UNIVERSITY OF TURKU
Turku School of Economics

Forecasting Equity Premium with Technical and Macroeconomic Indicators

Department of Economics
Master's thesis

Author:
Emil Lehti

Supervisor:
Professor Timo Kuosmanen

12.5.2023
Turku

Master's thesis

Subject: Economics
Author: Emil Lehti
Title: Forecasting Equity Premium with Technical and Macroeconomic Indicators
Supervisor: Professor Timo Kuosmanen
Number of pages: 84 pages + appendices 17 pages
Date: 12.5.2023

This thesis investigates the forecasting of the equity premium, a critical metric in financial economics, representing the difference between the expected return on a stock market portfolio and the risk-free rate. Accurate equity premium forecasts are paramount for asset allocation, risk management, and financial market regulation. While financial theory posits that stock prices should align with future discounted cash flows, empirical forecasting remains challenging.

In this study, a variety of forecasting methods and variables are assessed for their ability to predict the equity premium. The review of recent literature highlights the crucial role of model selection and parameterization in predicting the equity premium. The study acknowledges the contribution of alternative model specifications, which address model uncertainty and parameter instability. These models have demonstrated their potential to yield statistically and economically significant forecasts, outperforming the historical average forecast. The thesis corroborates the literature, suggesting that forecasting during recessions yields superior results compared to the historical average, while forecasting during expansion periods poses a greater challenge.

This research conducts a meticulous examination of macroeconomic predictors and technical indicators, utilized within diverse model specifications, to forecast the equity premium using updated data. Established macroeconomic predictors and technical indicators often fail to produce statistically or economically significant forecasts during expansion periods. Nevertheless, when macroeconomic predictors are employed within multiple-predictor models, investors can realize benefits surpassing those of the historical average forecast. During recessions, forecasting is comparatively less challenging, with technical indicators delivering the best forecasts both statistically and economically. Owing to the inherent stability of technical indicators, their incorporation into multiple-predictor models doesn't yield any additional value.

This study puts forth a strategic recommendation to enhance the economic advantage of equity premium forecasts. It suggests that an optimal approach could involve the deployment of multiple-predictor models that use macroeconomic predictors during periods of economic expansion, and individual technical indicators during recessions. This contribution to the discourse on equity premium forecasting advocates for a state-dependent forecasting methodology.

Future research could explore this state-dependent forecasting methodology further. This could involve the development and rigorous testing of state-dependent forecasting models, as well as the identification of the most suitable predictors for each economic state. While forecasting during recessions appears to be easier, it could be beneficial to examine different benchmark models depending on the current state of the economy. The widely used benchmark in current literature, the historical average model, consistently predicts positive equity premiums, even though these are generally negative in reality during recessions. Therefore, it might be prudent to develop a benchmark model that also depends on the state of the economy and compare the generated state-dependent forecasts to this model. This approach could provide a more accurate comparative measure for evaluating forecasting strategies during different economic conditions.

Key words: equity premium forecasting, technical indicators, macroeconomic predictors, business cycle

Pro gradu -tutkielma

Oppiaine: Taloustiede
Tekijä: Emil Lehti
Otsikko: Forecasting Equity Premium with Technical and Macroeconomic Indicators
Ohjaaja: Professori Timo Kuosmanen
Sivumäärä: 84 sivua + liitteet 17 sivua
Päivämäärä: 12.5.2023

Tässä tutkielmassa tarkastellaan osakemarkkinoiden tuotto-odotuksen ja riskittömän korkotuoton välistä eroa eli osakepreemion ennustamista. Tarkat osakepreemio -ennusteet ovat keskeisiä varainhoidossa, riskienhallinnassa ja rahoitusmarkkinoiden sääntelyssä. Vaikka rahoitusteorian, mukaan osakekurssien tulisi vastata tulevaisuuden diskontattuja kassavirtoja, osakepreemion empiirinen ennustaminen on haastavaa.

Tässä tutkimuksessa arvioidaan erilaisten ennustusmenetelmien ja muuttujien kykyä ennustaa osakepreemiota. Kirjallisuuskatsaus korostaa mallin valinnan ja parametrisoinnin olevan keskeisessä roolissa osakepreemiota ennustettaessa. Tutkimus tunnustaa vaihtoehtoisten mallisääntöjen merkityksen, jotka käsittelevät mallin epävarmuutta ja parametrien epävakautta. Näiden mallien on osoitettu pystyvän tuottamaan historiallista keskiarvoa tarkempia ennusteita niin tilastollisesti kuin taloudellisesti merkitsevästi. Tutkielma tukee kirjallisuutta, jonka mukaan taantumissa ennustaminen on suhteessa helpompaa verrattuna historialliseen keskiarvoon, kun taas laajentumiskausina ennustaminen on haastavampaa.

