate the Fed model alternatives in-sample. Out-of-sample, traditional CAPE exhibits very limited ability to generate accurate forecasts for long investment horizons of 20 years, while Fed model performs better at 1-year horizon. Alternative versions of the two models do not exhibit significant improvement in equity return predictability. So, for strategic asset allocation, CAPE and Fed should be used with caution as complementary tools and in conjunction with diversification.

Acknowledgements

We would like to express gratitude to our thesis supervisor, Associate Professor Costas Xiouros of the Department of Finance of BI Norwegian Business School. His guidance, attention, supervision and patience were essential to writing and completing our Master's thesis. We would also like to thank our university for providing ready access to facilities, equipment and data used in the process.

Georgii Meshcheriakov

Marina Poluianova

Contents

List of Tables	v
List of Figures	vi
1 Introduction	1
2 Literature Review	2
2.1 CAPM and Book-to-market ratio	2
2.2 Dividend yield	3
2.3 Earnings yield	4
2.4 Price-to-earnings ratio	6
2.5 Shiller's CAPE	6
2.6 Fed model	9
3 Theoretical Background	10
3.1 Price as common factor	10
3.2 CAPE Framework	11
3.3 Fed Model Framework	13
4 Methodology	16
4.1 Hypotheses	17
4.2 Regression analysis	17
4.3 Unit-root test	18
4.4 Forecasting Frameworks	19
4.4.1 Generic CAPE	19
4.4.2 After-tax Profit CAPE	19
4.4.3 Generic Fed model	19
4.4.4 Fed model adjusted for perceived risk	20
4.5 In-sample testing	21
4.6 Heteroskedasticity testing	22
4.7 Pseudo-out-of-sample testing	23
4.7.1 Check for outliers in forecast error time series	23
4.7.2 Comparison of forecast statistics	23

4.7.3	Testing equality of forecast accuracy	24
5	Data	26
6	Empirical findings and Analysis	27
6.1	Descriptive statistics	27
6.2	In-sample regression statistics	28
6.3	Pseudo-out-of-sample regression statistics	29
6.4	Check for outliers in forecast error time series	31
6.5	Forecast evaluation statistics	32
6.5.1	Generic CAPE	32
6.5.2	After-tax profits CAPE	34
6.5.3	Generic Fed model	36
6.5.4	Fed Model adjusted for perceived risk	38
6.6	Testing equality of forecast accuracy	41
6.6.1	MGN test of forecast Accuracy	41
6.6.2	DM test of forecast Accuracy	42
7	Conclusion	43
	Bibliography	44
	Appendices	49
Appendix 1 -	OLS Framework	49
Appendix 2 -	Out-of-sample forecast evaluation statistics	50
Appendix 3 -	Visual analysis of forecast time series	51
Appendix 4 -	Plots of Forecast error outlier time series	54
Appendix 5 -	Preliminary thesis report	57

List of Tables

Table 1: US Data in- and out-of-sample allocations	22
Table 2: Time series descriptive statistics	27
Table 3: In-sample regression statistics (Level and FD)	28
Table 4: Out-of-sample regression statistics (Level)	29
Table 5: Out-of-sample regression statistics (FD)	30
Table 6: CAPE forecast evaluation statistics (Level)	33
Table 7: CAPE forecast evaluation statistics (FD)	33
Table 8: CAPE with BY forecast evaluation statistics (Level)	34
Table 9: CAPE with BY forecast evaluation statistics (FD)	34
Table 10: NIPA CAPE forecast evaluation statistics (Level)	35
Table 11: NIPA CAPE forecast evaluation statistics (FD)	35
Table 12: NIPA CAPE with BY forecast evaluation statistics (Level)	36
Table 13: NIPA CAPE with BY forecast evaluation statistics (FD)	36
Table 14: Fed model forecast evaluation statistics (Level)	37
Table 15: Fed model forecast evaluation statistics (FD)	38
Table 16: Adjusted Fed model forecast evaluation statistics (Level)	39
Table 17: Adjusted Fed model forecast evaluation statistics (FD)	40
Table 18: MGN test results (Level)	41
Table 19: DM test results (Level)	42

List of Figures

Figure A.1: CAPE Predictions vs S&P 500 Return	50
Figure A.2: CAPE vs Alternatives	50
Figure A.3: Fed Predictions vs S&P 500 Return	51
Figure A.4: Fed vs Alternatives	51
Figure A.5: CAPE and Fed Predictions vs S&P 500 Return	52
Figure A.6: CAPE vs Fed	52
Figure A.7: Time series of CAPE forecast errors	53
Figure A.8: Time series of NIPA CAPE forecast errors	53
Figure A.9: Time series of CAPE with BY forecast errors	53
Figure A.10: Time series of NIPA CAPE with BY forecast errors	54
Figure A.11: Time series of Fed model forecast errors (EY_t)	54
Figure A.12: Time series of Fed model forecast errors ($EY_t - BY_t$)	54
Figure A.13: Time series of Fed model forecast errors (EY_t & BY_t)	55
Figure A.14: Time series of Adjusted Fed model forecast errors (EY_t)	55
Figure A.15: Time series of Adjusted Fed model forecast errors (EY_t^f)	55
Figure A.16: Time series of Adjusted Fed model forecast errors (EY_t^e)	56

1. Introduction

Expected returns on securities fluctuate over time and the ability to accurately determine this variation and accurately forecast security returns is fundamental for purposes of tactical and strategic asset allocation. Today, the investment community has access to a multitude of forecasting frameworks, most of which impose unrealistic assumptions and rarely provide a reliable forecast of future security returns. Usually, the frameworks are grouped in two broad categories: technical and fundamental analysis. Fundamental analysis is concerned with identifying security mispricing, which would enable the implementation of superior investment strategies. The objective is to measure security's intrinsic value by performing analysis of macro- and microeconomic data and factors that have the potential to affect security price. This involves extensive use of performance ratios made up of key figures from financial reports and other metrics.

In this thesis we compare two notable examples of valuation metrics within the domain of fundamental analysis, the Fed model and Shiller's cyclically adjusted price-to-earnings ratio (CAPE). The two are of significant interest as, despite strong criticism and disputed effectiveness, both remain popular in the finance industry. Our goal is to determine which model shows better in-sample and out-of-sample performance in predicting future returns based on the available data the US. We also seek to determine whether good model's forecasting performance can be attributed to solid theoretical foundation or misspecification.

Our work begins from the expanded version of Vanguard research on stock return forecastability (Davis, Aliaga-Díaz, & Thomas, 2012). The said analysis studies whether 16 popular metrics, including P/E 10 and the Fed model, have indeed exhibited correlation with the future stock returns in the period from January 1926 to June 2011. The findings suggest that Shiller's CAPE is the best performing valuation measure, explaining 43% of time variation in future real stock returns compared to 16% exhibited by the Fed model. Another important observation states that the smoothed P/E only

provides reasonable long-term forecasts. We proceed by studying the theoretical specifics, underlying assumptions, practical strengths and weaknesses of the two metrics in more detail based on, among others, Campbell and Shiller (2001), Fama and French (1989), Siegel (2016), Yardeni (1997).

With this thesis we seek to contribute to the body of work that compares valuation metrics in their ability to forecast long-term stock returns and time the market. The scope of the comparison encompasses the original CAPE and Fed models as well their improved versions proposed by Salomons (2004) and Siegel (2016).

In the next section, an overview of literature studying the most common metrics used for stock returns forecasting, including the two mentioned above. Section 3 contains theoretical background of the two models of choice. In section 4, our hypotheses are formulated, regression frameworks and respective variables used for testing are described. Data sets used and their sources follow in Section 5. In section 6, we present our findings and analysis, which are followed by our conclusions.

2. Literature Review

Academia and practitioners constantly strive to identify methods to efficiently and accurately forecast stock returns. Several factors showing moderate forecasting ability have been identified. In this section, we will provide an overview of the most notable factors, examining their relations and commonalities that lead up to the emergence of the two models of our interest: CAPE and Fed.

2.1. CAPM and Book-to-market ratio

In 1970, William Sharpe brought the capital asset pricing model (CAPM) to light, arguing that stock's beta, the measure of systematic risk, was the only relevant factor for explaining return fluctuations of a single stock, or a portfolio of stocks. He inferred that a high-beta stock will show a stronger

response to market movements (Sharpe, 1970). However, the model works under unrealistic assumptions and only takes into account two dates, disregarding any time variation in expected stock returns beyond. This renders CAPM inadequate for stock return forecasting.

Early studies by Fama and French (1992) showed evidence that book-to-market equity ratio (B/M), company's equity book value to its market value, explained stock return fluctuations leading to the idea that B/M could be a potentially solid metric for stock returns forecasting. It was further popularized by Fama and French 3 Factor model (FF), which improved upon the CAPM by accounting for company size and value as factors determining stock return. Authors argued that undervalued value stocks, characterized among other factors by $B/M > 1$, offer superior returns versus growth stocks as a reward for the financial distress and future earnings uncertainty (Fama and French, 1993).

B/M ratio of Dow Jones industrial Index (DJIA) and S&P 500 exhibited ability to predict market returns and small-cap excess returns between 1926 and 1996. Authors claimed that this could be attributed to the relation between book value and future earnings (Pontiff and Schall, 1998). However, based on the evidence from Australian and US stock market, a more recent analysis found that B/M ratio owes its stock return qualities to the absorption of explanatory power of market leverage as a risk factor (Dempsey, 2010).

2.2. Dividend yield

A significant body of academic research argues that dividend-price (D/P) ratio, or dividend yield, can be used to predict stock returns. Campbell and Shiller (1988) propose variance decomposition, arguing that log dividend yield variation is attributable to and positively correlated with the variance in expected future dividend growth and expected stock returns. The authors regressed returns on value- and equal-weighted portfolios of New York Stock Exchange (NYSE) on D/P concluding that the explanatory power (R^2) of the financial ratio increases significantly with investment horizon. Namely, R^2 was up to 5% for 1 - 3 month holding periods and exceeded 25% for 2 - 4 year holding periods. Authors argued that these mean-reverting components

of stock prices are a result of positive auto-correlation in long-run expected returns being offset by negative price shocks. Hodrick (1992) uses Monte Carlo simulations and VAR tests to avoid small sample bias and conclude that dividend yields forecast a significant portion of expected stock return variation. Cochrane (2007) reports strong evidence in favor of stock return forecasting ability of dividend yields based on US data. Interestingly, his analysis suggests that 1% increase in D/P results in a positive price shock of more than 3% instead of a 1% negative shock in response to extra dividends.

2.3. Earnings yield

Another common financial ratio deemed to have stock returns forecasting ability is the earnings yield, ratio of operating earnings to market value (E/P). Basu (1983) tests for empirical relationship between E/P , firm size and returns on the NYSE common stock with results indicating that firms with high E/P tend to exhibit higher risk-adjusted returns than low E/P firms. Author further shows that firm size effect on returns is negligible after accounting for the E/P .

Lewellen (2004) evaluates forecasting ability of D/P , B/M and E/P . He shows that regressing NYSE returns on natural logarithm of dividend yield, $\ln(D/P)$, offers evidence that dividend yield predicts aggregate market returns in the period 1946 – 2000 and other sub-periods. He also shows that $\ln(E/P)$ and $\ln(B/M)$ have limited ability to forecast nominal returns based on the 1963 – 2000 sample. The addition of 1995 – 2000 data strengthens the case for predictability by all three metrics.

Lamont (1998) uses the fact that E/P connects dividend yield to dividend payout ratio (D/E) and finds that D/P and D/E are positively correlated with future excess stock returns, with D/E being less robust. Lamont states that, during economic recovery or boom, reinvestment of profits is preferred to dividend payouts, making D/E a potent measure of business conditions, which in turn affect expected return variation. This observation can partly explain the potential of D/E to forecast expected equity returns. Overall, returns are expected to be low when prices and profits are high, or when dividend yield and payout ratio are low, respectively. However, management may not

be equally efficient at allocating savings from lower dividends to increase the company value and future earnings per share (EPS), as a result (Arnott and Asness, 2003).

Welch and Goyal (2007) show that multiple variables used in one-month ahead stock return predictions consistently fail to match and even underperform a simple historical average market return both in-sample and out-of-sample in the 1975 - 2004 period . Authors explain this problem by the lack of robustness in predictive models.

Campbell and Thompson (2007) used monthly data to predict simple monthly or annual S&P 500 Index returns and found that imposing restrictions on signs of coefficients and return forecasts never impedes and sometimes improves out-of-sample performance of a wide selection of forecasting variables in predictive regressions. This lead to the regressions yielding, on average, return figures superior to those of historical mean, with small but economically meaningful out-of-sample power.

An important take-away is that small R^2 values should not be neglected by investors due to potential for significant portfolio performance improvements due to improved market timing strategies. The study shows that an out-of-sample R^2 of 2% may lead to 8% return improvement annually, assuming optimal allocation in stocks and risk-free asset. Furthermore, in the short-run, higher R^2 should be carefully considered as they may indicate spurious regressions but is more realistic at longer horizons.

Da et al. (2014) propose a stock yield as factor with solid out-of-sample forecasting performance. Derived from Gordon constant growth model, it is a sum of the dividend yield (D/P) and a weighted average of expected future dividend growth rates (G). Using the 1977 – 2012 data, monthly predictive regressions for average monthly stock returns in-sample yielded future market returns with an adjusted R^2 of 13% and 54% at 1-year and 4-year horizons, respectively. Out-of-sample, R^2 for various horizons is consistently above 2%. This is superior to comparable stock return predictors, especially at times when recession is expected by the investors.

2.4. Price-to-earnings ratio

An inverse of E/P ratio is the price-to-earnings (P/E) ratio. Basu (1977) tests the relationship between P/E ratio and return performance of stocks listed on NYSE. Author finds that, on average, low P/E stocks outperformed high P/E securities in terms of risk-adjusted returns in the 1957 - 1971 period and explains that the contradiction of semi-strong form of efficient market hypothesis (EMH) can be explained by lagged response of the market to information reflected in P/E .

The long-run variation of stock returns can be observed as a response to altering market and economic conditions. At times of economic distress, a low P/E tends to be observed, owing to the high levels of risk-aversion exhibited by investors, prompting high expected rates of return and sinking prices. The opposite is observed when stock index prices surge as economy recovers (Fama and French, 1989). An alternative explanation stems from irrational investor behavior. Potential over-reliance on use of trailing returns for forecasting as well as overconfidence following a profitable investment may lead to a feedback loop, which, coupled with herd behavior, may lead to bubbles characterized by high P/E ratio (De Bondt and Thaler, 1990). In both scenarios, prices eventually revert to fair value.

2.5. Shiller's CAPE

Cyclically Adjusted Price Earnings ratio (CAPE) is a measure popularized by the collaborative effort of Campbell and Shiller (1998). The current value of Shiller's CAPE, P/E_{10} , is defined as the ratio of the real spot price and the 10-year arithmetic average of the lagged real earnings of a the Standard and Poor Composite Index (S&P 500) or any alternative broad equity index. CAPE's key distinction from the generic P/E measure, which uses 1-year trailing EPS, is the incorporation of a 10-year trailing EPS in the denominator to smooth the occasional extreme up- and downward spikes in corporate earnings resulting from cyclical effects. Additionally, the choice of 10-year average to represent the long horizon is justified by the predominant interest in long-term investments and that long-horizon returns are more

forecastable. This metric has been influenced by Graham and Dodd (1934), who pioneered the use of multi-year average EPS, stressing that at least 5-year average should be used.

In 2001, Shiller and Campbell published an update to their 1988 paper on use of valuation ratios. Alterations included expansion of time series for annual US stock market data (1871 - 2000) and inclusion of 30 years of quarterly data (starting in 1970) for 12 other countries. Their findings from US verify their earlier findings that P/E_{10} ratio is fit for use in forecasting changes in future stock prices. Study of foreign markets shows mixed evidence of CAPE's efficiency with some results hard to interpret owing to the lack of data. Most markets had roughly 30 years of data, compared to 129 in case of US. Due to this complication, authors resorted to using a 4-year smoothing as opposed to the typical 10-year alternative. The second major point discussed potential factors responsible for the high level of P/E_{10} ratio, roughly 41, and its non-reversion to the historical mean.

Faber (2012) extends the research of CAPE's efficiency in predicting future real stock returns. Authors analyze the US market along with 32 other countries. Despite the bias due to shorter time-series collected from international data, results are comparable to domestic, indicating that high-CAPE countries show lower future returns, while low CAPE indicate higher future returns outperforming the market by 4 - 7%. Keimling (2014) finds more evidence on CAPE performance internationally, in 14 stock markets, concluding that CAPE is not well suited for forecasting short- and medium-run, but can provide realistic long-run equity return expectations.

Siegel (2016) suggests that the current ratio levels are too high giving unreasonably pessimistic forecasts of future stock returns. The valuation bias can be attributed to the changes enforced by the Financial Accounting Standards Board (FASB) in 2001. It stated that the value of financial securities available for trading or sale as well as impairments to the value of fixed assets and intangibles was to be adjusted to reflect the fair market value. The effect of write-downs performed by large companies during the financial crisis of 2008 was even greater due to the so called "aggregation

bias” observed in S&P 500. It entails computation of index earnings as a net of profits and losses of each listed company irrespective of their market capitalization, leading to a few companies with substantial losses negatively affecting the P/E of the entire index. Siegel suggests using corporate National Income and Product Accounts (NIPA) after-tax profits which offer more consistent data at expense of data availability, going back to 1928. His study indicates that this method effectively increases the explanatory power of CAPE in forecasting 10-year real stock returns to 40.09%, compared to reported (35%) and operating (36%) S&P 500 earnings.

In addition, CAPE is assumed to have the capacity to be used as a market timing tool. This is based off the fact that Shiller accurately predicted the market crash resulting from the sub-prime mortgage crisis of 2007, after observing the metric in excess of 25. Similar levels were found in data in periods clustered around 1929 and 1999, notable for the Great Depression and dot-com bubble, respectively. In 1998, Campbell and Shiller suggested that the valuation may deviate from the historical ranges as a result of, among other things, changes in structure of national (US) industry and investor behavior. Today, Shiller’s P/E stands at even higher 28.26, which is roughly 69% higher than the historical mean of 16.7. The generally proposed strategy is investing in stocks when CAPE is below its historical average and selling when it is higher (Faber, 2012). However, in an interview with Business Insider, Robert Shiller states that CAPE is not a market timing mechanism (Blodget and Kava, 2013). Instead, he points out that the metric should be used together with an adequate algorithm for diversification so as to avoid selling off stocks completely. Hence, a lower (higher) CAPE should serve as a signal to increase (decrease) equity holdings in a portfolio.

Kantor and Holdsworth (2014) argue that no model based on P/E ratio produces sufficiently accurate return forecasts consistently over time. Test for mean-reversion using a unit-root and Bai-Perron test reveals CAPE’s failure to return to historic mean. Authors warn against using CAPE as a definitive market-timing tool, but point at metric’s feasibility as a signal to the investor to be cautious (confident) when market is overvalued (undervalued). Recent research by Dimitrov and Jain (2016) shows that market timing, on average

is not lucrative with CAPE value above 27.6.

2.6. Fed model

Before the Fed model had officially received its name, the concepts it proposes were actively used by financial market players worldwide. The Fed's "Humphrey-Hawkins Report" of July 22, 1997 states: "changes in this ratio [S&P 500 prices-to-earnings] have often been inversely related to changes in long-term Treasury yields..." The term "Fed model" was coined by Ed Yardeni in his report on stock market overvaluation (1997). Yet, there is no evidence of it being in official use by the US Federal Reserve (Greenspan, 1997). The model is defined as the stock-bond yield gap, which is the difference between the stock earnings yield (E/P), the inverse of the common P/E ratio, and the 10-year Treasury bond yield (BY_t).

This metric is based on the fact that stocks and bonds are considered to be competing assets by the investors and postulates that by checking whether the inequality $E/P = BY_t$ holds, an investor can determine the most attractive of the two investment vehicles.