Tutkimuksessa tarkastellaan makrotaloudellisia ennustajia ja teknisiä indikaattoreita, joita käytetään sekä yksittäisinä ennustajina että laajemmissa ennustemalleissa, joita käytetään osakepreemion ennustamiseen päivitetyllä aineistolla. Vakiintuneet makrotaloudelliset ennustajat ja tekniset indikaattorit eivät usein pysty tuottamaan tilastollisesti tai taloudellisesti merkitseviä ennusteita talouden nousukausina. Kuitenkin, kun makrotaloudellisia ennustajia käytetään laajemmissa ennustemalleissa, sijoittajat voivat saavuttaa taloudellisia etuja verrattaessa ennusteiden muodostamiseen historiallisen keskiarvon mukaisesti. Taantumissa ennustaminen on helpompaa, ja tekniset indikaattorit tuottavat tarkimmat ennusteet sekä tilastollisesti että taloudellisesti. Teknisten indikaattorien luontaisen vakauden vuoksi niiden käyttö laajemmissa ennustemalleissa ei tuota lisäarvoa. Erityisesti taantumissa on perusteltua käyttää teknisiä indikaattoreita itsenäisinä osakepreemion ennustajina.

Tutkielman mukaan optimaalinen lähestymistapa osakepreemion ennustamiseen on teknisten indikaattorien käyttö taantumissa ja makrotaloudellisten ennustajien käyttö laajemmissa ennustemalleissa talouden nousukausina. Tämä kannustaa tarkastelemaan osakepreemion ennustamista tilariippuvaisilla ennustemenetelmillä.

Katsoen eteenpäin, tulevat tutkimukset voisivat tutkia tätä tilasta riippuvaa ennustusmenetelmää tarkemmin. Tämä voisi sisältää tilasta riippuvien ennustusmallien kehittämisen ja perusteellisen testaamisen, sekä sopivimpien ennustajien tunnistamisen kullekin taloudelliselle tilalle. Nykyisessä kirjallisuudessa laajalti käytetty vertailumalli, historiallinen keskiarvo, ennustaa jatkuvasti positiivisia osakepreemioita, vaikka todellisuudessa taantumissa osaketuotot, ja osakepreemiot ovat negatiivisia. Siksi saattaisi olla mielekästä tarkastella ennusteita eri vertailumallilla, joka myös riippuu talouden tilasta. Tämä lähestymistapa voisi tarjota tarkemman vertailumittarin ennustestrategioiden arvioimiseksi eri taloudellisissa olosuhteissa.

Avainsanat: osakepreemion ennustaminen, tekniset indikaattorit, makrotaloudelliset ennustajat, suhdannevaihtelu

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 Introduction

The equity premium, the difference between the expected return on a stock market portfolio and the risk-free rate, is an essential measure for both investors and financial economists. Accurate forecasts of the equity premium have significant implications for asset allocation, risk management, and financial market regulation. According to theoretical asset pricing and financial economics, stock prices should match their future discounted cash flows. However, forecasting equity premium is not so black and white, as forecasting discounted future cash flows (dividends) is not so simple. Goyal and Welch (2008) found that forecasting equity premium with dividend yields, as well as with many other macroeconomic or financial variables, works very poorly. The present values of discounted cash flows depend on both future dividends and discount rates. Indeed, the empirical literature has also tried to answer the question which, dividends or discount rates, has a greater impact on the variation of the equity premium? (see, e.g. Doan & Lan 2022).

In their famous article "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction" Goyal and Welch (2008) argue that equity premium has not been predictable and that for a large number of variables, no agent could have obtained a prediction advantage over the historical average. However, Pesaran and Timmermann (1995) and Lamoureux and Zhou (2015) argue that the predictability of the equity premium is not necessarily a matter of choosing the right variables, but a problem of choosing the right model and its parametrization. Indeed, recent studies have shown that alternative model specifications, that address model uncertainty and parameter instability, can yield statistically and economically significant forecasts that are more accurate than the naive forecast (historical average) (see, e.g., Rapach & Zhou, 2013). These different model specifications (strategies) include, for example, economically motivated restrictions implemented on either predictors or resulting equity premium forecasts (Campbell & Thompson, 2008; Pettenuzzo et al., 2014; Pan et al., 2020), forecast combinations (Rapach et al., 2010), diffusion indices (Ludvigson & Ng, 2007; Kelly & Pruitt, 2013), technical indicators (Neely et al., 2014) and regime shifts (Guidolin & Timmermann, 2007; Henkel et al., 2011; Dangl & Halling, 2012) and machine learning (Gu et al. 2020).

As the number of suggested models and predictors grows, the need for a comprehensive comparison increases. In finance literature, it is often found that forecasting during a recession appears to be much easier than during an expansion (see, for example, Pan et al., 2020). Neely et al. (2014) introduce technical indicators as predictors alongside macroeconomic predictors. In this study, I examine how macroeconomic predictors and technical indicators with different model specifications can forecast the equity premium. In summary, I find that different sets of predictors have varying predictive abilities in different economic states. Macroeconomic predictors used in models that consider model uncertainty and parameter instability can generate added value for investors during expansions. Conversely, technical indicators can create economic benefits during recessions.

In Section 2, I examine linear regression, principal components analysis and model selection, that are essential for understanding statistical forecasting methods used in equity premium forecasts. In Section 3, I explore conventional models and theory on portfolio management, which are crucial in equity premium forecasting literature. In Section 4, I delve into technical analysis and provide economic justification for its use. In Section 5, I introduce the conventional data and technical indicators used in equity premium forecasting and review post-Goyal and Welch (2008) literature on equity premium forecasting. In Section 6, I present methods for forecasting evaluation and the forecasting models used in my analysis. In Section 7, I delve into the out-of-sample forecast results for equity premium. Utilizing updated data, I assess a variety of forecasting methods and variables. These results are then evaluated in terms of their forecasting ability. Both economic and statistical significance serve as key criteria in this comparative analysis. In Section 8, I present conclusions and motivation for further research.

2 On Forecasting Models and Model Selection

In this section, I introduce the basics of linear regression, principal component analysis, and model selection for the future purposes of this thesis. Section 2.1 is based on Friedman et al. (2009, 66, 79–80, 534–536) and Greene (2012, 52–56), while section 2.2 is based on Efron et al. (2004) and Konishi and Kitagawa (2008, 211–239).

2.1 Linear Regression and Principal Components

In financial econometrics, equity premium forecasts are typically made using linear regressions and principal component regressions (Goyal & Welch, 2008; Campbell & Thompson, 2008; Neely et al., 2014; Pan et al., 2020). Therefore, I will provide a more detailed explanation of the linear regression model and principal component regression.

Linear Regression

Linear regression is a statistical method that models the relationship between a dependent variable and one or more independent variables. The objective of a linear regression is to identify the line of best fit that accurately represents the observed data, enabling predictions to be made about the dependent variable based on the independent variables.

Formally, let Y be the dependent variable and $X_1, X_2, \dots, X_p$ be the independent predictive variables, such as technical indicators and macroeconomic predictors. The linear regression model is expressed as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon, \tag{1}$$

where β_0 represents the estimated value of the dependent variable when all independent variables are equal to zero. Coefficients β_i, $i = 1, .., p$, represent the estimated change in the dependent variable for a unit change in X_i, holding all other independent variables constant.

The coefficients are estimated by minimizing the sum of squared deviations between the observed data and the line of best fit. Once the coefficients have been estimated, the linear regression model can be employed to make predictions about Y based on the values of the independent variables.

Principal Components

PCA is a linear dimensionality reduction technique that converts the original data into a new set of uncorrelated variables, known as principal components (PCs), which retain as much of the original information as feasible. The first PC captures the maximum variance in the data, followed by the second PC, and so forth. The PCs enable data visualization in a lower-dimensional space, which is beneficial for exploratory data analysis, visualization, and pattern recognition. Mathematically, PCA is founded on the singular value decomposition (SVD) of the data's covariance matrix. The SVD of a data matrix X is expressed as $X=U\Sigma V'$, where U and V are orthogonal matrices, and Σ is a diagonal matrix of singular values. The columns of U, the left singular vectors, signify the new principal component axes in the transformed space, while the singular values in Σ represent the variance explained by each principal component.

As an unsupervised learning technique in machine learning, PCA does not utilize prior information about the relationships between the variables in the data. Instead, it leverages the covariance matrix structure to identify the most significant directions in the data. PCA can also be employed to reduce data noise by discarding PCs with smaller singular values.

PCR, a variant of linear regression, employs principal components as predictors rather than the original variables. The underlying concept of PCR is to use the first few PCs, which capture the majority of the data variability, as predictors in a linear regression model. This reduces the number of predictors and the risk of overfitting[1]. In PCR, the original data matrix X is initially transformed into a new matrix Z by projecting it onto the principal component axes. The transformed data Z is subsequently used as input for a linear regression model to predict a response variable Y. The regression coefficients can be estimated by minimizing the residual sum of squares. PCR can be viewed as a fusion of PCA and linear regression, with PCA reducing the data dimensionality and linear regression enabling predictions.

PCA and PCR are powerful tools for data analysis and modelling. By reducing the dimensionality of the data and preserving the most important information, PCA and PCR

[1] An overfitted model exhibits high accuracy on the training dataset but performs poorly on new, unseen data due to its inability to generalize from the training data to the broader population. This is because it captures not only the underlying patterns but also the noise or random fluctuations in the training data.

can make it easier to visualize and analyze complex datasets, as well as to build predictive models.