Salomons (2004) also shows that the Fed model is not feasible for forecasting equity returns. He suggests that model's predictive power over short horizons can be improved by adding an error term (EY_t^e), which accounts for the perceived risk of the two asset classes, to a regression. The interpretation is that the proposed term picks up market mispricing or changing investor risk aversion. Equity can be considered overvalued, indicating higher returns in short-term, if EY_t^e is high given a scenario with fixed volatility and interest rate. We will try to replicate and test the results of this study.

Koivu et al. (2005) introduce a quantitative dynamic version of Fed model, which based on co-integration analysis, can forecast changes in stock prices, earnings and bond yields. The model relies on use of logarithmic indicator, suggesting that by using logarithms of positive numbers can improve the model's fit better than initial data and in variations of positive variables can be more descriptive, compared to absolute values.

Clemens (2007) proposes that the inclusion of confidence intervals in the Fed framework improves the accuracy of predictions by reducing bias stemming from investor's irrational behaviour. However, he also points out that it is better suited for predicting relative returns of stocks versus bonds, as opposed to absolute stock returns. The author concludes that it is optimal to employ the Fed model for prediction over a short and medium horizon of up to 3 years, thereby complementing CAPE's long-run forecasts.

Finally, Davis et al. (2012) test forecasting abilities of 16 metrics using US stock returns data starting from 1926. The report indicates that dividend yield and Fed model have exhibited weak correlations with actual future stock returns. P/E ratios and CAPE have performed significantly better only at long investment horizons, explaining approximately 40% of variation in inflation-adjusted returns. Authors conclude that P/E ratios should not be used for short-term forecasts.

3. Theoretical Background

In this section, we will describe the theoretical foundation behind CAPE and Fed models in terms of Gordon Dividend Discount Model (DDM), covering the necessary assumptions and criticism against the two metrics

3.1. Price as common factor

Security price can be represented as present value of its future cash flows, $PV(FCF)$. For stocks holders, the most typical source of cash flow is dividends. In a simple case of value stocks, the long-term dividend growth rate, g , is assumed to be constant, as shown in the formula below:

$$P_0 = \frac{D_0(1+g)}{r-g} = \frac{D_1}{r-g} \quad (3.1)$$

Where P_0 is stock price at time 0; D_0 - dividend at time 0; D_1 - dividend at time 1; r - required rate of return on stock.

However, this is rarely the case. In reality, payouts may not be performed at all, as would be typical for growth stocks, or performed less frequently, Dividends may have a time-varying growth rate or not grow with time (formula 3.2).

$$P_0 = \frac{D_1}{r} \quad (3.2)$$

So, dividend is a choice variable. Let us examine this fact using the formula, where the dividend yield (D/P) is a function of earnings-to-price ratio (E/P) and dividend payout ratio (D/E).

$$\frac{D_0}{P} = \frac{E}{P} \frac{D_0}{E} \quad (3.3)$$

In good economic times with abnormal profits it is more reasonable to reinvest rather than pay out dividends (Lamont, 1998). This is represented by decreasing D/E , when profits (E) are higher. Alternatively, this can be justified by high stock prices (P), when economic conditions are favorable, leading to lower D/P .

Taking into account the formulas 3.1 and 3.3, P/E can be expressed as follows:

$$\frac{P}{E} = \frac{D_0}{E} \frac{(1+g)}{r-g} \quad (3.4)$$

3.2. CAPE Framework

The fundamental idea of stock return predictability from the simple P/E is extended to obtain the expression for current value of Shiller's $P/E10$:

$$CAPE_t = \frac{P_t}{\frac{1}{10} (E_{t-1} + E_{t-2} + \dots + E_{t-10})} \quad (3.5)$$

Where P_t represents the real spot price at time t and $E_{t-1}, E_{t-2}, \dots, E_{t-10}$ refer to the lagged real earnings. Consumer price index (CPI) data is used to allow conversion to real values.

Shiller posits that, based on the ratio, either numerator or denominator should be forecastable. Indeed, having that the stock price is the discounted sum of future earnings and the discount rate is the return required by investors, a valuation, P/E in this case, that relates current price with earnings can

give insight into the value of expected return. Overall, expected returns are negatively correlated with P/E , so that high P/E values resulting from either stock price growth or lower earnings would imply that expected returns should be low.

CAPE's framework contradicts the conventional random walk theory, which implies unpredictability of price changes in the stock market setting. Instead, CAPE works under assumption of valuation ratio stability and simple mean-reversion theory, whereby the stock prices fluctuate within historic boundaries and any significant deviation from the natural level relative to the earnings figures, is unlikely to persist.

CAPE has been drawing criticism for its failure to revert to historic mean. Siegel (2016) suggests that this can be attributed to CAPE not being robust to growth rate fluctuations in real earnings per share (EPS). High earnings growth causes a drop in ten-year average of past earnings compared to the current earnings. This drives CAPE up, which in turn results in downward bias of future stock returns forecast since the estimates were obtained from data with low earnings growth rates.

An alternative reason for metric's failure to mean-revert comes from the use of new accounting standards applying to investments into intangible assets or "e-capital" (Hall, 2000). These investments were considered current expenses and deducted from earnings, leading to a higher P/E ratio and lower expected future stock returns as a result. However, there is no confirmation that investments into intangible assets are in fact considerable enough in recent decades (historical time series data is not long enough).

Another theory focuses on baby-boomer (born 1946 - 1964) risk-tolerance. It posits that this demographic tends to favour stocks and is willing to pay higher price, resulting in P/E ratio increase. A version of this argument comes from Glassman and Hasset (1999), who suggest that the stock prices are too low, with zero risk-premium, leading to an abrupt surge in prices once investors realize this mispricing with baby-boomers being the first to see the errors in valuation ratios reflected in the stock prices of the day.

Additional argument is linked to inflation decline observed since 1980's. Authors express uncertainty about whether their bearish long-run forecast for stock-returns can be explained by the relationship between inflation and prices. Modigliani and Cohn (1979) argued about irrational behavior of market participants, namely that discounting real dividends at nominal interest rates leads to stocks undervaluation (overvaluation) when inflation is high(low), which was also confirmed by Ritter and Warr (2002).

Shiller (2014) expands the above list. He suggests that stock overvaluation could be explained by higher bond prices, resulting from low inflation (appx. 2%) and 10-year US Treasury notes yield (2.5%). The author suspects explanations may be rooted in psychology and sociology, with irrational exuberance coming into the equation. A less common explanation suggested that high CAPE may be a result of peoples' anxiety about the future lack of job security as a result of tech progress leading to layoffs.

3.3. Fed Model Framework

The Fed model can be expressed through the security's present value argument. The argument is based on the dividend discount model (DDM), stating that the price of stock is a discounted sum of future cash flows (FCF). Same principle can be applied to bonds trading at par value, where the income stream is represented by the annual yield. This relationship is shown in equation below:

$$\frac{P}{E} = \frac{1}{Y} \quad (3.6)$$

All else being equal, it then follows that when interest rates fall, $PV(FCF)$ rises leading to higher P/E for both stocks and bonds. Hence, the two values can be compared. This argument can be extended for use in determining the equilibrium price of stock, by solving the above equation for P . The received value can then be compared to the actual current market price (P^*). If the ratio P^*/P is above (below) 1, then the stock market is overvalued (undervalued). Lastly, the validity of this comparison is backed by the empirical evidence of historical co-movement of S&P 500 E/P and 10-year Treasury bond yields in

the period from 1965 to 2001. The two series exhibit a correlation of 0.81.

The competing assets argument mentioned earlier provides an alternative explanation to the Fed model. It follows that a scenario where stock yield (price) is higher (lower) than the bond yield (price), $E/P > BY$, implies that stocks are a more attractive investment alternative. On the other hand, $E/P < BY$ suggests stocks are more expensive, bearing lower yield, which should make them less attractive relative to bonds as in the period of 1991 - 1997 due to decrease in earnings during the recession. Finally, the $E/P = BY$ parity would suggest the two investment vehicles should be fairly valued and equally desirable (Asness, 2003).

The arguments in favour of Fed model were met with disapproval. The main criticism of the Fed model revolves around inflation illusion, whereby market participants tend to use real interest rates but nominal growth rates, leading to stock undervaluation when inflation rates are high, and vice versa (Modigliani and Cohn, 1979). Asness (2003) claims that stock's yield (E/P) is not its expected return. So, assuming all else is equal, when nominal expected return on stock moves simultaneously with bond yield, changes are reflected in expected nominal earning growth, not in E/P . This means that E/P does not describe investor's actual return as not all earnings are received by the investor. With earnings being related to inflation, E/P (or earnings yield) is a real return, whereas bond yield is clearly a nominal return. With respect to the PV argument, all else equal, when interest rates fall, it is true that PV of future CFs rises. Hence, current price rises and P/E should increase, too. However, all else is not equal. For instance, if inflation falls, future nominal CFs from equity also fall, which can offset the impact of lower rate, suggesting that the Fed model ignores this "counter-effect". Finally, if previous two arguments fail then the historic evidence of high correlation between S&P 500 E/P and 10-year US Treasury bond yield during the post-war period is just a proof of investors' blindness to biased model.

Thomas and Zhang (2008) suggest a counter-argument, stating that investors are rational, aware of inflation illusion and able to account for its three roots,

thereby avoiding it. The first root is that the earnings yield should not move with inflation. Thomas and Zhang respond that there are accounting policies which imply that the record of “inflationary holding gains” leads to co-movement of earnings (thus, earnings yield) with inflation. The second aspect of inflation illusion is that the nominal growth rates should move with inflation. The counter-argument is that the relevant growth rate should be a relatively stable perpetual dividend growth rate under a full, not current, payout policy. And according to accounting rules, this dividend growth rate does not exhibit significant variation with expected inflation, which contradicts the second argument. As their third response to inflation illusion, authors address the empirical observations in the “dynamic” market. They use 1976 - 2007 market data from US and 6 other countries. Findings indicate that the forecast errors, which used a “relatively constant” nominal growth rate, were uncorrelated with expected inflation. Additionally, the researchers argue that these forecasts rely on the logical outcome that in the periods of low inflation the expected equity market growth is higher, and vice versa. Their conclusion is that Fed model can be a useful valuation tool as it provides an intuition on risk premium and anticipated growth.

These findings were confirmed by Bekaert and Engstrom (2010), who are also very skeptical about the explanatory power of inflation illusion. In addition, they argue that the high positive correlation between real equity yields and nominal bond yields is a result of increased uncertainty, risk-aversion and higher expected inflation during economic downturn, which leads to higher stock and bond yields. These results were confirmed using a cross-sectional regression analysis of data from 20 countries, revealing that stagflation is highly correlated with high correlation between stock and bond yields.

Estrada (2009) shows that the Fed model can work, but the required assumptions are unrealistic. He considers a constant-growth dividend discount model:

$$P = \frac{D_0(1+g)}{r_f + RP - g} \quad (3.7)$$

With price (P), dividend (D_0), dividend growth (g), risk-free rate (r_f) and

risk premium for holding riskier stocks instead of bonds (RP). Both sides of the equation are divided by forward earnings (E) under the assumptions that all earnings are distributed as dividends ($D_0(1 + g) = E$) and that dividends do not grow ($g = 0$). The last assumption is that investors do not require higher returns from investing in stocks than in bonds ($RP = 0$). Finally, the Fed model itself is derived:

$$\frac{P}{E} = \frac{1}{r_f} \Rightarrow \frac{E}{P} = r_f \quad (3.8)$$

The result was supported by Siegel and Coxe (2002). Author argues that investor would always take into account inflation, and therefore consider higher growth and higher risk (approximately offsetting each other ($RP = g$)) of stocks rather than bonds. The Fed is derived as:

$$\frac{P}{E} = \frac{(1 + g)}{r_f} \quad (3.9)$$

Additionally, the Fed model can be adjusted to incorporate expected future inflation (I), as suggested by Arnott and Bernstein (2002), and a constant required risk-premium (RP), as suggested by Arnott and Bernstein (2002) and Asness (2003), respectively:

$$\frac{E}{P} = Y - I + RP \quad (3.10)$$

We will now proceed by providing the methodology for our empirical study.

4. Methodology

In-sample forecasts are believed to be comparably satisfactory as they use the same data set that was used to estimate the model's parameters (Brooks, 2008). Hence, the forecasting is conducted by splitting the data into an in-sample or fit period, which is used for model selection and parameter estimation, and an out-of-sample period, used to test forecast performance. Many researchers consider out-of-sample forecast performance as the "ultimate test of a forecasting model" since in-sample performance can be more responsive to data mining and outliers (Stock and Watson, 2007).

The main interest is to assess the models' pseudo out-of-sample h -step ($h \geq 0$) ahead forecast performance, and compare it with the parameters of in-sample fit. The process calls for splitting the available data (T) into a fit period and hold-out sample (H) which first forecast is estimated using the observations up to k , the second - up to $k + 1$, and the last is up to $T = k + H - 1$, given the total sample size equal $T + h = k + H - 1 + h$. If $E_{T+h}(\cdot)$, where $T + h$ is the sample size, is a loss function of the forecast errors, the framework implies absolute and relative forecasting performance measures.

In-sample tests imply using the fit period to estimate the model of interest. Predicted values are then compared to actual data. This type of tests has higher power (R^2) and are considered more credible. However, in-sample tests are also considered unreliable by some as they often erroneously exhibit spurious predictability (Kilian and Taylor, 2003).

Out-of-sample forecasting is conducted using the coefficient estimates obtained from prior in-sample analysis and running dynamic recursive 1-step regressions, where initial date (forecast origin) is fixed to point k while additional data points are added sequentially ($k + 1$, $k + 2$, etc). Obtained forecasts are compared to actual data in the hold-out period. In smaller samples, this type of test may fail to correctly detect predictability, thus having lower (R^2) than an in-sample test of same size.

4.1. Hypotheses

Based on the previous research, we formulate the following primary hypotheses:

H_0 : CAPE is equivalent to Fed model in forecasting stock returns in-sample;

H_0 : CAPE is equivalent to Fed model in forecasting stock returns out-of-sample.

4.2. Regression analysis

Empirical analysis of short- and long-term forecasting ability of chosen models will be performed using ordinary least squares (OLS) predictive regressions, which are subject to the assumptions of Classical Linear Regression Model

(CLRM) (see Appendix 1(a)). The general form of first-order unconstrained predictive multiple linear regression for our dependent variable, future stock returns (Y_{t+1}), for $t = 1, 2, \dots, T$ would take the following form:

$$Y_{t+1} = \alpha + \beta_1 X_{t,1} + \beta_2 X_{t,2} + \dots + \beta_T X_{t,T} + \varepsilon_t \quad (4.1)$$

Where constant α is the intercept, constants $\beta_1, \beta_2, \dots, \beta_T$ are coefficients, which represent the mean change of the dependent variable resulting from point increase in regressors $X_{t,1}, X_{t,2}, \dots, X_{t,T}$. ε_t is the error term, describing variation in the dependent variable unexplained by the model.

Running the regression generates fitted (predicted) values \widehat{Y}_{t+1} :

$$\widehat{Y}_{t+1} = \widehat{\alpha} + \widehat{\beta}_1 X_{t,1} + \widehat{\beta}_2 X_{t,2} + \dots + \widehat{\beta}_T X_{t,T} + \varepsilon_t \quad (4.2)$$

Regression estimators $\widehat{\alpha}$ and $\widehat{\beta}$ are obtained as described in Appendix 1(b).

In the OLS setting, the forecast model's goodness of fit, R^2 , is computed using total sum of squares (SST), sum of squared errors (SSE), as can be seen in Appendix 1(c).

4.3. Unit-root test

Non-stationarity is a common problem with time-series data. Generally, it is reflected in presence of time trend characterized by time-varying mean and variance, which does not revert to long-run average. Forecasting works under the assumption of data stationarity because lack thereof may result in upward biased R^2 for variables that have no effect on the dependent variable in reality. Primarily, this is the case for in-sample but applies out-of-sample to some extent. Hence, to yield meaningful results, data should be stationarized. To identify non-stationarity in the data used, Augmented Dickey-Fuller (ADF) test will be employed. ADF test will be supplemented by graphical analysis of the time-series plot.

If the data is identified as non-stationary, it will be detrended by taking first differences. However, the loss of data as a result of differencing may lead to inferior out-of-sample performance. So, the same regressions will be run using non-detrended data and compared to results obtained with detrended data.

4.4. Forecasting Frameworks

4.4.1. Generic CAPE

For generic version of CAPE, all constants in regression 4.1 except α and β_1 are considered zero and natural logarithm of CAPE ratio, $\ln CAPE_t$, replaces $X_{t,1}$. Then, the future real stock return (Y_{t+1}) is regressed on the logarithm of $CAPE_t$ using the following simple linear regression:

$$Y_{t+1} = \alpha + \beta_1 \ln CAPE_t + \varepsilon_t \quad (4.3)$$

Regression coefficients, intercept (α), slope (β_1), and the error term (ε_t) are estimated.

4.4.2. After-tax Profit CAPE

As suggested by Siegel (2016), after-tax profit data is used to replace reported and operating earnings of S&P 500 index and international alternatives where available.

We take the simple linear regression 4.3 used for generic CAPE and replace the independent variable with $\ln CAPE_t^*$:

$$Y_{t+1} = \alpha + \beta_1 \ln CAPE_t^* + \varepsilon_t \quad (4.4)$$

Additionally, the regressions (4.3) and (4.4) will be augmented with bond yield (BY_t) to check whether it provides additional predictive power:

$$Y_{t+1} = \alpha + \beta_1 \ln CAPE_t + \beta_2 BY_t + \varepsilon_t \quad (4.5)$$

$$Y_{t+1} = \alpha + \beta_1 \ln CAPE_t^* + \beta_2 BY_t + \varepsilon_t \quad (4.6)$$

4.4.3. Generic Fed model

Equity returns (Y_{t+1}) are forecasted using three regressions for the Fed model as proposed by Asness (2003). First two equations mimic the structure of in regression (4.1) with the regressor being replaced by equity yield (EY_t) and the equity premium ($EY_t - BY_t$) for equations 4.7 and 4.8 respectively:

$$Y_{t+1} = \alpha + \beta_1 EY_t + \varepsilon_t \quad (4.7)$$

$$Y_{t+1} = \alpha + \beta_1(EY_t - BY_t) + \varepsilon_t \quad (4.8)$$

Third regression (4.9) is bi-variate, taking into account (EY_t) and 10-year treasury bond yield (BY_t) as separate variables:

$$Y_{t+1} = \alpha + \beta_1 EY_t + \beta_2 BY_t + \varepsilon_t \quad (4.9)$$

We have α, β_1, β_2 , – constants, EY_t - equity yield, BY_t – bond yield, ε_t - error term.

Comparing the simple CAPE and Fed models, it becomes apparent that CAPE relies on the use of 10-year moving average (MA) of earnings. This difference leads to smoothed time-series and CAPE can be expected to have higher power (R^2) when forecasting long-run expected stock returns.

4.4.4. *Fed model adjusted for perceived risk*

As previously mentioned, the relation between bond yield and earnings yield is theoretically flawed and unstable, leading to Fed model exhibiting poor predictive power over long horizons. Seeking to improve the Fed model, Salomons (2004) proposed an alteration of the generic version, whereby an error term is added to account for volatility of bonds and equity. This adjustment for perceived risk indicates how earnings yield is set as function of bond yield by investors. Depending on value of error term, relative allocations between equity and bonds can be determined for an optimal portfolio strategy.

We will test whether adjustment for perceived risk indeed improves the predictive power of Fed model given the available US data. For this, two variables are added to the generic model: fitted value (EY_t^f) and error term (EY_t^e). They are obtained using the regressions presented below:

$$EY_t = \alpha + \beta_1 BY_t + \varepsilon_t \quad (4.10)$$

$$EY_t^f = \alpha + \beta_1 BY_t + \beta_2 \sigma_t^e + \beta_3 \sigma_t^b + \varepsilon_t \quad (4.11)$$

Where adjustments for risk σ_t^e and σ_t^b are the equity and bond volatility, respectively. Asness (2003) posits that investors assume E/P is a function of nominal interest rates (positive β_1) and require higher E/P compared to BY when stocks are more volatile than bonds (positive β_2 and negative β_3). Then, the error term is:

$$EY_t^e = EY_t - EY_t^f \quad (4.12)$$

Finally, the predictive power of EY_t , EY_t^f and EY_t^e is tested by running the three regressions:

$$Y_{t+1} = \alpha + \beta_1 EY_t + \varepsilon_t \quad (4.13)$$

$$Y_{t+1} = \alpha + \beta_1 EY_t^f + \varepsilon_t \quad (4.14)$$

$$Y_{t+1} = \alpha + \beta_1 EY_t^e + \varepsilon_t \quad (4.15)$$

More specifically, regression 4.14 tests the predictive ability of the perceived correct earnings yield, while regression 4.15 tests the error terms.