2.2 Model Selection

In a forecasting situation, the predictor space is often high dimensional, and choosing too many predictors can lead to overfitting, which may result in inaccurate forecasts. Therefore, econometricians need to employ various model selection methods to assess the suitability of different variables in the forecasting context. Here, model selection refers to the selection of variables instead of choosing the functional form of the model. In the case of a linear model, typical model selection algorithms include forward and backward stepwise, best subset selection, and machine learning-based model selection like LASSO (Least Absolute Shrinkage Selection Operator). In the case of all these model selection methods, except for LASSO, the suitability of predictors should be examined using various criteria, such as the coefficient of determination (R-squared) and adjusted R-squared, as well as information criteria such as the Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC). These criteria are suitable for evaluating nested[2] models, such as the selection of predictors in linear or time series models.

Following, for example, Cremers (2002), the Bayesian Information Criterion (BIC) is used for predictor selection in predictive linear regression in this thesis. Thus, we take a closer look at the BIC.

Bayesian information criterion

Bayesian Information Criterion (BIC), also known as Schwarz information criterion (SIC), is a widely used statistical criterion for model selection, particularly in the context of model comparison and choosing the best model among a set of candidate models. BIC is derived from the Bayesian perspective and provides a trade-off between model fit and model complexity. It is especially useful in situations where the sample size is large, as it

[2] Nested models refer to a set of models where one model is a simpler or restricted version of another. In other words, the simpler model can be derived from the more complex one by imposing constraints on its parameters. Nested models share the same basic structure and a common set of predictors, but some predictors are omitted or their effects are constrained in the simpler model. On the other hand, non-nested models are those that do not share a common structure or set of predictors and cannot be derived from one another by imposing constraints on their parameters.

tends to penalize more complex models more heavily compared to other criteria, such as the Akaike Information Criterion (AIC).

The BIC is defined as follows:

$$BIC = -2\, log(L) + p\, log(n), \tag{2}$$

where L represents the maximized likelihood of the model, p denotes the number of estimated parameters in the model, and n is the sample size. The first term, -2 log(L), measures the model fit, while the second term, p log(n), is a penalty for model complexity.

The intuition behind the BIC is to find a balance between the goodness of fit and the simplicity of the model. A model with more parameters may fit the data better, but it might also be more complex and harder to interpret. Moreover, complex models may overfit the data, capturing noise instead of the underlying structure, which could lead to poor generalization and prediction performance on new data.

When comparing multiple-predictor models, the model with the lowest BIC value is considered the best model. This is because a lower BIC value indicates a better balance between model fit and complexity. It is important to note that BIC is an asymptotic approximation and may not perform well in small sample sizes.

3 Conventional Models and Theory on Portfolio Management

In this section, I introduce the theory related to equity premium forecasts, including market, returns, and portfolio management perspectives. Section 3.1 is based on Franke et al. (2015, 54-55) and Welch (2017, 247-249, 277-312), Section 3.2 is based on Cochrane (2005, 143-173) and Welch (2017, 213-238), Section 3.3 is based on Welch (2017, 37-53), Section 3.4 is based on Markowitz (1952) and Cochrane (2005, 77-119), and Section 3.5 is based on Welch (2017, 281-287).

3.1 Efficient Market Hypothesis

According to Eugene Fama (1970), the ideal market would be one in which prices accurately signal all possible information available. Such a market is called an efficient market. The conditions of an efficient market can be divided into three conditions:

1. A weak condition where stock prices reflect all historical information such as historical prices and historical trading volumes. According to Fama (1970), a stock market does not have a memory and thus stock market returns are independently distributed. Given this case, no investor could benefit from technical analysis.

2. Semi-strong condition, where the information set reflecting current prices is extended with other publicly available data, such as macro data on stocks, stock splits, etc. In this case, fundamental analysis is unable to achieve excess returns.

3. Strong condition, which looks at whether some individual groups or institutions have potential price influences. Monopolistic information that is not publicly available to others. In this case, even insider information cannot achieve excess returns, and this information is also reflected in stock prices.

Consistent market conditions for market efficiency are (i) there are no transaction costs, (ii) all information is freely available to all market participants, (ii) all market participants agree on the effect of the information for each item. In practice, however, there is no market in which all these are perfectly realized. Fortunately, according to Fama, these conditions are not necessary.