4.5. In-sample testing

Available data has been split as presented in Table 1 and respective samples are used to run the predictive in-sample regressions from section 4.4. These provide us with necessary estimates for regression coefficients and allow to test, whether a specific model has predictive power in-sample. The ability of the four models to generate accurate stock return forecasts in-sample is evaluated based on their t-statistic and R^2 .

To test the predictability, we formulate the following null and one-sided alternative hypotheses for in-sample tests (Inoue and Kilian, 2005):

H_0 : Model exhibits no predictive ability, $\beta = 0$;

H_1 : Model exhibits predictive ability, $\beta > 0$.

Table 1: US Data in- and out-of-sample allocations

	In-sample		Out-of-sample		T(IS) + T(OS)	T(IS)/T(OS)
	Period	T(IS)	Period	T(OS)		
$CAPE_t$	1881:01-1970:01	1069	1970:02-2017:05	568	1637	1.88
$CAPE_t^*$	1957:02-1987:02	361	1987:03-2017:02	360	721	1.00
EY_t	1871:02-1960:02	1069	1960:03-2016:11	681	1750	1.57
$EY_t - BY_t$	1871:02-1960:02	1069	1960:03-2016:11	681	1750	1.57
$EY_t \& BY_t$	1871:02-1960:02	1069	1960:03-2016:11	681	1750	1.57
EY_t	1891:02 -1970:02	949	1970:02 - 2016:11	561	1510	1.69
EY_t^f	1891:02 -1970:02	949	1970:02 - 2016:11	561	1510	1.69
EY_t^e	1891:02 -1970:02	949	1970:02 - 2016:11	561	1510	1.69
<hr/>						
σ_t^e	1891:02 -2016:12					
σ_t^b	1891:02 -2016:12					
BY_t	1891:02 -2016:12					

Note: The table describes time-series samples used for the in-sample ($T(IS)$) and out-of-sample ($T(OS)$) regressions. Full sample and the ratio of the two samples are presented in the last two columns respectively.

Where $\beta^* = 0$. Test is performed at significance level of $\alpha = 0.05$ and H_0 is rejected if t-statistic (4.16) exceeds the corresponding critical value.

$$t - statistic = \frac{\hat{\beta} - \beta^*}{SE(\hat{\beta})} \quad (4.16)$$

4.6. Heteroskedasticity testing

The forecasting frameworks work under the assumption of homoskedasticity, i.e 2nd CLRM assumption mentioned previously. Conversely, when heteroskedasticity is present, standard errors (SE) are biased, leading to bias in test-statistics and confidence interval. So, the model may produce misleading parameter estimates.

White's test for heteroskedasticity is conducted. If the null hypothesis of homoskedasticity in error distribution is rejected, the data is corrected using Newey-West Heteroskedasticity and Autocorrelation Consistent (HAC) form. This provides heteroskedasticity robust SE , which relax the either or both IID assumptions imposed by OLS on errors: independent (3rd CLRM assumption) and identically distributed (5th CLRM assumption). This provides reasonable p-values while not changing coefficient estimates.

4.7. Pseudo-out-of-sample testing

4.7.1. Check for outliers in forecast error time series

As the data contains periods of Great Depression (1929 - 1939), dot-com bubble (1997 - 2001), Global financial crisis (2007 - 2008), and other unusual events, it is inevitable to have outliers in it. Increased errors and distortions of parameters and estimates are some of the effects of the outliers on statistical analysis (Zimmerman, 1994). In addition, outliers can potentially reduce the power of statistical tests, decrease normality, increasing the risk of making the Type I and II errors as a result, and influence estimates (Rasmussen, 1988). Thus, visual inspection of the out-of-sample forecast errors needs to be conducted. Scatter plots of time-series of errors will be presented and analyzed. The outliers should be considered when making conclusions and will not be removed from the data to retain the representativeness of population (Orr et al., 1991).

4.7.2. Comparison of forecast statistics

Using coefficients obtained by running the in-sample regressions, we run appropriate unrestricted regressions of form 4.2 to generate out-of-sample forecasts for the hold-out sample.

Many researchers have proposed guidance on what measure one should use to compare the forecast performance of different models for time-series data. However, some of them can be inapplicable and produce misleading results for the given data set, as the measures can be infinite or undefined (Hyndman and Koehler, 2006). Given the different scales across data sets for the two models and the presence of outliers, various parameters are used for the comparison of effects. These are: mean error (ME), root mean squared error ($RMSE$), mean absolute error (MAE), mean percentage error (MPE), mean absolute percentage error ($MAPE$) and Theil's U (U). Formulas for obtaining the parameters are presented in Appendix 2.

ME , $RMSE$ and MAE are scale-dependent measures, in other words these can be useful to compare different models for the same data set. Hence, the

parameters can be relied upon for comparing forecasting methods using level and differenced data, but cannot be used to compare, for instance, CAPE and Fed which use sets of observations with different lengths. ME and $RMSE$ have been commonly used in statistical modeling due to their theoretical relevance, but they are more affected by outliers than MAE is (Hyndman and Koehler, 2006).

MPE and $MAPE$ are scale-independent measures based on percentage errors. These two are considered more reliable when comparing models that use different data sets. The disadvantage of these parameters is that they are “asymmetric”, whereby they penalize positive errors (forecasted value exceeds actual) more than the negative (forecasted value is less than actual) (Armstrong and Collopy, 1992).

All the measures above need to be as low as possible to indicate a superior forecast performance of the model. In addition, positive values indicate that the variable was underestimated, and vice versa.

Final parameter is Theil’s U . It compares $RMSE$ of the forecast obtained from the model with the $RMSE$ from the benchmark model (random walk). The desirable value of Theil’s U statistic is less than 1, as it implies that the forecasting technique is better than guessing. If $U = 1$, the forecasting model is as good as guessing while $U > 1$ would imply that guessing offers results superior to the forecast model.

Examination of predictive power will be conducted over investments of the following lengths: 3 months, 6 months, 1 year, 5 years, 10 years and 20 years.

4.7.3. Testing equality of forecast accuracy

In addition to the forecast metrics, Morgan-Granger-Newbold (MGN) and Diebold-Mariano (DM) tests will be conducted to determine whether the select pairs of predictive models in fact show equal forecast accuracy out-of-sample (Diebold and Mariano, 2002).

Morgan-Granger-Newbold test

MGN uses forecast errors, e_{it} and e_{jt} obtained from two forecasting models to form orthogonalization of the form $X_t = e_{it} + e_{jt}$ and $Z_t = e_{it} - e_{jt}$. This test functions under the assumptions that loss is quadratic, forecast errors are white noise, have zero mean and are serially as well as contemporaneously uncorrelated. The null hypothesis for MGN test states that two forecasting models have equal forecast accuracy, which is equivalent to zero correlation between X and Z ($T \times 1$ vectors of X_t and Z_t , respectively), $\rho_{X,Z} = 0$.

The test-statistic for the MGN test (distributed as Student's with $T - 1$ degrees of freedom (df)), takes the following form:

$$MGN = \frac{\rho_{X,Z}}{\sqrt{\frac{(1 - \rho_{X,Z}^2)}{T - 1}}} \quad (4.17)$$

Where correlation $\rho_{X,Z}$ is computed as shown in the formula 4.18. X' and Z' are transposed vectors.

$$\rho_{X,Y} = \frac{X'Z}{\sqrt{(X'X)(Z'Z)}} \quad (4.18)$$

Test will be conducted for same out-of-sample horizons as well as the equal full sample from 1987:03 to 2016:11. The sign of test-statistic will not be interpreted to tell which of the forecasts is more accurate. Instead, the results will be judged in conjunction with respective forecast statistics.

Diebold-Mariano test

Diebold-Mariano method is arguably simpler than the MGN test. Assumptions are made directly about the forecast errors (e_t) and associated loss at time t , $L(e_t) = e_t^2$. Namely, method of Diebold and Mariano (2002) is suitable for non-quadratic loss functions with non-Gaussian forecast errors, which are not zero-mean and are serially and contemporaneously correlated. The only requirement is that loss differential between forecast obtained from two models, $d_{ijt} = L(e_{it}) - L(e_{jt})$, is covariance stationary.

DM is an asymptotic z-test with a null hypothesis that the loss differential is zero, $E(d_{ijt}) = 0$, which would imply that two forecasts are equally accurate. The null is rejected when $E(d_{ijt}) \neq 0$ if the absolute value of DM test-statistic exceeds the critical value of a standard unit Gaussian distribution. The DM-statistic is obtained as a fraction of sample mean loss differential, $\overline{d_{ij}} = \frac{1}{T} \sum_{t=1}^T (d_{ijt})$, and the consistent estimate of its standard deviation ($\widehat{\sigma d_{ij}}$):

$$DM = \frac{\overline{d_{ij}}}{\widehat{\sigma d_{ij}}} \quad (4.19)$$

Where ($\widehat{\sigma d_{ij}}$) is equal to regression standard error obtained by running simple OLS regressions of the form: $d_{ijt} = \alpha + \varepsilon_t$.

Just as MGN test, DM will be conducted for 6 horizons out-of-sample and the the full sample from 1987:03 to 2016:11. The sign of DM test-statistic is can help identify the best forecast of the two. For instance, a negative test-statistic would imply that the forecast from the first model is better (lower loss associated with forecast errors).

5. Data

Data on S&P 500 Index prices, earnings, dividends and consumer price index (CPI) for this study has been sourced from Robert Shiller's web site. The sample includes monthly observations in the period of January 1871 through December 2016, giving the ability to measure intrinsic values of securities. Monthly US 10-year constant maturity Treasury bond yield observations have also been obtained from Shiller's web site. The sample of actual historical observations is 1962:01 to 2016:12. Shiller compensated the lack of data prior to 1962:01 by interpolation.

Monthly National Income and Product Accounts (NIPA) US corporate after-tax profits data for use in alternative CAPE version proposed by Siegel is obtained from quarterly using cubic spline interpolation method, as opposed to linear interpolation used by Shiller. Resulting sample period is 1953:02 to 2017:05.

Adjustment for perceived risk is accomplished using the monthly equity volatility (σ_t^e) time series, obtained from a rolling window of S&P 500 returns with a look-back period of 20 years. Monthly Treasury bond yield volatility (σ_t^b) is computed identically using Shiller’s interpolated data. Resulting data sets are both from 1891:02 to 2016:12.

6. Empirical findings and Analysis

In this section, we will provide the results of our empirical study, their analysis and interpretations.

6.1. Descriptive statistics

Descriptive statistics for variables used in the study are presented in Table 2.

Table 2: Time series descriptive statistics

	Mean	Median	Stdev	Min	Max	Jarque-Bera	ADF(level)	ADF(FD)	T
R	4.65	4.76	4.45	-23.08	64.05	32833.9 (0)	1.41E-25	3.93E-38	1750
$CAPE_t$	2.74	2.78	0.40	1.57	3.79	15.72 (0,0004)	0.26	6.07E-20	1637
$CAPE_t^*$	0.37	0.45	0.42	-0.70	1.16	33.0545 (6,64E-08)	0.81	3.6E-20	721
BY_t	4.41	3.91	2.41	-3.78	14.94	742.45 (6,005E-162)	0.73	5.28E-31	1750
EY_t	7.40	6.82	2.70	0.81	18.82	374.26 (5,39E-82)	0.0008	4.78E-22	1750
σ_t^e	0.02	0.01	0.00	0.01	0.03	135.19 (4,41E-30)	0.59	3.04E-07	1511
σ_t^b	0.04	0.04	0.01	0.03	0.08	1031.03 (1,30E-224)	0.70	1.70E-12	1511

Note: The table contains descriptive statistics for the time series used in empirical analysis. Mean, median, standard deviation, minimum, maximum and number of observations are reported for each variable. Results of Jarque-Bera normality test and p-values for Augmented Dickey-Fuller (ADF) test for unit-root in level and first difference with constant and trend are also presented for all listed variables.

According to the standard deviation, S&P 500 return has the highest volatility compared to other variables. For Jarque-Bera normality test (joint test for 0 skewness and 0 excess kurtosis), based on the statistically significant (high) test-statistics and respective p-values close to zero, a conclusion can be drawn that the time series for variables are not normally distributed, as would be

expected for financial metrics and economic data. The condition for rejection of the unit-root null is $P < 0.05$, so only equity returns (R) and earnings yield (EY_t) time series are stationary in level (ADF(level)). Differencing completely removed the time trend from all series as can be seen from the extremely low p-values in the ADF(FD) column.

6.2. In-sample regression statistics

The in-sample forecast accuracy comparison among the models is based on R^2 and t-statistic measures. Results, corrected for heteroscedasticity and serial correlation, are presented in Table 3 for ten regressions in level and eight regressions in first difference (FD).

Table 3: In-sample regression statistics (Level and FD)

Regression	Regressor	Level		FD	
		t-statistic	R squared	t-statistic	R squared
(4.3)	$CAPE_t$	-6,081***	0.0779	6,369***	0.0439
(4.4)	$CAPE_t^*$	-3,701***	0.0529	5,56***	0.0619
(4.5)	$CAPE_t$ & BY_t	-6,082***	0.0779	6,455***	0.0459
		0.01499		-1,753	
(4.6)	$CAPE_t^*$ & BY_t	-3,666***	0.0750	0.4933	0.0017
		-1,974*		-0,4651	
(4.7)	EY_t	4,595***	0.0440	-	-
(4.8)	$EY_t - BY_t$	3,155***	0.0134	-	-
(4.9)	EY_t & BY_t	4,867***	0.0592	1.45	0.0066
		4,219***		1.253	
(4.13)	EY_t	4,924***	0.0507	1.211	0.0041
(4.14)	EY_t^f	2,024*	0.0063	1.418	0.0014
(4.15)	EY_t^e	4,04***	0.0336	0.1421	0.0000

Note: The table reports in-sample regression statistics for regressions in level (4.3 - 4.9 and 4.13 - 4.15) and first difference (4.3 - 4.6, 4.9 and 4.13 - 4.15). Regression's t-statistic is stated with the significance level denoted by (*), (**) and (***) for 5%, 1% and 0.1%, respectively. Regression's power denoted in R squared (R^2).

CAPE model augmented with bond yield in level (eq. 4.5) has R^2 of 7.8%, the highest among the models. Generic CAPE (4.3) has R^2 of 7.79% which is 3.4% higher than its differenced alternative. The lowest model fit is shown by the differenced Fed model adjusted for perceived risk, which uses error term as a regressor (4.15). After-tax corporate profits CAPE (4.4) has R^2 of 7.5% in level but its fit in difference is 43 times lower, indicating significant loss of

information due to detrending. Almost all in-sample regressors show 99.95% significant t-statistic results in level, except for the bond yield (4.6) and fitted value in Fed model (4.14) which show 97.5% significance. Bond yield (4.5) is not statistically significant. However, only three variables show significant results in difference: generic CAPE (4.3), after-tax profit CAPE (4.4), generic CAPE with bond yield (4.5).

6.3. Pseudo-out-of-sample regression statistics

Table 4 contains the pseudo-out-of-sample regression statistics in level. It can be seen that, most coefficients are significant at 0.1% level and only the bond yield coefficient in regression (4.5) and intercept in (4.9) show no statistical significance at the 3 levels.

Table 4: **Out-of-sample regression statistics (Level)**

Sample period	Regression	α	β_1	β_2	R^2
1881:02 - 1970:01	(4.3)	0,143*** (9,004)	-0,035*** (-6,081)		0.044
1957:02 - 1987:02	(4.4)	0,043*** (15,65)	-0,016*** (-3,701)		0.062
1881:02 - 1970:02	(4.5)	0,143*** (8,644)	-0,035*** (-6,082)	0.002 (0,015)	0.078
1957:02 - 1987:02	(4.6)	0,069*** (5,003)	-0,033*** (-3,666)	-0,330** (-1,974)	0.075
1871:02 - 1960:03	(4.7)	2,394*** (3,523)	0,387*** (4,595)		0.044
1871:02 - 1960:03	(4.8)	4,700*** (15,06)	0,177*** (3,155)		0.013
1871:02 - 1960:03	(4.9)	1.041 (1,383)	0,40*** (4,867)	0,364*** (4,219)	0.007
1891:02 - 1970:03	(4.13)	0.02*** (3.064)	0.403*** (4.924)		0.051
1891:02 - 1970:03	(4.14)	0.024* (1.912)	0.363** (2.024)		0.006
1891:02 - 1970:03	(4.15)	0.05*** (24.87)	0.31*** (4.04)		0.034

Note: The table reports regression statistics for regressions 4.3 - 4.9 and 4.13 - 4.15 in level. Respective sample periods are specified in the first column. Regression coefficients (α, β_1, β_2) are presented with a significance level, denoted by (*), (**) and (***) for 5%, 1% and 0.1%, respectively. T-statistics are presented in the parentheses below the appropriate coefficients. Regression's power denoted by R squared (R^2) is specified in the last column.

Detrending non-stationary data for all regressions, except 4.7 and 4.8 yields the regression statistics presented in Table 5. It can be noted that while all intercepts are statistically significant at 0.01%, only β_1 coefficients of generic

CAPE (4.3), after-tax CAPE (4.4) and generic CAPE with bond yield (4.5) are significant at 0.01% level. β_2 is only significant at 0.05% for (4.5). So, it can be observed that while differencing effectively stationarizes the time series, the resulting loss of information in regressions 4.6, 4.9 and 4.13 - 4.15 is considerable.

Table 5: **Out-of-sample regression statistics (FD)**

Sample Period	Regression	α	β_1	β_2	R^2
1881:02 - 1970:01	(4.3)	0,049*** (31,73)	0,207*** (6,369)		0.044
1957:02 - 1987:02	(4.4)	0,039*** (18,94)	0,245*** (5,560)		0.062
1881:02 - 1970:02	(4.5)	0,049*** (31,84)	0,211*** (6,455)	-0,141* (-1,753)	0.046
1871:02 - 1987:03	(4.6)	0,039*** (15,82)	0.03 (0,493)	-0,201 (-0,465)	0.002
1871:02 - 1960:03	(4.9)	5,511*** (28,73)	0.891 (1,450)	0.092 (1,253)	0.007
1891:03 - 1970:03	(4.13)	0.052*** (24.02)	0.797 (1.211)		0.004
1891:03 - 1970:03	(4.14)	0.052*** (24.77)	0.327 (1.418)		0.001
1891:03 - 1970:03	(4.15)	0.052*** (24.73)	0.045 (0.142)		0.000

Note: The table reports regression statistics for regressions 4.3 - 4.6, 4.9 and 4.13 - 4.15 using first differences of variables. Respective sample periods are specified in the first column. Regression coefficients (α , β_1 , β_2) are presented with a significance level, denoted by (*), (**), and (***) for 5%, 1% and 0.1%, respectively. T-statistics are presented in the parentheses below the appropriate coefficients. Regression's power denoted by R squared (R^2) is specified in the last column.

As it was expected, R^2 of after-tax corporate profits CAPE is substantially higher than that of the generic version, which is consistent with Siegel (2016). Addition of bond yield increases R^2 of both generic and after-tax corporate profits CAPE models by 2-3% in level. Interestingly, bond yield is significantly negative on 99% for after-tax profits CAPE but insignificant for the generic version. As for detrended data results, CAPE with bond yield added increases model fit in comparison to generic CAPE model. However, addition of bond yield into after-tax corporate profits CAPE reduces R^2 by significant amount of 6% in difference.