In the early literature Fama (1970) conducts empirical tests for weak, semi-strong and strong conditions, which are supported by the empirical findings of several other researchers. Fama (1965) finds consistently positive autocorrelation between day-to-day stock prices, but still very close to zero. Indeed, Fama finds that day-to-day stock prices can be assumed to be independently distributed. In other words, in real life stock returns can in fact be assumed to follow a random-walk pattern. Cootner (1964) also finds a very weak, in this case negative, correlation in weekly stock returns. However, these correlations are also so weak that, rather than correlated stock returns, their analysis also supports the view that stock returns are not correlated with these lags.

The fact that stock returns are not autocorrelated indeed supports the assumption that stock prices follow a random walk process. Clearly, stock prices consist of their historical returns. Consider stock returns, R, as identically, and as the lack of autocorrelation advocates, independently distributed random variables. In this case, the stock price at time t can be represented as the sum of historical returns. This can be written as

$$P_t = \sum_{j=0}^{t} R_j, \tag{3}$$

where P_t is price at time t and R_j is the stock return at time j, $j = (0, \ldots, t)$. Alternatively, stock price can be formulated as

$$P_t = P_{t-1} + R_t, \tag{4}$$

where stock price at time t is just the previous price plus the return on from period $t-1$ to t. Indeed, this is the exact form of Random Walk (see Appendix 1. Basic Concepts).

The Efficient Market Hypothesis is a crucial theory when examining equity premium forecasts. If the conditions of the EMH are met, neither macroeconomic predictors nor technical indicators would be capable of predicting the equity premium at all. In Section 4, I explore the critiques of the efficient market hypothesis and provide arguments in favor of technical analysis.

3.2 The Capital Asset Pricing Model

The Capital Asset Pricing Model (CAPM) is a widely used financial theory that helps investors understand the relationship between risk and expected return. The CAPM model does not have a single clear inventor, but the scientific discoveries of Sharpe (1964), John

V. Lintner (1965), and Mossin (1966) are often mentioned as the greatest contributors. Markowitz's (1952) earlier scientific breakthrough in the field of portfolio theory also contributed to the development of the CAPM theory. The theory has become a cornerstone of modern financial economics, providing a framework for pricing and managing financial assets.

At its core, the CAPM states that the expected return on an asset is a function of its beta, or the degree to which it is correlated with the overall market, and the market equity premium, or the excess return required by investors for bearing market risk. Specifically, the expected return on an asset is equal to the risk-free rate plus the product of its beta and the market risk premium.

According to the CAPM, the expected return on an asset can be calculated by adding the risk-free rate to the product of the asset's beta and the expected market risk premium. The risk-free rate is the return an investor expects to receive on an investment that is considered risk-free, such as a U.S. Treasury bond. Beta measures the volatility of the asset relative to the overall market, with a beta of 1 indicating that the asset moves in tandem with the market, while a beta greater than 1 suggests that the asset is more volatile than the market. Finally, the market risk premium is the additional return investors expect to receive for holding a risky asset compared to a risk-free asset.

The formula for calculating the expected return on an asset using the CAPM is as follows:

$$r_a = r_f + \beta_a(r_m - r_f), \tag{5}$$

where r_f is the risk-free rate, β_a is the beta of asset a, r_m is the expected market return and $r_m - r_f$ is the equity premium.

The CAPM is widely used in finance to estimate the expected return on an asset or portfolio. For instance, an investor could use the CAPM to estimate the expected return on a stock by calculating its beta and the market risk premium. This information could then be used to make investment decisions and to build diversified portfolios.

While the CAPM has its limitations, it remains an important tool for investors and financial economists. For instance, critics have noted that the CAPM relies on several assumptions that may not hold in the real world, such as perfect information and rational investor behaviour. Additionally, the CAPM assumes that all investors have the same expectations and that markets are efficient, which may not be true in practice.

Overall, the CAPM provides a useful framework for understanding the relationship between risk and expected return of an asset and can help investors make informed investment decisions.

3.3 Dividend Discount Model

The Dividend Discount Model (DDM) is a theoretical model used to value stocks based on the future stream of dividends that they are expected to pay. The DDM assumes that the value of a stock is equal to the present value of its future dividends. For that reason, dividends are often suggested as a viable equity premium predictor.

Under the DDM, the value of a stock can be calculated as

$$P= \left(D_1/(1+r)\right) + (D_2/(1+r)^2) + \cdots + (D_n/(1+r)^n), \tag{6}$$

where P is the value of the stock, $D_1, D_2, \dots, D_n$ are the dividends expected to be paid in each of the next n periods and r is the discount rate, which represents the opportunity cost of investing in the stock.

The DDM is based on the idea that investors value stocks based on the future stream of dividends that they expect to receive. The model assumes that the dividends will grow at a constant rate over time and that the discount rate used to value the dividends will remain constant.

One of the main benefits of the DDM is that it provides a straightforward way to value stocks based on the expected future stream of dividends. This can be useful for investors who are focused on income from dividends, as well as for investors who are considering the sustainability of a company's dividend payments over the long term.