Figure A.1 shows the time series of actual S&P 500 returns, its long-run mean, and returns forecasted by generic CAPE model, after-tax corporate profits

CAPE and versions of the two augmented with bond yield from 1987:03 to 2016:11. It can be observed that all of the model versions follow a similar pattern, fluctuating around the mean of S&P 500 return (2.63%) within a 1-5% range. Figure A.2 offers a closer look at the time series of forecasting models, which appear to be picking up the major shocks of 1987, 1990, 1998 and 2009 but with a magnitude different from that of observed returns. Both CAPE with bond yield and NIPA CAPE with bond yield exhibit a stronger reaction to shocks.

Figure A.3 plots the time series of observed S&P 500 return with its long-run historic mean and returns predicted by generic Fed and risk-adjusted Fed model. Generic Fed model and its alternatives fluctuate in the range from 3% to approximately 5%. They move above the mean return (2.63%) and do not intersect it at any point of time examined. By zooming in on the time-series of forecasting models in figure A.4, one can say that generic model tends to follow the shocks but in lower magnitude than the observed values in periods of 1987, beginning of 2000's, 2009 and 2011.

Figure A.5 compares the time series of observed returns of S&P 500 and those predicted by traditional CAPE and Fed models from 1970:02 to 2016:11. Fed moves mostly above the mean return, except for 2009 when it predicts returns slightly below the mean. CAPE is more volatile, and it crosses mean return straight line few times. It can be seen on the figure A.6, that starting far above the mean in 1970, it intersects the line in May of 1995 and plunges to the largest negative spike in February, 2000. Both models correctly follow the shocks in the periods of Global financial crisis 2007-2009.

6.4. Check for outliers in forecast error time series

Visual analysis of the out-of-sample forecast errors with outliers more than 10% (in blue) is provided for the CAPE, CAPE with bond yield, after-tax profits CAPE, after-tax profits CAPE with bond yield, Fed model and Fed model adjusted for perceived risk. The plots are provided in Appendix 4.

Figure A.7 shows the scatter plot of CAPE model forecast errors for the period 1970:02-2017:05. It is clear that the most distant point (approx. -22.5%) is

September of 2008, which is known as a critical stage in the Global financial crisis. Unsurprisingly, the periods from June 1974 to October 1987, known as periods of stock market crash, oil crisis and US recession, are outlying points on the plot (approx. -15%). Recession of early 2000's is also depicted on the plot at February-August 2001 points at approx. -10% to -13%. July-August 2011 US stock market fall due to the threat of European sovereign debt crisis is at -12.5% point on the plot. All outliers are below -10%, meaning that CAPE, in general, shows upward bias when predicting expected equity returns. This is consistent with Shiller's most recent observations (Shiller, 2014).

As shown in Figure A.8, forecast errors of the after-tax corporate profits CAPE for the period of 1987:03 to 2017:02, follow the same pattern as generic CAPE model. One exception is January 1991 (Persian Gulf War) point which apparently has an impact on NIPA corporate profits (10%) but not on the S&P 500 equity index.

Figure A.9 shows the scatter plot of errors of the CAPE model with US Treasury bond yield for the period of 1970:02 to 2016:12. The lower part of the plot resembles that of generic CAPE, but the upper part has two outliers at 10%: August 1982 and January 1991 (drop of Dow Jones Industrial Average). In contrast, the upper part of the after-tax profits CAPE with bond yield (Figure A.10) has only one outlier in December 1990 (drop of Nikkei index). The lower part is the same.

Three plots of generic Fed model (Figures A.11 - A.13) have the same outlying points as previously described models, and also one additional point at February 2009 (14%). Similar structure of outlying forecast errors can be observed for the Fed model adjusted for perceived risk (Figures A.14 - A.16) with the outlying point in February 2009 now at 10%.

6.5. Forecast evaluation statistics

6.5.1. *Generic CAPE*

Forecast regression statistics for generic CAPE model for different horizons using variables in level and first difference are presented in Tables 6 and 7, respectively.

In both cases only the model which uses a 1-year and 20-year horizon forecast shows positive R^2 . Mean error (ME) has the lowest value in the shortest horizon used (3 months) in both level and difference. However, when it comes to the root mean squared error ($RMSE$) and mean absolute error (MAE), a 1-year and 20-year forecasts have the lowest values. Mean percentage error (MPE) statistics are distributed from highest to lowest across horizons, with positive values for the short-term horizons and negative for the long-term. Similarly, Theil's U values are distributed from highest to lowest as horizon increases. Mean absolute percentage error ($MAPE$) shows mixed results, yet has the lowest value in a 1-year horizon. Overall, the results for level and difference variables are almost identical and have the same pattern in all error statistics.

Table 6: CAPE forecast evaluation statistics (Level)

		3m	6m	1yr	5yr	10yr	20yr
$CAPE_t$	R^2	-0.29126	-0.23039	0.080558	-0.15715	-0.15982	0.000843
	ME	-0.05778	-0.03126	-0.00381	-0.01919	-0.01873	-0.01354
	RMSE	0.079389	0.057429	0.044206	0.048452	0.042819	0.039823
	MAE	0.059436	0.038134	0.030723	0.036226	0.032333	0.02987
	MPE	739.15	357.44	176.04	-11.797	-42.138	-117.76
	MAPE	739.15	390.51	199.39	340.78	266.79	300.45
	Theil's U	1.5854	1.584	1.5818	1.0858	1.0759	0.34344

Note: The table contains R^2 , ME , $RMSE$, MAE , MPE , $MAPE$ and Theil's U values for the in level regression 4.3 of the generic CAPE model. Forecast valuation statistics are available for 6 forecasting horizons. Values arranged from worst (red) to best (green) performance.

Table 7: CAPE forecast evaluation statistics (FD)

		3m	6m	1yr	5yr	10yr	20yr
$CAPE_t$	R^2	-0.2588	-0.18069	0.155728	-0.11262	-0.05624	0.078255
	ME	-0.05714	-0.02526	-0.00596	-0.01841	-0.01231	-0.00746
	RMSE	0.078385	0.056257	0.04236	0.047511	0.040862	0.038249
	MAE	0.061586	0.036892	0.027336	0.036364	0.030549	0.02821
	MPE	844.51	427.77	200.05	-23.308	-28.065	-103.71
	MAPE	844.51	435.12	216.46	336.18	244.81	280.16
	Theil's U	1.5268	1.5254	1.5232	1.0395	1.0002	0.33806

Note: The table contains R^2 , ME , $RMSE$, MAE , MPE , $MAPE$ and Theil's U values for the first difference regression 4.3 of the generic CAPE. Forecast valuation statistics are available for 6 forecasting horizons. Values arranged from worst (red) to best (green) performance.

Addition of the bond yield to the Generic CAPE model (Tables 8 and 9), both in level and in first difference, reduces R^2 for 1-year time horizon but increases it for 3 months, 6 months, 5 years, 10 years and 20 years. The effect on ME and MPE is negative: the adjusted model shows higher values than generic on all horizons. However, the model shows lower $RMSE$, MAE for all horizons except the 1-year. Looking at $MAPE$ values, addition of bond yield increases the forecast performance in long horizons. Theil's U shows that the adjusted model shows better results in all except 20-year forecast horizons.

Table 8: CAPE with BY forecast evaluation statistics (Level)

		3m	6m	1yr	5yr	10yr	20yr
$CAPE_t$ & BY_t	R^2	-0.28732	-0.22573	0.075831	-0.12737	-0.08174	0.06775
	ME	-0.05777	-0.03074	-0.00329	-0.01784	-0.01491	-0.00602
	RMSE	0.079318	0.057342	0.044302	0.047829	0.041346	0.038451
	MAE	0.059388	0.037938	0.030828	0.035878	0.031046	0.028796
	MPE	739.96	359.12	177.29	-10.887	-32.965	-89.736
	MAPE	739.96	390.14	199.4	332.99	252.31	264.38
	Theil's U	1.5836	1.5821	1.5799	1.0761	1.0542	0.4239

Note: The table contains R^2 , ME , $RMSE$, MAE , MPE , $MAPE$ and Theil's U values for the in level regression 4.5 of the generic CAPE model with addition of bond yield (BY). Forecast valuation statistics are available for 6 forecasting horizons. Values arranged from worst (red) to best (green) performance.

Table 9: CAPE with BY forecast evaluation statistics (FD)

		3m	6m	1yr	5yr	10yr	20yr
$CAPE_t$ & BY_t	R^2	-0.25441	-0.17835	0.155664	-0.10945	-0.05438	0.081537
	ME	-0.05746	-0.02525	-0.00612	-0.0184	-0.01226	-0.00747
	RMSE	0.078298	0.056223	0.042346	0.047448	0.04082	0.038165
	MAE	0.06129	0.03704	0.027512	0.036286	0.030504	0.028151
	MPE	847.64	430.09	200.8	-23.459	-27.97	-105.18
	MAPE	847.64	437.78	217.92	334.46	243.78	281.03
	Theil's U	1.5239	1.5225	1.5204	1.037	0.99815	0.35336

Note: The table contains R^2 , ME , $RMSE$, MAE , MPE , $MAPE$ and Theil's U values for the in level regression 4.5 of the After-tax profit CAPE model with addition of bond yield (BY). Forecast valuation statistics are available for 6 forecasting horizons. Values arranged from worst (red) to best (green) performance.

6.5.2. After-tax profits CAPE

Out-of-sample forecast performance results for CAPE model, using the after-tax corporate profits, are presented in Tables 10 and 11. In general, results

for level and detrended data follow a different pattern except all statistics for 1-year horizon (excl. R^2). In comparison to generic CAPE, the model using NIPA corporate profits, shows substantially higher model fit for all horizons (except for 1-year).

Table 10: NIPA CAPE forecast evaluation statistics (Level)

		3m	6m	1yr	5yr	10yr	20yr
$CAPE_t$	R^2	0.081042	0.135459	-0.10546	0.042804	0.066846	0.023828
	ME	-0.00563	-0.00177	-0.02092	-0.00401	-0.00218	-0.00604
	RMSE	0.023614	0.03183	0.060606	0.039426	0.031058	0.034738
	MAE	0.022987	0.029036	0.043491	0.028169	0.021741	0.02527
	MPE	-82.804	43.611	29.319	-287.29	-149.37	-77.142
	MAPE	109.14	139.58	113.65	403.37	238.46	241.31
	Theil's U	0.8381	0.82823	1.4386	0.28599	0.28854	0.37683

Note: The table contains R^2 , ME , $RMSE$, MAE , MPE , $MAPE$ and Theil's U values for the in level regression 4.4 of the After-tax profit CAPE model. Forecast valuation statistics are available for 6 forecasting horizons. Values arranged from worst (red) to best (green) performance.

Taking the first difference significantly improves the model fit for 1, 5 and 10-year horizons with the highest value of 12% at the 5-year time span, but also increases ME , $MAPE$ and Theil's U for the respective time-horizons. Similarly to generic CAPE, $MAPE$ larger than 100% in all horizons in both level and difference with the largest value in a 5-year time span (3 months for generic model), shows that forecasting models are not able to explain the volatility of the out-of-sample data (Brooks, 2008).

Table 11: NIPA CAPE forecast evaluation statistics (FD)

		3m	6m	1yr	5yr	10yr	20yr
$CAPE_t^*$	R^2	-0.20802	-0.08502	0.095097	0.120259	0.105861	-0.03405
	ME	-0.00679	-0.00767	-0.01819	-0.00553	-0.00551	-0.01101
	RMSE	0.027074	0.035658	0.054833	0.037797	0.030402	0.035753
	MAE	0.024993	0.029795	0.040258	0.027831	0.022165	0.026483
	MPE	-100.74	52.349	43.041	-232.61	-131.83	-66.239
	MAPE	128.37	166.9	122.47	356.35	221.56	251.52
	Theil's U	0.7351	0.76003	1.3608	0.41218	0.41381	0.48902

Note: The table contains R^2 , ME , $RMSE$, MAE , MPE , $MAPE$ and Theil's U values for the in level regression 4.4 of the After-tax profit CAPE model. Forecast valuation statistics are available for 6 forecasting horizons. Values arranged from worst (red) to best (green) performance.

Addition of the bond yield increases R^2 in all horizons except 6-month in level, but decreases model fit for all horizons except for shortest time horizon in difference. It improves ME only for the longest time horizon in level and significantly worsens values in all horizons for the de-trended data. However, addition of the bond yield has almost no or slightly negative effect on $RMSE$ and MAE values. An adjusted model shows significantly lower $MAPE$ values for all horizons except for 1-year in level and for the short-horizons in difference.

Table 12: NIPA CAPE with BY forecast evaluation statistics (Level)

		3m	6m	1yr	5yr	10yr	20yr
$CAPE_t^*$ & BY_t	R^2	0.474374	-0.01678	-0.00965	0.086211	0.085223	0.038871
	ME	0.009386	-0.01764	-0.01799	0.001178	0.002251	-0.00379
	RMSE	0.02059	0.062119	0.058042	0.038577	0.031514	0.034478
	MAE	0.019908	0.047347	0.042783	0.02763	0.02236	0.025187
	MPE	2.8028	99.77	8.6566	-227.41	-115.3	-55.077
	MAPE	45.736	121.24	130.68	343.87	207.71	222.31
	Theil's U	0.6721	1.4497	1.3966	0.35037	0.35259	0.40419

Note: The table contains R^2 , ME , $RMSE$, MAE , MPE , $MAPE$ and Theil's U values for the in level regression 4.6 of the After-tax profit CAPE model with addition of bond yield (BY). Forecast valuation statistics are available for 6 forecasting horizons. Values arranged from worst (red) to best (green) performance.

Table 13: NIPA CAPE with BY forecast evaluation statistics (FD)

		3m	6m	1yr	5yr	10yr	20yr
$CAPE_t^*$ & BY_t	R^2	0.456371	-0.12058	-0.10929	0.034737	0.058982	-0.08135
	ME	0.009591	-0.02048	-0.01965	-0.00367	-0.00353	-0.01055
	RMSE	0.020939	0.065212	0.060839	0.039649	0.031963	0.036571
	MAE	0.019973	0.048809	0.044313	0.028552	0.022559	0.026702
	MPE	3.1537	107.67	13.387	-294.66	-159.13	-97.908
	MAPE	45.515	128.85	134.85	414.22	248.58	277.19
	Theil's U	0.70613	1.5369	1.4776	0.24784	0.25174	0.38717

Note: The table contains R^2 , ME , $RMSE$, MAE , MPE , $MAPE$ and Theil's U values for the in level regression 4.6 of the After-tax profit CAPE model with addition of bond yield (BY). Forecast valuation statistics are available for 6 forecasting horizons. Values arranged from worst (red) to best (green) performance.

6.5.3. Generic Fed model

Tables 14 and 15 contain the forecast statistics for generic Fed model. Generic Fed model regressions 4.7 - 4.9 in level exhibit R^2 of 22% in 1-year

horizon. Taking first differences for same horizon has negative effect on R^2 for regression 4.9, yielding 13%. Regressions 4.7 and 4.9 show R^2 of 0.0365 and 0.019 in 5-year horizon, respectively. Very low R^2 of -26% for 3-month and -60% for 6-month horizons are in accordance with Asness (2003) and Fama and French (1988), as "short-term market timing is hard". However, negative R^2 for 10- and 20-year horizons are inconsistent with the findings of Asness. When it comes to first difference, regression 4.9 exhibits even lower R^2 s for all horizons.

Table 14: Fed model forecast evaluation statistics (Level)

		3m	6m	1yr	5yr	10yr	20yr
EY_t	R^2	-0.25996	-0.5995	0.217944	0.036515	-0.13013	-0.16834
	ME	-1.281	-1.9409	0.091616	-0.57981	-1.2476	-1.6014
	RMSE	2.8193	3.0332	2.62	3.0458	3.4291	3.921
	MAE	2.7157	2.6911	2.2756	2.1811	2.4646	2.9111
	MPE	-154.53	-301.7	-125.48	-36.198	-135.85	-101.84
	MAPE	175.42	312.85	159.32	114.4	305.96	286.26
	Theil's U	0.53251	1.7966	1.1071	0.81627	0.90385	0.87427
	$EY_t - BY_t$	R^2	-0.36595	-0.85173	0.221283	-0.05859	-0.17227
ME		-1.5735	2.2918	-0.32202	-1.0329	-1.4931	-1.2163
RMSE		2.9355	3.2636	2.6144	3.1926	3.4924	3.764
MAE		2.7787	2.8945	2.2131	2.2943	2.523	2.7972
MPE		-169.48	-332.56	-144.05	-47.851	-145.23	-95.851
MAPE		187.02	341.33	170.55	127.06	311.36	271.41
Theil's U		0.48819	1.9944	1.2075	0.83526	0.89151	0.84548
EY_t & BY_t		R^2	-0.41908	-0.74187	0.196333	0.019211	-0.25757
	ME	-1.5234	-2.1185	-0.02943	-0.71012	-1.5614	-1.9742
	RMSE	2.992	3.1653	2.6559	3.073	3.6172	4.107
	MAE	2.8553	2.7844	2.2895	2.2165	2.6174	3.0764
	MPE	-169.69	-316.58	-132.93	-40.834	-153.95	-117.7
	MAPE	189.08	326.27	165.18	118.27	337.52	313.84
	Theil's U	0.5201	1.8436	1.13	0.82189	0.99923	0.96991

Note: The table contains R^2 , ME, RMSE, MAE, MPE, MAPE and Theil's U values for the in level regressions 4.7 - 4.9 of generic Fed model. Forecast valuation statistics are available for 6 forecasting horizons. Values arranged from worst (red) to best (green) performance.

Both in level and first difference, lowest ME is shown by regressions 4.7 and 4.9 at 6-month horizon. The consistently lowest RMSE is roughly 2.6 at 1-year horizon. First difference yields an RMSE increase to 2.76. In level, the lowest average MAE is 2.26 observed at 1 year horizon. Regression 4.7 shows the lowest MAE and differencing has an adverse effect on MAE.

Lowest negative MPE of below -300 are observed at 6 month horizon. Differencing does improve MPE , lowering it to -377.44 for the same period. $MAPE$ is lowest at 5-year horizon both in level and differenced. Highest $MAPE$ is obtained from running all 6-month regressions.

Lowest Theil's U below 1 is observed at 3-month horizon. First difference does provide the lowest U for this horizon only, while increasing it marginally for all other horizons. Interestingly, highest U above 1 is observed at 6-month horizon.

Table 15: Fed model forecast evaluation statistics (FD)

		3m	6m	1yr	5yr	10yr	20yr
EY_t & BY_t	R^2	-0.72526	-1.32688	0.131188	-0.23194	-0.44169	-0.25126
	ME	-2.0323	-2.7193	-0.71755	-1.5914	-2.1642	-1.8115
	RMSE	3.2991	3.6584	2.7615	3.4441	3.873	4.0578
	MAE	3.0511	3.2287	2.2304	2.4975	2.8601	3.0564
	MPE	-197.48	-377.44	-168.02	-63.502	-180.54	-125.16
	MAPE	212.32	384.86	188.92	143.76	367.06	317.85
	Theil's U	0.47588	2.1892	1.3071	0.86512	1.0866	1.0397

Note: The table contains R^2 , ME , $RMSE$, MAE , MPE , $MAPE$ and Theil's U values for the in first difference regressions 4.7 - 4.9 of generic Fed model. Forecast valuation statistics are available for 6 forecasting horizons. Values arranged from worst (red) to best (green) performance.

6.5.4. Fed Model adjusted for perceived risk

Regression 4.11 used to obtain the fitted value, EY_t^f , takes the following form:

$$EY_t^f = 0.065 + 0.404BY_t + 0,4579\sigma_t^e - 1.897\sigma_t^b; \quad R^2 = 0.1419$$

(6.352)*** (4.401)*** (2.610)*** (-4.711)***

Signs of the coefficients match those obtained by Asness (2003). Based on the t-statistics presented in the parentheses under the regression, all respective coefficients are statistically significant at 0.1% level as can be seen. The relationship between EY_t^f and BY_t , as represented by the $\beta_1 = 0.404$, is not consistent with the level expected by the Fed model ($\beta_1 = 1$).

Tables 16 and 17 contain the forecast statistics for Fed model adjusted for perceived risk. Highest average R^2 of 0.1 for in-level regressions is observed for 1-year horizon.