3.4 Modern Portfolio Theory

Modern Portfolio Theory is a framework for constructing portfolios that seeks to maximize expected return for a given level of risk. It was developed by Harry Markowitz in 1952 and is considered a cornerstone of modern finance.

The theory states that investors should not simply invest in individual assets, but instead should create a well-diversified portfolio of assets. By combining different assets with different risk and return characteristics, an investor can reduce the overall risk of their portfolio while still achieving a higher expected return.

Markowitz proposed a mean-variance optimization model to determine the optimal portfolio, where the expected return is the mean of the portfolio's return distribution, and the risk is measured by the standard deviation (i.e., the volatility) of the portfolio's returns. The optimization process involves finding the weights for each asset in the portfolio that minimize the risk for a given level of expected return or maximize the expected return for a given level of risk.

The mean-variance optimization problem can be formalized as follows:

Maximize $$E(R_p) = \mu_p' w \tag{7}$$

subject to $$\begin{cases} w'\Sigma w \leq \sigma_p^2 \\ \sum w_i = 1 \\ w_i \geq 0 \end{cases}, \tag{8}$$

where w is a column vector of portfolio weights, μ_p is a column vector of expected returns for each asset in the portfolio Σ is covariance matrix of returns and σ_p^2 total portfolio variance. We can use the Lagrange method to solve this problem by introducing a Lagrange multiplier λ and forming the Lagrangian:

$$L(w, \lambda) = \mu_p' w - \lambda(w'\Sigma w - \sigma_p^2). \tag{9}$$

The solution is found by setting the gradient of the Lagrangian with respect to *w* and λ to zero:

$$\begin{cases} \nabla L(w, \lambda) = \mu_p - 2\lambda\Sigma w = 0, \\ w'\Sigma w - \sigma_p^2 = 0. \end{cases} \tag{10}$$

Solving for w and λ, we get:

$$w = (1/2\lambda)\Sigma^{-1}\mu_p. \tag{11}$$

Substituting this back into the second equation, we can solve for λ:

$$\lambda = (\mu_p'\Sigma^{-1}\mu_p)/(2\sigma_p^2). \tag{12}$$

The optimal portfolio weights are then given by:

$$w = (1/2\lambda)\Sigma^{-1}\mu_p = \left(1/2\sigma_p^2\right)\left(\Sigma^{-1}\mu_p\right)/\left(\mu_p'\Sigma^{-1}\mu_p\right). \tag{13}$$

Modern portfolio theory is implemented to equity premium forecasts when examining mean-variance investor who allocates their wealth according to optimal portfolio weights, constructed by implementing modern portfolio theory. Portfolio weights are conducted using the equity premium forecast.

3.5 Certainty Equivalent Return

Certainty equivalent return (CER) is a concept used in finance to represent the minimum return an investor would require from an investment to be indifferent between accepting the uncertain returns of the investment and the guaranteed return of an alternative investment. In other words, the CER represents the minimum return an investor would require to be compensated for the uncertainty associated with the investment.

The CER is based on the idea that investors are risk-averse and prefer certainty over uncertainty. For instance, an investor may prefer a guaranteed return of 5% to the uncertain returns of a stock with an expected return of 10% and a standard deviation of 20%. In this case, the CER of the stock is 5%. The CER can be thought of as the investor's reservation price for the investment below which the investment is not considered attractive. By determining the CER for each investment, an investor can make informed decisions about the trade-off between expected returns and risk. For instance, if the CER of a stock is higher than its expected return, the stock may not be considered attractive due to the high level of uncertainty associated with its returns.

Certainty equivalent return provides a useful tool for analyzing the trade-off between expected returns and risk in investments. The certainty equivalent return (CER) can be mathematically expressed as follows:

$$CER = E(R) - \frac{1}{2}\gamma\sigma^2(R), \tag{14}$$

where $E(R)$ is the expected return of investment on portfolio, $\sigma^2(R)$, is the variance of the returns, and γ is the risk aversion coefficient of the investor. The risk aversion coefficient is a measure of the investor's preference for certainty over uncertainty and represents the rate at which the investor is willing to trade-off expected return for reduced risk. If the investor is more risk-averse, the risk premium will be larger, resulting in a lower CER.

In the financial econometrics literature, CERs are used to determine the economic benefits (economic significance) produced by forecasting models. I present literature that examines both the economic and statistical predictability. Assuming an investor can invest in two different assets, risk-free interest rate or stock market, the investor uses their forecast to determine mean-variance sense optimal portfolio weights, from which the CER is calculated when the next period's returns are realized. These CERs are typically compared to a CER calculated using only the historical average as the forecasting model. (See e.g. Campbell & Thompson, 2008; Ferreira & Santa-Clara, 2011; Neely et al., 2014; Pan et al., 2020.)