Table 16: **Adjusted Fed model forecast evaluation statistics (Level)**

		3m	6m	1yr	5yr	10yr	20yr
EY_t	R^2	-0.26233	-0.00872	0.105908	-0.11205	-0.14729	0.020392
	ME	-0.05063	-0.01268	0.001937	-0.01715	-0.01842	-0.01146
	RMSE	0.073491	0.056124	0.044035	0.049248	0.044003	0.04064
	MAE	0.050627	0.037947	0.032047	0.037116	0.033489	0.030573
	MPE	26.191	28.598	4.9276	-41.546	-60.523	-128.45
	MAPE	78.161	54.582	62.724	312.56	256.53	299.29
	Theil's U	0.28103	0.55405	0.73329	0.66871	0.72217	0.24273
EY_t^f	R^2	-0.22159	-0.00612	0.100814	-0.09106	-0.07892	0.059466
	ME	-0.05921	-0.02044	-0.00739	-0.02165	-0.01614	-0.00929
	RMSE	0.080146	0.059344	0.044887	0.051308	0.04381	0.041031
	MAE	0.059206	0.039015	0.030008	0.039772	0.033423	0.030722
	MPE	14.139	17.504	-16.304	-87.777	-69.305	-142.09
	MAPE	98.158	59.973	69.209	359.75	266.49	310.46
	Theil's U	0.46502	0.55258	0.65435	0.79958	0.80047	0.21989
EY_t^e	R^2	-0.22159	-0.00612	0.100814	-0.09106	-0.07892	0.059466
	ME	-0.04815	-0.00991	0.003377	-0.01554	-0.01474	-0.00917
	RMSE	0.072296	0.056052	0.04416	0.048781	0.042672	0.039822
	MAE	0.048145	0.038237	0.032241	0.037105	0.03232	0.029758
	MPE	31.033	33.223	6.8411	-44.189	-53.938	-130.04
	MAPE	71.997	53.705	62.734	309.59	245.42	296.05
	Theil's U	0.1954	0.57888	0.75216	0.65835	0.68974	0.21133

Note: The table contains R^2 , ME , $RMSE$, MAE , MPE , $MAPE$ and Theil's U values for the in level regressions 4.13 - 4.15 of generic Fed model. Forecast valuation statistics are available for 6 forecasting horizons. Values arranged from worst (red) to best (green) performance.

Average R^2 for regressions that used detrended data is 8.6% and is also noted at 1 year horizon. While in-level, regression 4.15 with error term has a lower power, than 4.13 and 4.14, the 4.15 with FD has highest positive R^2 both for 1- and 20-year horizons. Positive R^2 values observed for 20-year horizon are now consistent with the findings of Asness (2003) and Salomons (2004). All other forecast periods exhibit negative R^2 , meaning that the regression fits worse than a horizontal line.

Both in level and first difference, 3-year horizon ME is the lowest while the highest is for 6-year horizon. $RMSE$ is lowest for 20-year horizon and highest for 3-month horizon. First differences do not consistently improve $RMSE$, as only 4.14 has a drop in average $RMSE$ compared to in level (0.05342 vs 0.0542). MAE around 0.03 at its lowest for 20-year horizon for both level and first difference.

Table 17: **Adjusted Fed model forecast evaluation statistics (FD)**

		3m	6m	1yr	5yr	10yr	20yr
EY_t	R^2	-0.39688	-0.08576	0.081612	-0.20952	-0.11256	0.005974
	ME	-0.05734	-0.01697	-0.0034	-0.01982	-0.01393	-0.00775
	RMSE	0.077309	0.058228	0.044629	0.051361	0.043332	0.040938
	MAE	0.057344	0.040376	0.031575	0.040027	0.032921	0.030558
	MPE	13.302	20.857	-9.1541	-83.845	-62.582	-139.27
	MAPE	94.978	61.695	69.104	357.21	259.56	305.87
	Theil's U	0.4894	0.61703	0.71032	0.81376	0.80393	0.21077
EY_t^f	R^2	-0.40072	-0.08779	0.08804	-0.17185	-0.09935	0.016358
	ME	-0.05558	-0.01649	-0.00401	-0.01939	-0.01353	-0.00772
	RMSE	0.077415	0.058283	0.044473	0.050555	0.043074	0.040724
	MAE	0.05558	0.039091	0.030623	0.039352	0.032878	0.030429
	MPE	19.436	23.168	-9.8049	-81.103	-62.065	-141.71
	MAPE	89.697	58.298	67.54	348.83	256.16	306.31
	Theil's U	0.37593	0.56462	0.68144	0.76495	0.76206	0.20945
EY_t^e	R^2	-0.40355	-0.08569	0.088871	-0.17337	-0.10186	0.017342
	ME	-0.05528	-0.0164	-0.00415	-0.01941	-0.01351	-0.00775
	RMSE	0.077493	0.058226	0.044453	0.050588	0.043123	0.040704
	MAE	0.055282	0.038882	0.030552	0.039311	0.032867	0.030378
	MPE	20.571	23.752	-10.184	-80.977	-62.103	-140.96
	MAPE	88.78	57.856	67.792	348.71	256.35	306.65
	Theil's U	0.3545	0.54701	0.67716	0.76607	0.76288	0.20623

Note: The table contains R^2 , ME , $RMSE$, MAE , MPE , $MAPE$ and Theil's U values for the in first difference (FD) regressions 4.13 - 4.15 of Fed model adjusted for perceived risk. Forecast valuation statistics are available for 6 forecasting horizons. Values arranged from worst (red) to best (green) performance.

MPE lowest for 20 yr both level and first difference but improves significantly for 4.13 and 4.15: approximately 2-fold reduction. 3- and 6-month MPE are highest. $MAPE$ values for the periods from 3 months to 6 months show results lower than 100%, implying that adjusted Fed can explain volatility in the out-of-sample data. High levels observed for long-horizons, reaching maximum at 5 years for all equations, both level and FD.

Theil's U is better than guessing if it is lower than 1. Values are consistently below 1, and are the lowest for 20 year horizon, closely followed by 3m horizon. Minimal Theil's U obtained in level for 3m horizon

6.6. Testing equality of forecast accuracy

6.6.1. MGN test of forecast Accuracy

MGN tests have been conducted, for three pairs of models forecast models: 1) generic CAPE (eq. 4.3) versus after-tax profit CAPE (eq. 4.4); 2) generic Fed (eq. 4.8) vs the version adjusted for perceived risk using error term (eq. 4.15); 3) generic CAPE vs generic Fed. Results presented in Table 18 indicate that we fail to reject the null hypothesis of forecast accuracy equivalence most of the time.

Table 18: MGN test results (level)

		3m	6m	1yr	5yr	10yr	20yr	Full Sample
$CAPE_t$ vs Generic FED	t-stat	-0.065	-0.290	1.233	1.437	0.803	-3.249**	-4.805***
	p-value	0.477	0.392	0.878	0.922	0.788	0.001***	0.000***
$CAPE_t$ vs $CAPE_t^*$	t-stat	0.378	-0.009	0.682	0.493	1.036	-1.002	-2.297*
	p-value	0.629	0.497	0.745	0.688	0.849	0.159	0.011*
Generic Fed vs Adjusted Fed	t-stat	0.307	0.125	1.237	1.609	2.544	2.011	-1.911*
	p-value	0.606	0.544	0.879	0.944	0.994	0.977	0.028*

Note: The table contains *t*-statistics and *p*-values for MGN test are presented with a significance level, denoted by (*), (**) and (***) for 5%, 1% and 0.1%, respectively. The values are available for 6 forecasting horizons and the full sample 1987m3 - 2016m11 ($T = 357$).

Specifically, according to *p*-values of set of generic CAPE and Fed models in level, rejection takes place at 20 year horizon and at full sample at 0.1% significance level. Based on the analysis of forecast statistics conducted in subsection 6.4, a conclusion can be drawn that indeed there is a statistical difference in forecast accuracy between the two model and that it is in favour of the modified CAPE (NIPA).

Other two pairs only reject the null at 5% significance at full sample of 357 observations, implying that for finite samples, forecasting horizons between 3 months and 20 years do not indicate existence of significant forecast accuracy differential between model sets 1) and 2). Hence, suggested improvements to generic CAPE and Fed are not recognized unless sample is large enough.

6.6.2. DM test of forecast Accuracy

Forecasts, obtained using same models used in MGN analysis, have been used and tested for equality of forecast precision using Diebold-Mariano test. Results are presented in Table 19.

Table 19: DM test results (level)

	3m	6m	1yr	5yr	10yr	20yr	Full Sample
$CAPE_t$ vs Gen FED	-0.038	-0.132	0.328	0.129	0.043	-0.203	-0.282
$CAPE_t$ vs $CAPE_t^*$	0.221	-0.004	0.305	0.071	0.067	-0.061	-0.134
Gen FED vs Adj FED	-0.398	0.283	0.341	0.142	0.138	0.109	-0.093

Note: The table contains t -statistics for DM, which are presented with a significance level, denoted by (*), (**) and (***) for 5%, 1% and 0.1%, respectively. The values are available for 6 forecasting horizons and the full sample 1987m3 - 2016m11 ($T = 357$).

When it comes to determining the best forecast, the results appear to be mixed. For instance, the sign (negative) of DM test statistic indicates that the forecast generated by the generic CAPE model for 3- to 6-month horizons is more accurate than that of the generic Fed, which is not in line with the theory. The same unexpected result persists through 1-, 5- and 10-year horizons where the generic Fed appears to yield a more accurate forecast. At longer 20-year and full sample, forecasts made by generic CAPE outperform those of generic Fed. In the comparison between the traditional and after-tax corporate profits CAPE, the latter performs best at 3-month, 1-year, 5-year and 10-year horizons. For the pair of two alternative Fed models, the generic version dominates the modified at 3-month and full sample horizons.

However, the results are not indicative of significant forecast difference as the DM-statistics for all pseudo-out-of-sample sub-horizons as well as the full sample are not statistically significant at 5%, 1% and 0.1% significance levels. This leads to non-rejection of null hypothesis and a conclusion that the pairs of models provide expected equity return forecasts of equal precision. These findings are consistent with arguments presented by Diebold (2015), who posits that statistical comparison of pseudo-out-of-sample forecast models is of little value due to reduced power stemming from finite samples and vulnerability to data-mining. Overall, full-sample model comparisons are preferable.

7. Conclusion

Historic data indicates that returns on equity investments are subject to considerable short-and medium-run fluctuations but, over long horizons, equity holders benefit from a safer, high yielding and liquid investment alternative. Therefore, accurately forecasting expected stock returns is crucial for making strategic investment decisions.

In-sample, all CAPE model versions dominate the traditional Fed and its alternatives in-level and in first difference. Detrending of data causes significant loss of explanatory power, putting Fed model at and even greater disadvantage.

Our findings out-of-sample give mixed results. According to them, generic Fed model can be seen as a useful tool for generating out-of-sample forecasts for horizons no shorter than 1 year. Taking perceived risk into account does improve the predictive ability of Fed model marginally judging from some forecast statistics (*MAPE* less than 100%), while others indicate that addition has negative effects. We confirm the findings by Asness (2003) and Salomons (2004) that the adjusted version of Fed model can be employed for forecasting of expected stock returns at 20 year horizon.

On the other hand, generic CAPE model appears to be better suited for forecasting returns in 1 year and 20-year horizon. The altered version of CAPE, which uses after-tax profits data seems to improve generic CAPE forecast performance, as it shows better results for 3m - 5yr horizons.

Based on the aforementioned analysis, a conclusion can be drawn that no model is superior all around. We agree with Clemens (2007) on that the two traditional models can be combined to complement each other in providing semi-accurate forecasts of expected stock returns for different investment horizons instead. Furthermore, we agree with the points presented by Vanguard Forecasting and Shiller, whereby investors should not focus on forecasting expected stock returns for single points in time but take into account a larger time-series distribution, which would give a more realistic insight into the possible future outcomes. This should provide a broader

picture when constructing an assets allocation strategy. Irrespective of forecast model's performance, market tends to be unpredictable most of the time. Therefore, diversification should be employed.

Models and methods employed in our analysis were chosen deliberately as reasonably simple. Further improvements could be achieved by imposing theoretical constraints onto CAPE and Fed models as well as by employing quantitative dynamic version of Fed model proposed by (Koivu et al., 2005). We recognize the limitation of selected data samples from US market. Compelling results could be obtained with tests conducted for different in-sample-to-out-of-sample ratios. Research could be extended further by examining the model performance in the setting of other developed markets of MSCI World Index. Finally, a more in-depth comparison could be executed for specific periods of market activity, to see if models are capable of pointing at upcoming drastic shifts in the stock market.

Bibliography

- Armstrong, J. S. and Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International journal of forecasting*, 8(1):69–80.
- Arnott, R. D. and Asness, C. S. (2003). Surprise! higher dividends= higher earnings growth. *Financial Analysts Journal*, 59(1):70–87.
- Arnott, R. D. and Bernstein, P. L. (2002). What risk premium is “normal”? *Financial Analysts Journal*, 58(2):64–85.
- Asness, C. (2003). Fight the fed model. *Journal of Portfolio Management*, 30(1):11.
- Basu, S. (1977). Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. *The journal of Finance*, 32(3):663–682.
- Basu, S. (1983). The relationship between earnings’ yield, market value and return for nyse common stocks: Further evidence. *Journal of financial economics*, 12(1):129–156.
- Bekaert, G. and Engstrom, E. (2010). Inflation and the stock market: Understanding the “fed model”. *Journal of Monetary Economics*, 57(3):278–294.
- Blodget, H. and Kava, L. (2013). Shiller: Stocks are priced for (relatively) crappy returns. retrieved from. <http://www.businessinsider.com/robert-shiller-on-stocks-2013-1?r=US&IR=T&IR=>.
- Brooks, C. (2008). Introductory econometrics for finance. *Cambridge, Cambridge University*.
- Campbell, J. Y. and Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *The Review of Financial Studies*, 1(3):195–228.
- Campbell, J. Y. and Thompson, S. B. (2007). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, 21(4):1509–1531.
- Clemens, M. (2007). A behavioral defense of the fed model. *Available at SSRN 958800*.
- Cochrane, J. H. (2007). The dog that did not bark: A defense of return

- predictability. *The Review of Financial Studies*, 21(4):1533–1575.
- Da, Z., Jagannathan, R., and Shen, J. (2014). Growth expectations, dividend yields, and future stock returns. Technical report, National Bureau of Economic Research.
- Davis, J., Aliaga-Díaz, R., and Thomas, C. J. (2012). Forecasting stock returns: What signals matter, and what do they say now. *Valley Forge, Pa.: The Vanguard Group*.
- De Bondt, W. F. and Thaler, R. H. (1990). Do security analysts overreact? *The American Economic Review*, pages 52–57.
- Dempsey, M. (2010). The book-to-market equity ratio as a proxy for risk: evidence from australian markets. *Australian Journal of Management*, 35(1):7–21.
- Diebold, F. X. (2015). Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold–mariano tests. *Journal of Business & Economic Statistics*, 33(1):1–1.
- Diebold, F. X. and Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & economic statistics*, 20(1):134–144.
- Dimitrov, V. and Jain, P. C. (2016). Shiller’s cape: Market timing and risk.
- Estrada, J. (2009). The fed model: The bad, the worse, and the ugly. *The Quarterly Review of Economics and Finance*, 49(2):214–238.
- Faber, M. T. (2012). Global value: Building trading models with the 10 year cape. *Cambria Quantitative Research*, (5).
- Fama, E. F. and French, K. R. (1988). Permanent and temporary components of stock prices. *Journal of political Economy*, 96(2):246–273.
- Fama, E. F. and French, K. R. (1989). Business conditions and expected returns on stocks and bonds. *Journal of financial economics*, 25(1):23–49.
- Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. *the Journal of Finance*, 47(2):427–465.
- Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56.
- Glassman, J. K. and Hassett, K. A. (1999). *Dow 36,000: The new strategy for profiting from the coming rise in the stock market*. Times Business New York.
- Graham, B. and Dodd, D. L. (1934). *Security analysis: principles and*

technique. McGraw-Hill.

- Greenspan, A. (1997). Federal reserve board humphrey hawkins testimony.
- Hall, R. E. (2000). E-capital: The link between the stock market and the labor market in the 1990s. *Brookings Papers on Economic Activity*, 2000(2):73–102.
- Hodrick, R. J. (1992). Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *The Review of Financial Studies*, 5(3):357–386.
- Hyndman, R. J. and Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4):679–688.
- Inoue, A. and Kilian, L. (2005). In-sample or out-of-sample tests of predictability: Which one should we use? *Econometric Reviews*, 23(4):371–402.
- Kantor, B. and Holdsworth, C. (2014). 2013 nobel prize revisited: Do shiller’s models really have predictive power? *Journal of Applied Corporate Finance*, 26(2):101–108.
- Keimling, N. (2014). Cape: Predicting stock market returns. *StarCapital Research Publication, February*.
- Kilian, L. and Taylor, M. P. (2003). Why is it so difficult to beat the random walk forecast of exchange rates? *Journal of International Economics*, 60(1):85–107.
- Koivu, M., Pennanen, T., and Ziemba, W. T. (2005). Cointegration analysis of the fed model. *Finance Research Letters*, 2(4):248–259.
- Lamont, O. (1998). Earnings and expected returns. *The journal of Finance*, 53(5):1563–1587.
- Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics*, 74(2):209–235.
- Modigliani, F. and Cohn, R. A. (1979). Inflation, rational valuation and the market. *Financial Analysts Journal*, 35(2):24–44.
- Orr, J. M., Sackett, P. R., and Dubois, C. L. (1991). Outlier detection and treatment in i/o psychology: A survey of researcher beliefs and an empirical illustration. *Personnel Psychology*, 44(3):473–486.
- Pontiff, J. and Schall, L. D. (1998). Book-to-market ratios as predictors of market returns. *Journal of Financial Economics*, 49(2):141–160.

- Rasmussen, J. L. (1988). Evaluating outlier identification tests: Mahalanobis d squared and comrey dk. *Multivariate Behavioral Research*, 23(2):189–202.
- Ritter, J. R. and Warr, R. S. (2002). The decline of inflation and the bull market of 1982–1999. *Journal of financial and quantitative analysis*, 37(01):29–61.
- Salomons, R. (2004). A tactical implication of predictability: Fighting the fed model. *Available at SSRN 517322*.
- Sharpe, W. F. (1970). *Portfolio theory and capital markets*. McGraw-Hill College.
- Shiller, R. J. (2014). The mystery of lofty stock market elevations. *The New York Times*.
- Siegel, J. J. (2016). The shiller cape ratio: A new look. *Financial Analysts Journal*, 72(3):41–50.
- Siegel, J. J. and Coxe, D. G. (2002). *Stocks for the long run*, volume 3. McGraw-Hill New York.
- Stock, J. H. and Watson, M. W. (2007). Introduction to econometrics.
- Thomas, J. and Zhang, F. (2008). Don't fight the fed model. *Unpublished paper, Yale University, School of Management*.
- Welch, I. and Goyal, A. (2007). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 21(4):1455–1508.
- Yardeni, E. (1997). Fed's stock market model finds overvaluation. *Topical study*, 38.
- Zimmerman, D. W. (1994). A note on the influence of outliers on parametric and nonparametric tests. *The journal of general psychology*, 121(4):391–401.

Appendices

Appendix 1 - OLS Framework

(a) Classical Linear Regression Model (CLRM) assumptions

1. $E(v_t) = 0$ (error terms have zero mean)
2. $Var(v_t) = \sigma^2$ (variance of the errors is constant and finite over all Y_{t+1})
3. $Cov(v_t, v_j) = 0$ (error terms are linearly independent from one another);
4. $Cov(v_t, X_t) = 0$ (error terms and corresponding X_t are uncorrelated);
5. $v_t \sim N(0, \sigma^2)$ (error terms are normally distributed)
6. No perfect multicollinearity

(b) Regression coefficient estimates

$$\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X} \quad (7.1)$$

$$\hat{\beta} = \frac{\sum (X_t - \bar{X}) Y_t}{\sum (X_t - \bar{X})^2} \quad (7.2)$$

(c) Goodness of Fit (R^2) of the regression model

$$R^2 = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{t=1}^T (Y_t - \hat{Y}_t)^2}{\sum_{t=1}^T (Y_t - \bar{Y})^2} \quad (7.3)$$

Where Y_t - observed stock return observation at time t , \bar{Y} - mean stock return, \hat{Y}_t - stock return forecast at time t .

Appendix 2 - Out-of-sample forecast evaluation statistics

$$\text{ME} = \frac{1}{T - (T_1 - 1)} \sum_{t=T_1}^T (Y_t - \hat{Y}_t)$$

$$\text{MAE} = \frac{1}{T - (T_1 - 1)} \sum_{t=T_1}^T |Y_t - \hat{Y}_t|$$

$$\text{RMSE} = \sqrt{\frac{1}{T - (T_1 - 1)} \sum_{t=T_1}^T (Y_t - \hat{Y}_t)^2}$$

$$\text{MPE} = \frac{100}{T - (T_1 - 1)} \sum_{t=T_1}^T \left(\frac{Y_t - \hat{Y}_t}{Y_t + \hat{Y}_t} \right)$$

$$\text{MAPE} = \frac{100}{T - (T_1 - 1)} \sum_{t=T_1}^T \left| \frac{Y_t - \hat{Y}_t}{Y_t + \hat{Y}_t} \right|$$

$$\text{Theil's U (U)} = \frac{\sqrt{\sum_{t=T_1}^T \left(\frac{Y_t - \hat{Y}_t}{Y_t} \right)^2}}{\sqrt{\sum_{t=T_1}^T \left(\frac{Y_t - \widehat{Y}_t^{RW}}{Y_t} \right)^2}}$$

T - total sample size;

T_1 - first out-of-sample forecast observation hold-out sample;

$(T - (T_1 - 1))$ - hold-out sample;

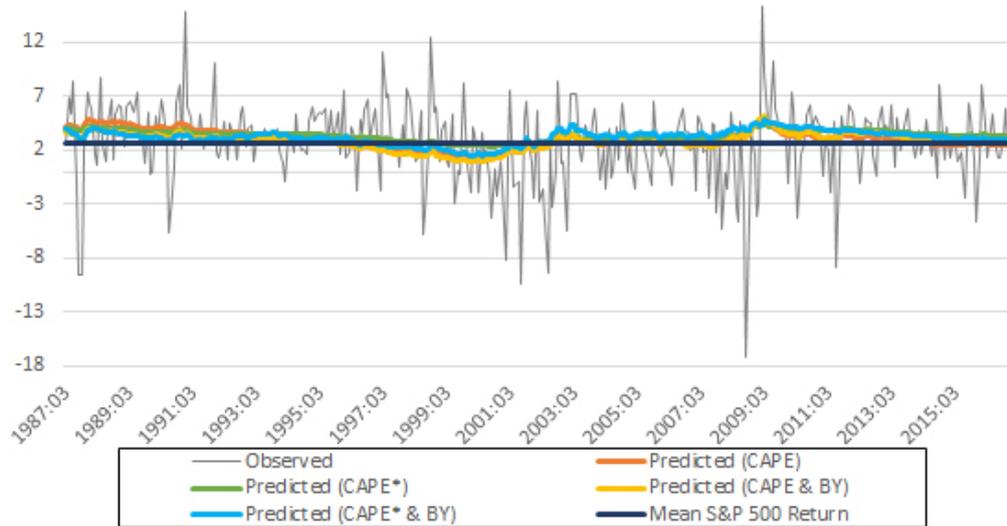
Y_t - actually observed value at time t ;

\hat{Y}_t - forecast value for time t obtained using a predictive model;

\widehat{Y}_t^{RW} is the forecast obtained from random walk.

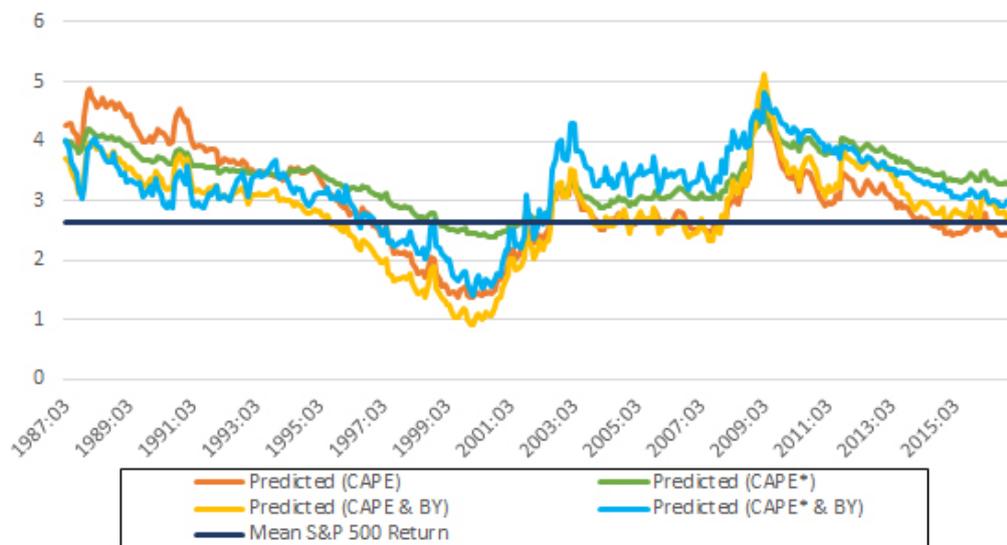
Appendix 3 - Visual analysis of forecast time series

Figure A.1: CAPE Predictions vs S&P 500 Return



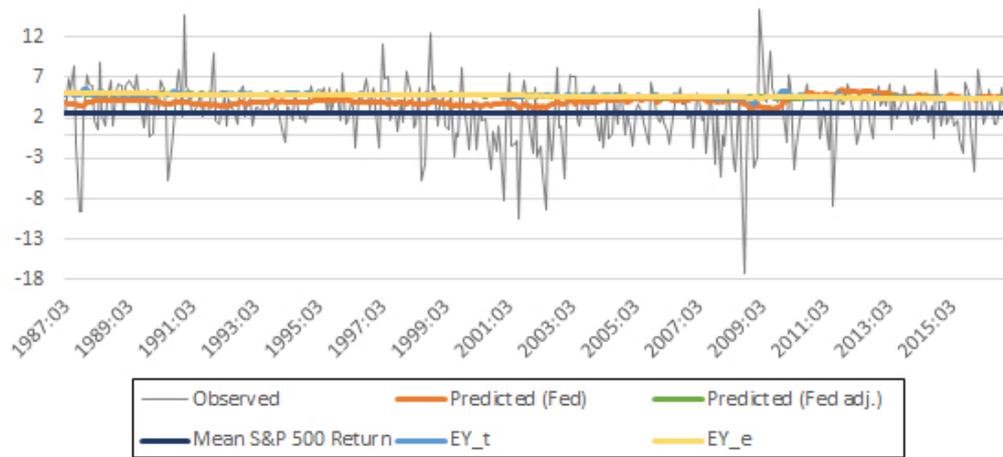
Note: The figure plots the time series of observed S&P 500 returns with its the long-run mean against the expected equity return forecasts from the generic CAPE, after-tax profit (NIPA) CAPE (CAPE*), generic CAPE with bond yield (CAPE&BY) and NIPA CAPE with bond yield (CAPE*&BY) for the 1987:03 - 2016:11 period.

Figure A.2: CAPE vs Alternatives



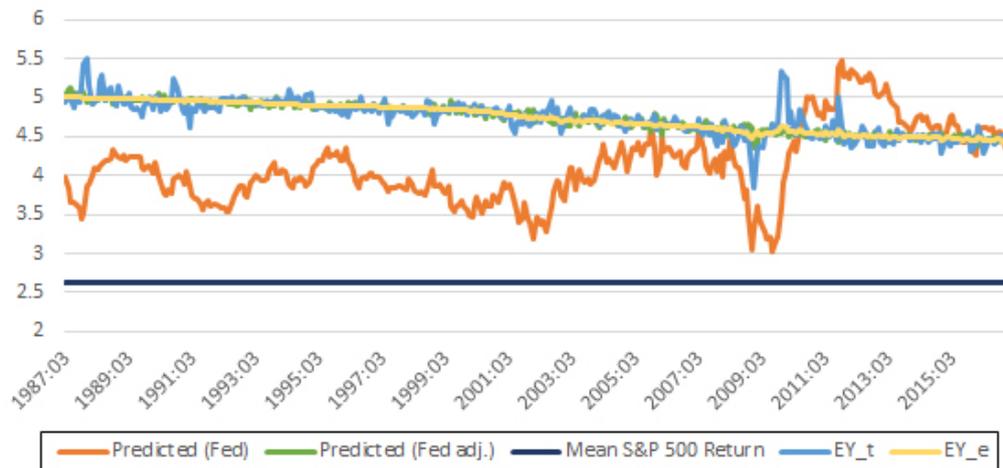
Note: The figure zooms in on the plots of the long-run mean return of S&P 500 and the time series of expected equity return forecasts from the generic CAPE, after-tax profit (NIPA) CAPE (CAPE*), generic CAPE with bond yield (CAPE&BY) and NIPA CAPE with bond yield (CAPE*&BY) for the 1987:03 - 2016:11 period.

Figure A.3: Fed Predictions vs S&P 500 Return



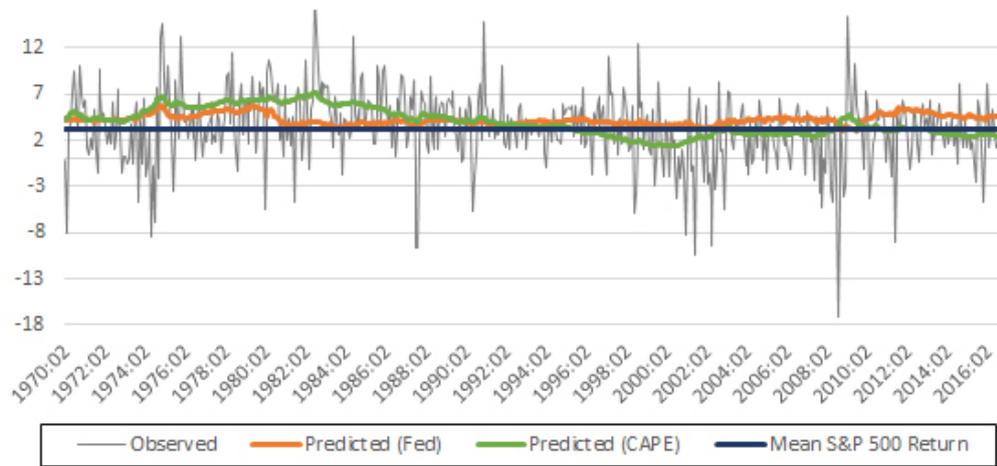
Note: The figure plots the time series of observed S&P 500 returns with its the long-run mean against the expected equity return forecasts from the generic Fed model and its risk-adjusted version, including the yield (EY_t) from regression 4.13 and error term (EY_e) from regression 4.15, for the 1987:03 - 2016:11 period.

Figure A.4: Fed vs Alternatives



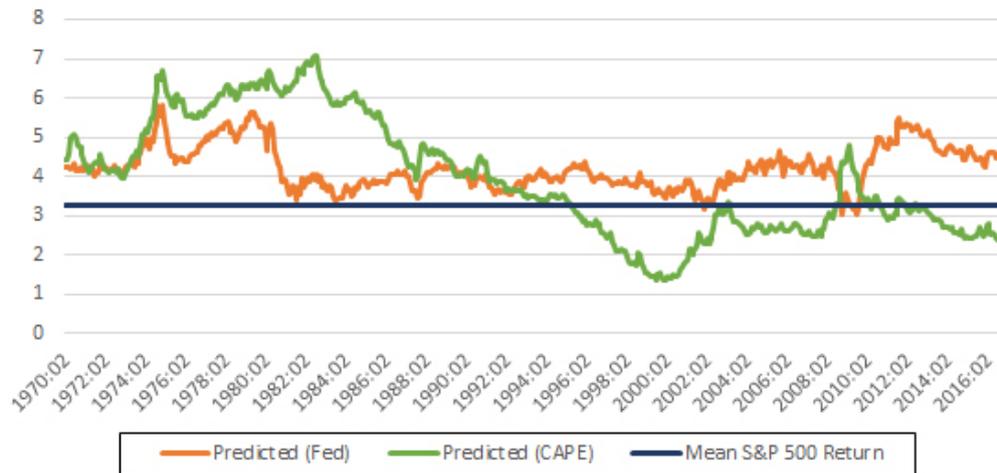
Note: The figure zooms in on the plots of the long-run mean return of S&P 500 and the time series of expected equity return forecasts from the generic Fed model and its risk-adjusted version, including the yield (EY_t) from regression 4.13 and error term (EY_e) from regression 4.15, for the 1987:03 - 2016:11 period.

Figure A.5: CAPE and Fed Predictions vs S&P 500 Return



Note: The figure plots the time series of observed S&P 500 returns with its the long-run mean against the expected equity return forecasts from the generic CAPE and Fed models for the 1970:02 - 2016:11 period.

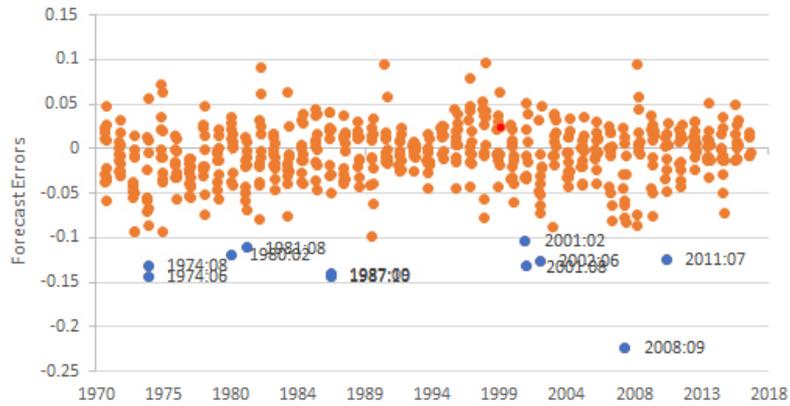
Figure A.6: CAPE vs Fed



Note: The figure zooms in on the plots of the long-run mean return of S&P 500 and the time series of expected equity return forecasts from the generic CAPE and Fed models for the 1970:02 - 2016:11 period.

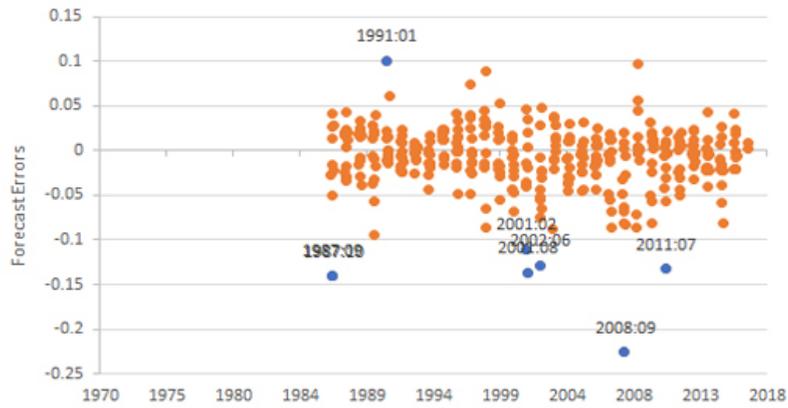
Appendix 4 - Plots of Forecast error outlier time series

Figure A.7: Time series of CAPE forecast errors



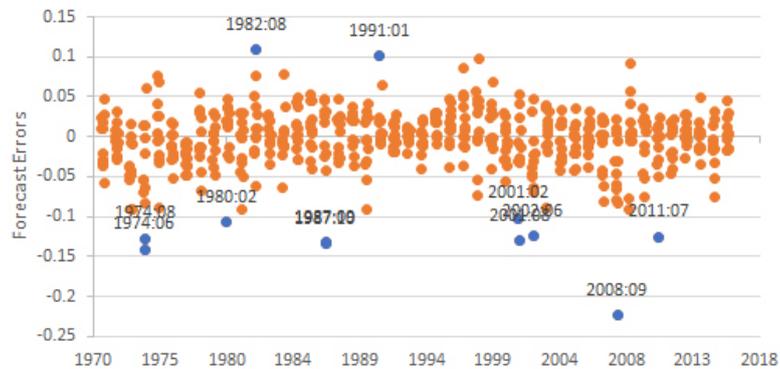
Note: The figure plots the forecast errors obtained by running regression 4.3 of generic CAPE model for the 1970:02 - 2017:05 period. Outliers are printed in blue with date level on the right.

Figure A.8: Time series of NIPA CAPE forecast errors



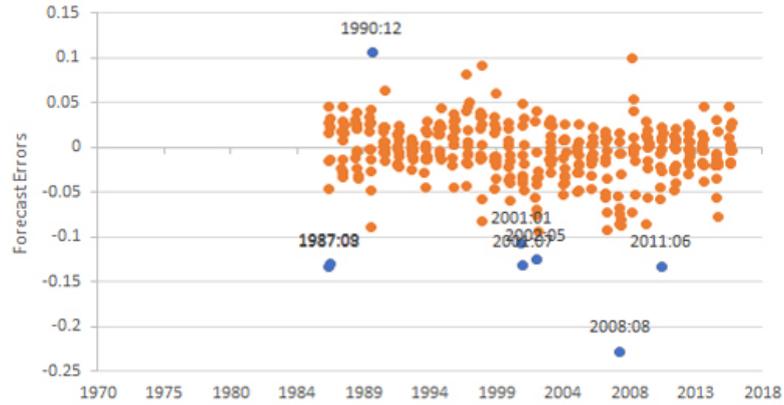
Note: The figure plots the forecast errors obtained by running regression 4.4 of after-tax profits (NIPA) CAPE model for the 1987:03 - 2017:02 period. Outliers are printed in blue with date level on top.

Figure A.9: Time series of CAPE with BY forecast errors



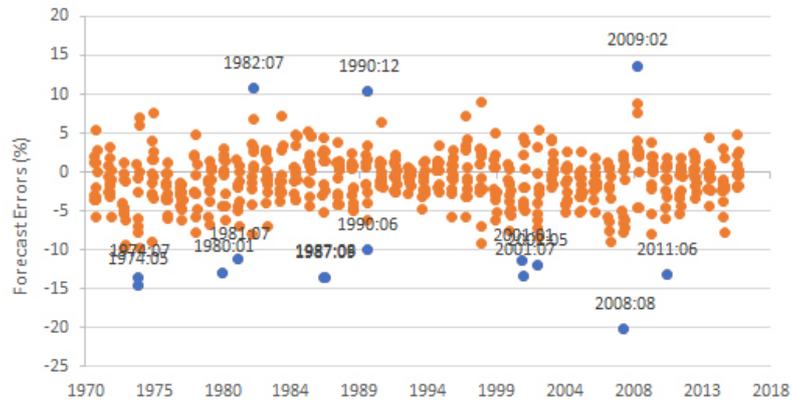
Note: The figure plots the forecast errors obtained by running regression 4.5 of generic CAPE, with $(CAPE_t + BY_t)$ as regressor for the 1970:02 - 2016:12 period. Outliers are printed in blue with date level on top.

Figure A.10: **Time series of NIPA CAPE with BY forecast errors**



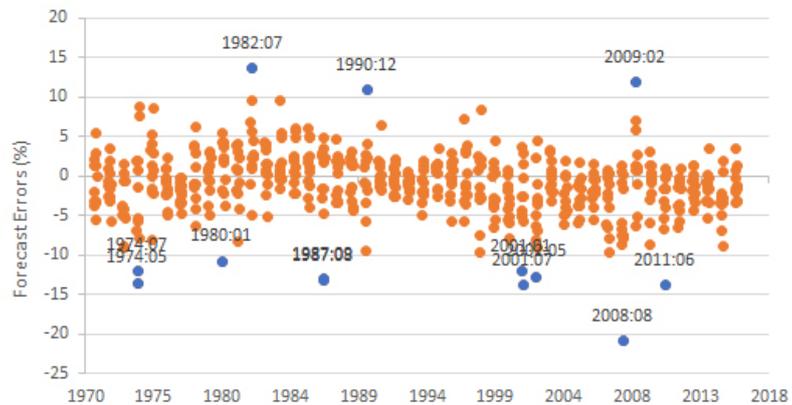
Note: The figure plots the forecast errors obtained by running regression 4.6 of after-tax profits (NIPA) CAPE model with BY_t as the second regressor for the 1987:03 - 2017:02 period. Outliers are printed in blue with date level on top.