4 Technical Analysis

In this section I introduce criticism on the EMH and arguments in favour of technical analysis to motivate the use of technical indicator. Section 4.1 is based on Malkiel (2003), and Welch (2017, 353-354).

4.1 Criticism on Efficient Market Hypothesis

If the efficient market hypothesis is accepted, it is impossible to consistently achieve returns that are higher than the market average by using information that is publicly available. In such a situation, future stock returns would be impossible to predict. However, despite its widespread acceptance, the EMH has been subject to criticism and there is a growing body of evidence that suggests that financial markets may not be as efficient as the theory suggests.

One of the main criticisms of the EMH is that it assumes that all market participants have access to the same information and that they all react to it in the same way. In reality, information is not equally distributed and there are often significant barriers to entry that prevent some market participants from accessing certain types of information. Furthermore, even when market participants have access to the same information, they may interpret it differently and react to it differently. This means that prices may not always reflect all available information, and that there may be opportunities for some market participants to consistently achieve returns that are higher than the market average.

Another criticism of the EMH is that it assumes that financial markets are rational. However, market participants are often influenced by a variety of psychological and emotional factors that can lead to irrational behaviour. For example, market participants may overreact to news events or be overly influenced by short-term events and ignore long-term trends. This can lead to price bubbles and market crashes, which would not occur in an efficient market.

The EMH assumes that financial markets are rational and therefore consists of only rational investors. In reality, market participants are not all rational but are often influenced by a variety of psychological and emotional factors that can lead to irrational behaviour. For example, market participants can overreact news or can put more weight on short-term events and ignore long-term trends.

EMH states that all market participants have access to the same information. In reality, some market participants have access to non-public information and can use it to their advantage. In this case, although expected with efficient markets, prices do not fully reflect all available information. In addition, this can result in some market participants consistently achieving higher returns than the average market returns, which would not be possible, if the markets were efficient.

4.2 Arguments in Favour of Technical Analysis

There are essentially four different types of theoretical models that support technical analysis, which I will outline below.

First type of theoretical model takes into account differences in the time it takes for investors to receive information. Treynor and Ferguson (1985) argue that technical analysis can be useful in determining whether information has been fully incorporated into equity prices under this friction, while Brown and Jennings (1989) demonstrate that past prices can help investors make better inferences about price signals. Additionally, Grundy and McNichols (1989) and Blume et al. (1994) show that trading volume can provide information beyond prices.

The next type of model proposes different responses to information by heterogeneous investors. Cespa and Vives (2012) show that asset prices can deviate from their fundamental values if there is uncertainty in asset residual payoff and/or persistence in liquidity trading. In such a setting, rational long-term investors follow trends. In the real world, different responses to information are more likely during recessions, when households experiencing job losses engage in consumption-smoothing asset sales and some investors liquidate margined assets. These factors help explain why technical indicators display enhanced predictive ability during recessions.

The third type of model considers underreaction and overreaction to information. Hong and Stein (1999) explain that at the start of a trend, investors underreact to news due to behavioural biases; as the market rises, they subsequently overreact, leading to even higher prices. Similarly, positive feedback traders - who buy (sell) after asset prices rise (fall) - can create price trends that technical indicators detect. Soros (2015) argues that positive feedback can alter firm fundamentals, justifying to some extent the price trends.

Edmans et al. (2015) shows that such feedback trading can occur in a rational model of investors with private information.

Finally, models of investor sentiment shed light on the efficacy of technical analysis. Researchers have analyzed how investor sentiment can drive asset prices away from their fundamental values since Keynes (1936). De Long et al. (1990) show that in the presence of limits to arbitrage, noise traders with irrational sentiment can cause prices to deviate from their fundamentals, even when informed traders recognize the mispricing. Baker and Wurgler (2006, 2007) find that measures of investor sentiment help to explain the cross-section of U.S. equity returns. Monthly sentiment-changes index from Baker and Wurgler (2007) is significantly and positively contemporaneously correlated with the realized equity risk premium, and that technical indicators significantly predict the sentiment-changes index, while macroeconomic predictors do not. Therefore, the differential information that technical indicators provide for predicting the equity risk premium appears to be related to their ability to anticipate changes in investor sentiment.

In summary, theoretical models based on information frictions help to explain the predictive value of technical indicators. Empirically, Moskowitz et al. (2012) recently found that pervasive price trends exist across commonly traded equity index, currency, commodity, and bond futures. Since the stock market is not a pure random walk and exhibits periodic trends, technical indicators should prove informative because they are primarily designed to detect trends.