Figure A.11: **Time series of Fed model forecast errors (EY_t)**



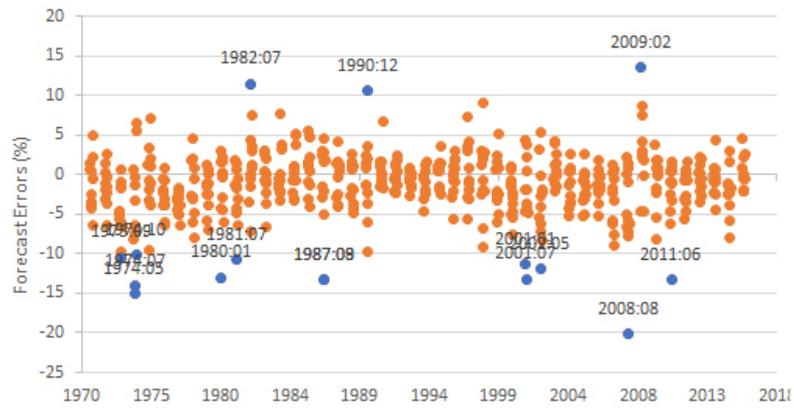
Note: The figure plots the forecast errors obtained by running regression 4.7 of generic Fed model with regressor EY_t for the 1960:03 - 2016:11 period. Outliers are printed in blue with date level on top.

Figure A.12: **Time series of Fed model forecast errors ($EY_t - BY_t$)**



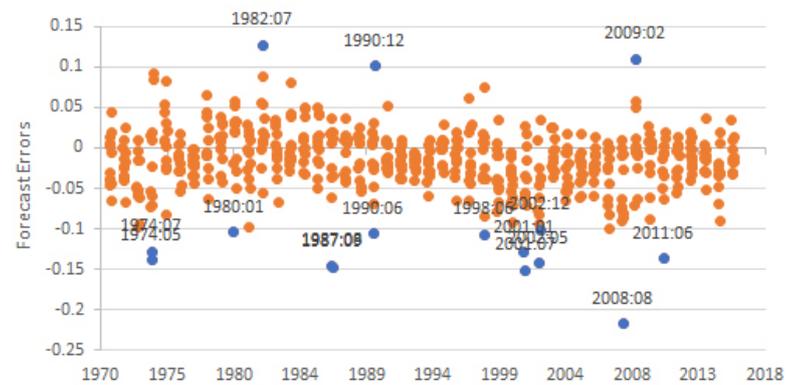
Note: The figure plots the forecast errors obtained by running regression 4.8 of generic Fed model with regressors, $EY_t - BY_t$, for the 1960:03 - 2016:11 period. Outliers are printed in blue with date level on top.

Figure A.13: **Time series of Fed model forecast errors (EY_t & BY_t)**



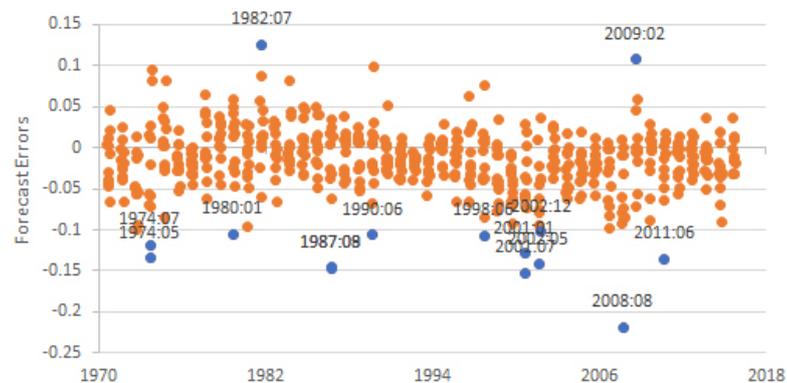
Note: The figure plots the forecast errors obtained by running regression 4.9 of generic Fed model with two regressors, EY_t and BY_t , for the 1960:03 - 2016:11 period. Outliers are printed in blue with date level on top.

Figure A.14: **Time series of Adjusted Fed model forecast errors (EY_t)**



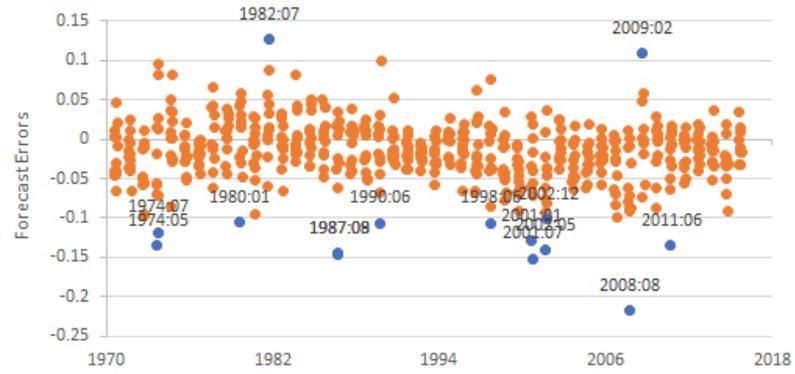
The figure plots the forecast errors obtained by running regression 4.13 of Fed model adjusted for perceived risk using the earnings yield (EY_t) data for the 1960:03 - 2016:11 period. Outliers are printed in blue with date level on top.

Figure A.15: **Time series of Adjusted Fed model forecast errors (EY_t^f)**



Note: The figure plots the forecast errors obtained by running regression 4.14 of Fed model adjusted for perceived risk using the fitted term (EY_t^f) data for the 1960:03 - 2016:11 period. Outliers are printed in blue with date level on top.

Figure A.16: Time series of Adjusted Fed model forecast errors (EY_t^e)



Note: The figure shows the scatter plot of forecast errors obtained by running regression 4.15 of Fed model adjusted for perceived risk using the error term (EY_t^e) data for the 1960:03 - 2016:11 period. Outliers are printed in blue with date level on top.

Appendix 5 - Preliminary thesis report

BI Norwegian Business School - campus Oslo

GRA 19502

Master Thesis

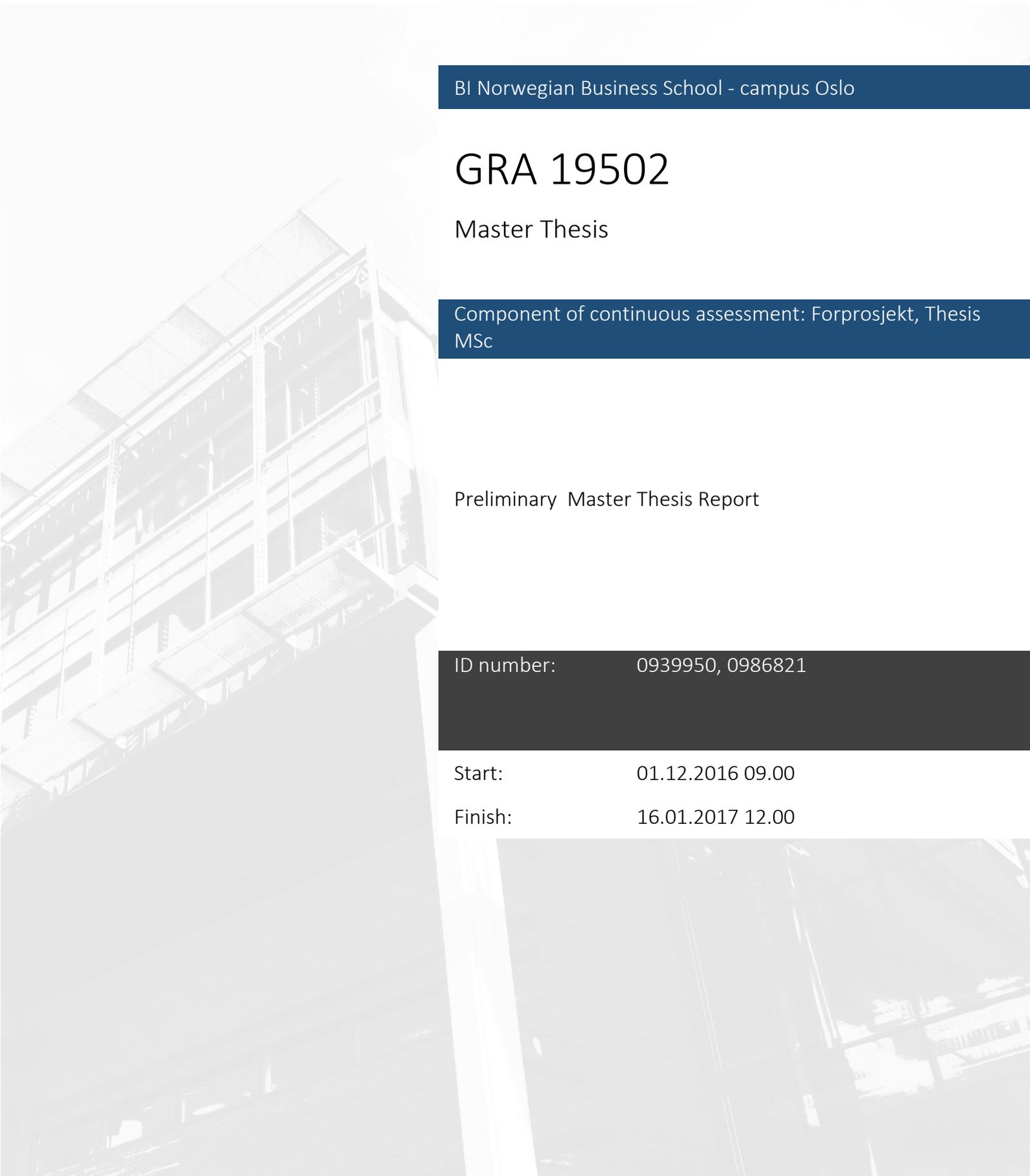
Component of continuous assessment: Forprosjekt, Thesis
MSc

Preliminary Master Thesis Report

ID number: 0939950, 0986821

Start: 01.12.2016 09.00

Finish: 16.01.2017 12.00



Preliminary Thesis Report

CAPE vs the Fed Model:

**Comparative analysis of the out-of-sample
performance in predicting future stock returns**

Supervisor:

Costas Xiouros

Campus:

BI Oslo

Examination code and name:

GRA19502 Master Thesis

Program:

Master of Science in Business with Major in Finance

Abstract

The goal of this thesis is to compare the out-of-sample performance in predicting future stock returns of the Fed model and Shiller's cyclically adjusted price-to-earnings ratio (CAPE). The two metrics found both proponents and critics, who suggested alternative versions that improve the forecasting ability compared to the original. Therefore, in our analysis, we consider a version of CAPE that uses after-tax profits and two versions of Fed model: quantitative dynamic and that accounts for perceived risk. The five models are tested using regression analysis, vector equilibrium correction (VEC), Dickey-Fuller unit-root test using data from 23 developed countries in a time period ranging from 1871 to 2016.

Contents

1	Introduction	2
2	Literature Review	3
2.1	Shiller’s CAPE	3
2.1.1	Theoretical Background	3
2.1.2	Previous Research	5
2.2	Fed model	7
2.2.1	Theoretical Background	7
2.2.2	Previous Research	9
3	Methodology	12
3.1	Hypotheses	12
3.2	Regression Framework and Testing	12
3.2.1	Generic CAPE	12
3.2.2	After-tax Profit CAPE	13
3.2.3	Generic Fed model	13
3.2.4	Fed model adjusted for perceived risk	13
3.2.5	Quantitative Dynamic Version of Fed model	14
3.2.6	Comparing the metrics	15
4	Data	15
5	Time Plan for Completing Thesis	16
	Bibliography	16

1. Introduction

The ability to accurately forecast security returns for purposes of tactical and strategic asset allocation has always been of great value. Today, the investment community has access to a multitude of forecasting frameworks, most of which impose unrealistic assumptions and rarely provide a reliable forecast of future security returns. Usually, the frameworks are grouped in two broad categories: technical and fundamental analysis. Fundamental analysis is concerned with identifying security mispricing, which would enable the implementation of superior investment strategies. The objective is to measure security's intrinsic value by performing analysis of macro- and microeconomic data and factors that have the potential to affect security price. This involves extensive use of performance ratios made up of key figures from financial reports and other metrics.

In this thesis we compare two notable examples of valuation metrics within the domain of fundamental analysis, the Fed model and Shiller's cyclically adjusted price-to-earnings ratio (CAPE). The two are of significant interest as, despite strong criticism and disputed effectiveness, both remain popular in the finance industry. Our goal is to determine which model shows better out-of-sample performance in predicting future stock returns in based on the available data from developed markets.

Our work begins from the expanded version of Vanguard research on stock return forecastability (Davis, Aliaga-Díaz, & Thomas, 2012). The said analysis studies whether 16 popular metrics, including P/E10 and the Fed model, have indeed exhibited correlation with the future stock returns in the period from January 1926 to June 2011. The findings suggest that Shiller's CAPE is the best performing valuation measure, explaining 43% of time variation in future real stock returns compared to 16% exhibited by the Fed model. Another important observation states that the smoothed P/E only provides reasonable long-term forecasts. We proceed by studying the theoretical specifics, underlying assumptions, practical strengths and weaknesses of the two metrics in more detail based on, among others,

Campbell and Shiller (2001), Fama and French (1989), Siegel (2016), Yardeni (1997).

With this thesis we seek to contribute to the body of work that compares valuation metrics in their ability to forecast long-term stock returns and time the market. The scope of the comparison encompasses the original CAPE and Fed models as well their improved versions proposed by Salomons (2004), Koivu, Pennanen, and Ziemba (2005), Clemens (2007) and Siegel (2016).

In the next section, a summary of the theoretical background of the two models is given, followed by an overview of previous research related to the two aforementioned metrics. Next, in Section 3 our hypotheses are formulated and regression frameworks and respective variables used are described, followed by Section 4 presenting the datasets and their sources. The time-plan for completing thesis is given in Section 5.

2. Literature Review

2.1. Shiller's CAPE

2.1.1. Theoretical Background

Cyclically Adjusted Price Earnings ratio (CAPE) is a measure popularized by Robert Shiller following his collaborative effort with John Campbell on a 1998 article "Valuation ratios and the long-run stock market outlook". The current value of Shiller's P/E10, CAPE_t, is defined as the ratio of the real spot price and the 10-year arithmetic average of the lagged real earnings of a the Standard and Poor Composite Index(S&P500) or any alternative broad equity index. Consumer price index data is used to allow conversion to real values. This measure has been influenced by Graham and Dodd (1934). The authors pioneered the use of multi-year earnings per share (EPS) for P/E ratio to smooth the occasional extreme up- and downward spikes in corporate earnings resulting from cyclical effects.

Shiller's work contradicts the conventional random-walk theory, which, implies unpredictability of price changes in the stock market setting. Instead, CAPE

works under assumption of valuation ratio stability and simple mean-reversion theory, whereby the stock prices fluctuate within historic boundaries and any significant deviation from the natural level relative to the earnings figures, is unlikely to persist. Shiller posits that based on the ratio either numerator or denominator should be forecastable. Indeed, having that the stock price is a discounted sum of future earnings and the discount rate is a return required by investor, a valuation, P/E_{10} in this case, that relates current price to earnings can give insight into the value of expected return.

The long-run variation of stock returns can be observed as a response to altering market and economic conditions. At times of economic distress, a low P/E tends to be observed owing to the high levels of risk-aversion exhibited by investors, prompting high expected rates of return and sinking prices. The opposite is observed when stock index prices surge as economy recovers (Fama and French, 1989). An alternative explanation stems from irrational investor behavior. Potential overreliance on use of trailing returns for forecasting as well as overconfidence following a profitable investment may lead to a feedback loop, which, coupled with herd behavior, may lead to bubbles characterized by high P/E ratio (De Bondt and Thaler, 1990). In both scenarios, prices eventually revert to fair value.

In addition, CAPE is assumed to have the capacity to be used as a market-timing tool. This is based off the fact that Shiller accurately predicted the market crash resulting from the subprime mortgage crisis of 2007, after observing the metric in excess of 25. Similar levels were found in data in a periods clustered around 1929 and 1999, notable for the Great Depression and dot-com bubble, respectively. In 1998, Campbell and Shiller suggested that the valuation may deviate from the historical ranges as a result of, among other things, changes in structure of national (US) industry and investor behavior. Today, Shiller's P/E stands at even higher 28.26, which is roughly 69% higher than the historical mean of 16.7. The generally proposed strategy is investing in stocks when CAPE is below its historical average and selling when it is higher (Faber, 2012).

However, in an interview with Business Insider, Robert Shiller states that CAPE is not a market timing mechanism (Blodget and Kava, 2013). Instead, he points out that the metric should be used together with an adequate algorithm for diversification so as to avoid selling off stocks completely. Hence, a lower (higher) CAPE should serve as a signal to increase (decreasing) equity holdings in a portfolio.

2.1.2. Previous Research

In 2001, Shiller and Campbell published an update to their 1988 paper on use of valuation ratios. Alterations included expansion of time-series for annual US stock market data (1871-2000) and inclusion of 30 years of quarterly data (starting in 1970) from for 12 other countries. Their findings from US verify their earlier findings that P/E 10 ratio is fit for use in forecasting changes in future stock prices, which contradicts the simple efficient market models. Study of foreign markets shows mixed evidence of CAPE's efficiency with some results hard to interpret owing to the lack of data. Most markets had roughly 30 years of data, compared to 129 in case of US. Due to this complication the authors resorted to using a 4-year smoothing as opposed to the typical 10-year alternative. The second major point discussed potential factors responsible for the high level of P/E10 ratio, roughly 41, and its non-reversion to the historical mean.

One explanation states that earnings can be biased by the use of new accounting standards applying to investments into intangible assets or "e-capital" (Hall, 2000). The problem came from the fact these investments were considered current expenses and deducted from earnings, leading to a higher P/E ratio. There is no confirmation that investments into intangible assets are in fact so huge in recent decades (no long enough historical time-series data).

Another theory focuses on baby-boomer (born 1946-1964) risk-tolerance. It posits that this demographic tends to favour stocks and willing to pay higher price, resulting in P/E ratio increase in. A version of this argument comes from

Glassman and Hassett (1999), who suggest that the stock prices are too low, with zero risk-premium, leading to an abrupt surge in prices once investors realize this mispricing and baby-boomers could have been the first to see the errors in valuation ratios reflected in the stock prices of the day.

Last argument is linked to inflation decline observed since 1980's. Authors express uncertainty of whether their bearish long-run forecast for stock-returns can be explained by the relationship between inflation and prices. Modigliani and Cohn (1979) argued about irrational behavior of market participants, namely that discounting real dividends at nominal interest rates leads to stocks undervaluation (overvaluation) when inflation is high(low), which was also confirmed by Ritter and Warr (2002).

Shiller (2014) expands the above list. He suggests that stock overvaluation could be explained by higher bond prices, resulting from low inflation (appx. 2%) and 10-year treasury notes yield (2.5%). The author suspects explanations may be rooted in psychology and sociology, with irrational exuberance coming into the equation. A less common explanation suggested that high CAPE may be a result of peoples' anxiety about the future lack of job security as a result of tech progress leading to layoffs.

Faber (2012) extends the research of CAPE's efficiency in predicting future real stock returns. Authors analyse the US market along with 32 other countries. Despite the bias due to shorter time-series collected from international data, results are comparable to domestic, indicating that high-CAPE countries show lower future returns, while low CAPE indicate higher future returns outperforming the market by 4-7%. Keimling (2014) finds more evidence on CAPE performance internationally, in 14 stock markets, concluding that CAPE is not well suited for forecasting short- and medium-run, but can provide realistic long-run equity return expectations.

While CAPE has its proponents, it certainly has its critics. Siegel (2016) suggests that the current ratio levels are too high giving unreasonably pessimistic forecasts of future stock returns. The valuation bias can be attributed to the changes enforced by the Financial Accounting Standards

Board (FASB) in 2001. It stated that the value of financial securities available for trading or sale as well as impairments to the value of fixed assets and intangibles was to be adjusted to reflect the fair market value. The effect of write-downs performed by large companies during the financial crisis of 2008 was even greater due to the so called “aggregation bias” observed in S&P500. It entails computation of index earnings as a net of profits and losses of each listed company irrespective of their market capitalization, leading to a few companies with substantial losses negatively affecting the P/E of the entire index. Siegel suggests using corporate National Income and Product Account (NIPA) after-tax profits which offer more consistent data at expense of data availability, going back to 1928. His study indicates that this method effectively increases the explanatory power of CAPE in forecasting 10-year real stock returns to 40.09%, compared to reported (35%) and operating (36%) S&P earnings.

Kantor and Holdsworth (2014) provide more arguments against using CAPE, suggesting no model based on P/E ratio produces sufficiently accurate returns consistently over time. First, a test for mean reversion using a unit-root and Bai-Perron test uncovered CAPE’s failure to return to historic mean. Second, a major limitation of CAPE is serial correlation, meaning that the regression’s error terms do not have a zero mean and the model fails to provide unbiased and reliable coefficient estimates. A remedy to this problem, was proposed by Lars Peter Hansen (1982). In conclusion, authors recommend that CAPE is not used as definitive market-timing tool, but as a signal to the investor to be cautious (confident) when market is overvalued (undervalued). Recent research by Dimitrov and Jain (2016) shows that market timing, on average is not lucrative with CAPE value above 27.6.

2.2. Fed model

2.2.1. Theoretical Background

Before the Fed model had officially received its name, the concepts it proposes were actively used by financial market players worldwide. The Fed’s “Humphrey-Hawkins Report” of July 22, 1997 states: “changes in this ratio

[S&P500 prices to earnings] have often been inversely related to changes in long-term Treasury yields...” The term “Fed model” was coined by Ed Yardeni in his report on stock market overvaluation (1997). Yet, there is no evidence of it being in official use by the US Federal Reserve (Greenspan, 1997). The model is defined as the stock-bond yield gap, which is the difference between the stock earnings yield (E/P), the inverse of the common P/E ratio, and the 10-year Treasury bond yield (Y).

This metric is based on the fact that stocks and bonds are considered to be competing assets by the investors and postulates that $E/P = Y$. The underlying idea is that by checking whether the aforementioned equality holds, an investor can determine which of the investment vehicles is more attractive. A case where $E/P > Y$ indicates stock yield (price) is higher (lower) than the bond yield (price), making the stocks a more attractive investment alternative. On the other hand, $E/P < Y$ suggests stocks are more expensive, bearing lower yield, which should make them less attractive relative to bonds as in the period of 1991-1997 due to decrease in earnings during the recession. Finally, the $E/P = Y$ parity would suggest the two investment vehicles should be fairly valued and equally desirable (Asness, 2003).

Another explanation revolves around security’s present value argument. The argument is based on the dividend discount model (DDM), stating that the price of stock is a discounted sum of future cash flows (FCF). Same principle can be applied to bonds trading at par value, where the income stream is represented by the annual yield. This relationship is shown in equation below:

$$\frac{P}{E} = \frac{1}{Y} \quad (2.1)$$

All else being equal, it then follows that when interest rates fall, PV(FCF) rises leading to higher P/E for both stocks and bonds. Hence, the two values can be compared. This argument can be extended for use in determining the equilibrium price of stock, by solving the above equation for P. The received value can then be compared to the actual current market price (P^*). If the ratio P^*/P is above (below) 1, then the stock market is overvalued (undervalued).

Lastly, the validity of this comparison is backed by the empirical evidence of historical co-movement of S&P500 E/P and 10-year Treasury bond yields in the period from 1965 to 2001. The two series exhibit a correlation of 0.81.

2.2.2. Previous Research

The arguments in favour of Fed model were met with disapproval. The main criticism of the Fed model revolves around inflation illusion, whereby market participants tend to use real interest rates but nominal growth rates, leading to stock undervaluation when inflation rates are high, and vice versa (Modigliani and Cohn, 1979). Asness (2003) claims that stock's yield (E/P) is not its expected return. So, assuming all else is equal, when nominal expected return on stock moves simultaneously with bond yield, changes are reflected in expected nominal earning growth, not in E/P. This means that E/P does not describe investor's actual return as not all earnings are received by the investor. As earnings are related to inflation, E/P (or earnings yield) is a real return, whereas bond yield is clearly a nominal return. With respect to the PV argument, it is true, all else equal, when interest rates fall, PV of future CFs rises, hence current price rises, and P/E should also increase. However, all else is not equal. For instance, if inflation falls, future nominal CFs from equity also fall which can offset the impact of lower rate, suggesting that the Fed model ignores this "counter-effect". Finally, if previous two arguments fail then the historic evidence of high correlation between S&P500 E/P and 10-year US Treasury bond yield during the post-war period is just a proof of investors' blindness to biased model.

Thomas and Zhang (2008) suggest a counter-argument, stating that investors are rational and being aware of the inflation illusion are able to account for its three roots, thereby avoiding it. The first root is that the earnings yield should not move with inflation. Thomas and Zhang respond that there are accounting policies which imply that the record of "inflationary holding gains" leads to co-movement of earnings (thus, earnings yield) with inflation. The second aspect of inflation illusion is that the nominal growth rates should move with inflation. The counter-argument is that the relevant growth rate should be

a relatively stable perpetual dividend growth rate under a full, not current, payout policy. And according to accounting rules, this dividend growth rate does not exhibit significant variation with expected inflation, which contradicts the second argument. As their third response to inflation illusion, authors address the empirical observations in the “dynamic” market. They use 1976-2007 market data from US and 6 other countries. Findings indicate that the forecast errors, which used a “relatively constant” nominal growth rate, were uncorrelated with expected inflation. Additionally, the researchers argue that these forecasts rely on the logical outcome that in the periods of low inflation the expected equity market growth is higher, and vice versa. Their conclusion is that Fed model can be a useful valuation tool as it provides an intuition on risk premium and anticipated growth.

These findings were confirmed by Bekaert and Engstrom (2010), who are also very skeptical about the explanatory power of inflation illusion. In addition, they argue that the high positive correlation between real equity yields and nominal bond yields is a result of increased uncertainty, risk-aversion during and higher expected inflation during economic downturn, which leads to higher stock and bond yields. These results were confirmed using a cross-sectional regression analysis of data from 20 countries, revealing that stagflation is highly correlated with high correlation of stock - bond yields.

Estrada (2009) shows that the Fed model can work, but the required assumptions are unrealistic. He considers a constant-growth dividend discount model:

$$P = \frac{\delta(1 + g)}{r_f + RP - g} \quad (2.2)$$

With price (P), dividend (δ), dividend growth(g), risk-free rate (r_f) and risk premium for holding riskier stocks instead of bonds (RP). Both sides of the equation are divided by forward earnings (E) under the assumptions that all earnings are distributed as dividends ($\delta(1 + g) = E$) and that dividends do not grow ($g = 0$). The last assumption is that investors do not require higher returns from investing in stocks than in bonds (RP = 0). Finally, the Fed

model itself is derived:

$$\frac{P}{E} = \frac{1}{r_f} \quad \Rightarrow \quad \frac{E}{P} = r_f \quad (2.3)$$

The result was supported by Siegel and Coxe (2002). The Fed is derived as $\frac{P}{E} = \frac{(1+g)}{r_f}$, arguing that investor would always take into account inflation, and therefore consider higher growth and higher risk (approximately offsetting each other ($RP = g$)) of stocks rather than bonds.

Salomons (2004) also shows that the Fed model is not a feasible for forecasting equity returns. He suggests that model's predictive power over-short horizons can be improved by adding an error term (EY_e), which accounts for the perceived risk of the two asset classes, to a regression. The interpretation is that the proposed term picks up market mispricing or changing investor risk aversion. Equity can be considered overvalued, indicating higher returns in short-term, if EY_e is high given a scenario with fixed volatility and interest rate. We will try to replicate and test the results of this study.

Koivu et al. (2005) introduce a quantitative dynamic version of Fed model, which based on co-integration analysis, can forecast changes in stock prices, earning and bond yields. The model relies on use of logarithmic indicator, suggesting that by using logarithms of positive numbers can improve the model's fit better than initial data and in variations of positive variables can be more descriptive, compared to absolute values.

Clemens (2007) proposes that the inclusion of confidence intervals in the Fed framework improves the accuracy of predictions by reducing bias stemming from investor's irrational behaviour. Investment decisions made with consideration of However, he also points out that it is better suited for predicting relative returns of stock versus bonds, as opposed to absolute stock returns. The author concludes that it is optimal to employ the Fed model for prediction over a short and medium horizon of up to 3 years, thereby complementing CAPE's long-run forecasts.

3. Methodology

3.1. Hypotheses

Based on the previous research, we formulate the following hypotheses:

Ho: Using NIPA after-tax profits significantly improves the power of CAPE in forecasting 10-year real stock returns in comparison to reported and operating S&P500 earnings data (international alternatives);

Ho: Adding a dummy variable to account for transaction costs improves ;

Ho: Quantitative dynamic version of Fed model shows improved forecasting power compared to the original;

Ho: Using “error-term” significantly improves the forecasting power Fed model;

Ho: CAPE outperforms Fed model in short- and middle-term forecasts;

Ho: CAPE outperforms Fed model in long-term forecasts;

Ho: Bond yield has significant explanatory power in forecasting stock returns;

3.2. Regression Framework and Testing

3.2.1. Generic CAPE

In accordance with CAPE methodology, the current value of Shiller’s P/E10 is:

$$CAPE_t = \frac{P_t}{\frac{1}{10}(E_{t-1} + E_{t-2} + \dots + E_{t-10})} \quad (3.1)$$

Where P_t represents the real spot price and E_{t-1} refers to the lagged real earnings. Then, the future real stock return (R_t) is regressed on the logarithm of $CAPE_t$:

$$R_t = \alpha + \beta \log CAPE_t + \epsilon_t \quad (3.2)$$

Regression coefficients, intercept (α), slope (β), and the error term (ϵ_t) are estimated.

3.2.2. *After-tax Profit CAPE*

As suggested by Siegel (2016), after-tax profit data is used to replace reported and operating earnings of S&P500 index and international alternatives (if available). Addition of a dummy variable (D_1) would make the model robust to changes in real earnings growth and remove potential endogeneity problem. Transaction costs (TR) can serve as one such factor, as lower transaction costs lead to returns closer to those obtained from broad equity index. Then, a modified version of equation 3.2 would take the following form:

$$R_t = \alpha + \beta \log CAPE_t + D_1 TR + \epsilon_t \quad (3.3)$$

3.2.3. *Generic Fed model*

Equity returns (R_{t+i}) are forecasted using the regressions for the Fed model as proposed by Asness (2003) are presented below:

$$R_{t+i} = \alpha + \beta_1 EY_t + \epsilon_t \quad (3.4)$$

$$R_{t+i} = \alpha + \beta_1 (EY_t - BY_t) + \epsilon_t \quad (3.5)$$

$$R_{t+i} = \alpha + \beta_1 EY_t + \beta_2 BY_t + \epsilon_t \quad (3.6)$$

We have α, β_1, β_2 , - constants, EY_t - equity yield (E/P), BY_t - bond yield (Y), ϵ_t - error term.

3.2.4. *Fed model adjusted for perceived risk*

We test whether the adding the error term, which account for perceived risk, proposed by Salomons (2004) improves the predictive power of Fed model given the available US and international data. For this, two variables are added to the generic model: fitted value (EY_f). an error term (EY_e), obtained using the regressions below.

$$EY = \alpha + \beta_1 BY_t + \epsilon_t \quad (3.7)$$

$$EY_f = \alpha + \beta_1 BY_t + \beta_2 \sigma_t^e + \beta_3 \sigma_t^b + \epsilon_t \quad (3.8)$$

where σ_t^e and σ_t^b are the equity and bond volatility, respectively. Then, the error term is:

$$EY_e = EY - EY_f \quad (3.9)$$

3.2.5. *Quantitative Dynamic Version of Fed model*

The linear time-series model is built around the log indicator. A test of whether the indicator improves the predictive power of the model is conducted. The method is based on the work by Koivu et al. (2005). We study whether Fed model is able to forecast changes in future stock prices, bond yields and earnings yields using the following “logarithmic indicator”:

$$I_t = \ln\left(\frac{Y_t}{E_t/P_t}\right) = \ln Y_t - \ln E_t + \ln P_t \quad (3.10)$$

I_t is log indicator, Y_t is bond yield, E_t is earnings, P_t is stock price index. According to Engle and Granger (1987), I_t should be an “equilibrium correction term” in a VEC model (Vector Equilibrium Correction), that is I_t is linear in all given logarithmic variables. Alternative is to use VAR (Vector Autoregressive) model, but Koivu et al. (2005) suggests that VEC does better.

First, time-series (log bond yield, log stock price index, log earnings) should be organized accordingly with trend, stationarity and drift using augmented Dickey-Fuller methodology (Dickey and Fuller, 1981).

Second, indicator I_t is tested for stationarity using augmented Dickey-Fuller test to determine whether there is a trend across the markets. Third, if I_t is found to be (fairly) stationary in all markets, which means that the time-series might be cointegrated, then the VEC models can be estimated. Using VEC model we test if I_t can be used to explain variations in the time-series. VEC can be obtained from VAR (Campbell and Shiller, 1998) by adding an “equilibrium correction term”. Vector process:

$$x_t = \begin{bmatrix} \ln Y_t \\ \ln P_t \\ \ln E_t \end{bmatrix} \quad (3.11)$$

We get a linear time-series VEC model:

$$\Delta x_t = c + \sum_{i=1}^k A_i \Delta x_{t-i} + \alpha(\beta' x_{t-1} - \mu) + \epsilon_t, \quad \epsilon_t \sim N(0, \Sigma) \quad (3.12)$$

where a covariance matrix is represented by $c \in \mathbb{R}^3$, $A_i \in \mathbb{R}^{3 \times 3}$, $\beta \in \mathbb{R}^{3 \times l}$, $\mu \in \mathbb{R}^l$, $\alpha \in \mathbb{R}^{3 \times l}$ (speed of adjustment to equilibrium) and $\Sigma \in \mathbb{R}^{3 \times 3}$. The long-run movement of x_t is explained by $\beta' x = \mu$, which suggests that assuming $E[\beta' x_t] = \mu$ in the long-run, x_t reverts to equilibrium following initial deviation from them.

3.2.6. Comparing the metrics

The explanatory power (R^2) of the given metrics is compared. Examination of predictive power will be conducted over investments of the following lengths: 3 months, 6 months, 1 year, 2 years, 3 years, 5 years, 7 years, 10 years, 15 years.

We test for statistical significance of results to account for potential overlapping observations when a horizon of less than one year is used as different points are not independent statistically. In case of lack of data for developed markets, the smoothing period will be reduced to be under 10 years.

Also, to compare short-term effects of the models, variance decompositions will be estimated. Similar approach was suggested by Durré and Giot (2007).

4. Data

Data on S&P500 Index prices, earnings, dividends and consumer price index (CPI) for this study has been sourced from Robert Shiller's web site. The sample includes monthly observations in the period of January 1871 through December 2016, giving the ability to measure intrinsic values of securities.

Monthly US 10-year constant maturity treasury bond yield observations have been obtained from the Federal Reserve Economic Data (FRED). The sample is 1962:01 to 2016:12. International analogues are yet to be found.

Using Datastream software, MSCI World index benchmark data is obtained, covering the large and mid-cap equity performance across all segments in 22 developed markets (US excluded). Respective historic CPI figures will also be sourced using the aforementioned database.

National Income and Product Account (NIPA) after-tax profits for use in alternative CAPE version proposed by Siegel. International analogues are yet to be found.

5. Time Plan for Completing Thesis

Our first step to completing the thesis will be to perform additional research on methodology and gathering international data. Depending on data availability, we will select developed countries for our analysis from the pool of the 22 available (US excluded) from MSCI World Index. Necessary adjustments and additions to the thesis based on the feedback from the preliminary report presentation and our supervisor will be made. Finally, we will begin writing a code for our regressions and tests. We estimate that this step will be completed by middle of May 2017.

Next, in mid-June we plan on running the regressions, analysing the results and drawing conclusions with respect to the thesis goals. We will work towards compiling the first draft of the final thesis to have it submitted by the beginning of July 2017. Final changes and editing will be done to meet the deadline of September 1st 2017.

Bibliography

- Asness, C. (2003). Fight the fed model. *Journal of Portfolio Management*, 30(1):11.
- Bekaert, G. and Engstrom, E. (2010). Inflation and the stock market: Understanding the “fed model”. *Journal of Monetary Economics*, 57(3):278–294.
- Blodget, H. and Kava, L. (2013). Shiller: Stocks are priced for (relatively) crappy returns. retrieved from. <http://www.businessinsider.com/robert-shiller-on-stocks-2013-1?r=US&IR=T&IR=>.
- Campbell, J. Y. and Shiller, R. J. (1998). Valuation ratios and the long-run stock market outlook. *The Journal of Portfolio Management*, 24(2):11–26.
- Clemens, M. (2007). A behavioral defense of the fed model. *Available at SSRN 958800*.
- De Bondt, W. F. and Thaler, R. H. (1990). Do security analysts overreact? *The American Economic Review*, pages 52–57.
- Dickey, D. A. and Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica: Journal of the Econometric Society*, pages 1057–1072.
- Dimitrov, V. and Jain, P. C. (2016). Shiller’s cape: Market timing and risk.
- Durré, A. and Giot, P. (2007). An international analysis of earnings, stock prices and bond yields. *Journal of Business Finance & Accounting*, 34(3-4):613–641.
- Engle, R. F. and Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, pages 251–276.
- Estrada, J. (2009). The fed model: The bad, the worse, and the ugly. *The Quarterly Review of Economics and Finance*, 49(2):214–238.
- Faber, M. T. (2012). Global value: Building trading models with the 10 year cape. *Cambria Quantitative Research*, (5).
- Fama, E. F. and French, K. R. (1989). Business conditions and expected returns on stocks and bonds. *Journal of financial economics*, 25(1):23–49.
- Glassman, J. K. and Hassett, K. A. (1999). *Dow 36,000: The new strategy for profiting from the coming rise in the stock market*. Times Business New

York.

- Graham, B. and Dodd, D. L. (1934). *Security analysis: principles and technique*. McGraw-Hill.
- Greenspan, A. (1997). Federal reserve board humphrey hawkins testimony.
- Hall, R. E. (2000). E-capital: The link between the stock market and the labor market in the 1990s. *Brookings Papers on Economic Activity*, 2000(2):73–102.
- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, pages 1029–1054.
- Kantor, B. and Holdsworth, C. (2014). 2013 nobel prize revisited: Do shiller’s models really have predictive power? *Journal of Applied Corporate Finance*, 26(2):101–108.
- Keimling, N. (2014). Cape: Predicting stock market returns. *StarCapital Research Publication, February*.
- Koivu, M., Pennanen, T., and Ziemba, W. T. (2005). Cointegration analysis of the fed model. *Finance Research Letters*, 2(4):248–259.
- Modigliani, F. and Cohn, R. A. (1979). Inflation, rational valuation and the market. *Financial Analysts Journal*, 35(2):24–44.
- Ritter, J. R. and Warr, R. S. (2002). The decline of inflation and the bull market of 1982–1999. *Journal of financial and quantitative analysis*, 37(01):29–61.
- Salomons, R. (2004). A tactical implication of predictability: Fighting the fed model. *Available at SSRN 517322*.
- Shiller, R. J. (2014). The mystery of lofty stock market elevations. *The New York Times*.
- Siegel, J. J. (2016). The shiller cape ratio: A new look. *Financial Analysts Journal*, 72(3):41–50.
- Siegel, J. J. and Coxe, D. G. (2002). *Stocks for the long run*, volume 3. McGraw-Hill New York.
- Thomas, J. and Zhang, F. (2008). Don’t fight the fed model. *Unpublished paper, Yale University, School of Management*.
- Yardeni, E. (1997). Fed’s stock market model finds overvaluation. *Topical study*, 38.