5 Data and Literature Review

This paper examines the predictive power of macroeconomic predictors and technical indicators on the equity premium. Goyal and Welch (2008) argued that macroeconomic predictors cannot be used to produce forecasts that are accurate enough for anyone to benefit from them. These variables were widely accepted macroeconomic predictors as predictors of the equity premium already before the publication of Goyal and Welch (2008), after which the debate on the predictive ability of these variables has remained a topic of interest for financial econometricians. In this section, I delve into an extensive review of the existing literature and data pertinent to this topic, focusing on how these variables have been integrated into diverse econometric models. This examination encompasses the introduction of macroeconomic predictors delineated by Goyal and Welch (2008), and the technical indicators advocated by Neely et al. (2014), demonstrating their significant contributions to the field.

5.1 Data

Goyal and Welch (2008) use the following variables as predictors, the data frame of which Amit Goyal is still updating to make it more accessible and available to the public (Goyal, n.d.). I also use their data in my empirical phase.

The dependent variable, equity premium, is the stock return minus the risk-free rate. The equity premium can be seen as compensation for the risk an investor takes on over a risk-free investment, such as a government bond or another risk-free interest rate.

Stock Return: Goyal and Welch (2008) use continuously compounded S&P 500 month-end returns from the Center for Research in Security Press (CRSP) as their stock returns. This return series begins in 1927 and in their original article, Goyal and Welch (2008) end it in 2005. Their updated dataset extends all the way to the end of 2021. In their article, they also use wider frequencies such as quarterly and yearly, in addition to monthly. However, in my study, I focus mainly on monthly forecasts as they are the most meaningful from an investor's perspective. Therefore, the data at the center of my attention should also be monthly.

Risk-free rate: As the risk-free rate, Goyal and Welch (2008) use the Treasury-bill rate. Treasury-bill rates are highly liquid investments and are backed by the U.S. government, making them practically considered risk-free.

Figure 1 shows the monthly equity premiums of the S&P 500 stock market from January 1951 to December 2021.

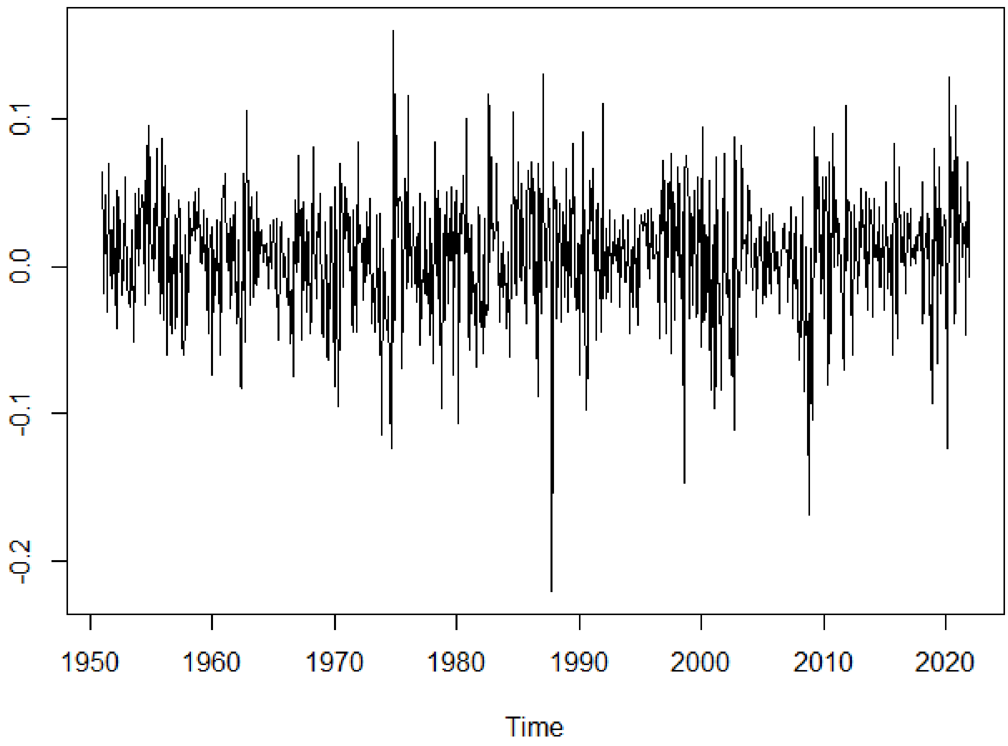

Figure 1. The monthly equity premium for the S&P 500 stock market index from January 1951 to December 2021.

In Figure 1, it can be observed that the equity premium has a very noisy nature during the observation period. The equity premium appears to fluctuate randomly around its historical mean, which is slightly above zero. In Figure 2, the risk-free interest rates (US Treasury Bill rates) are presented from January 1951 to December 2021.

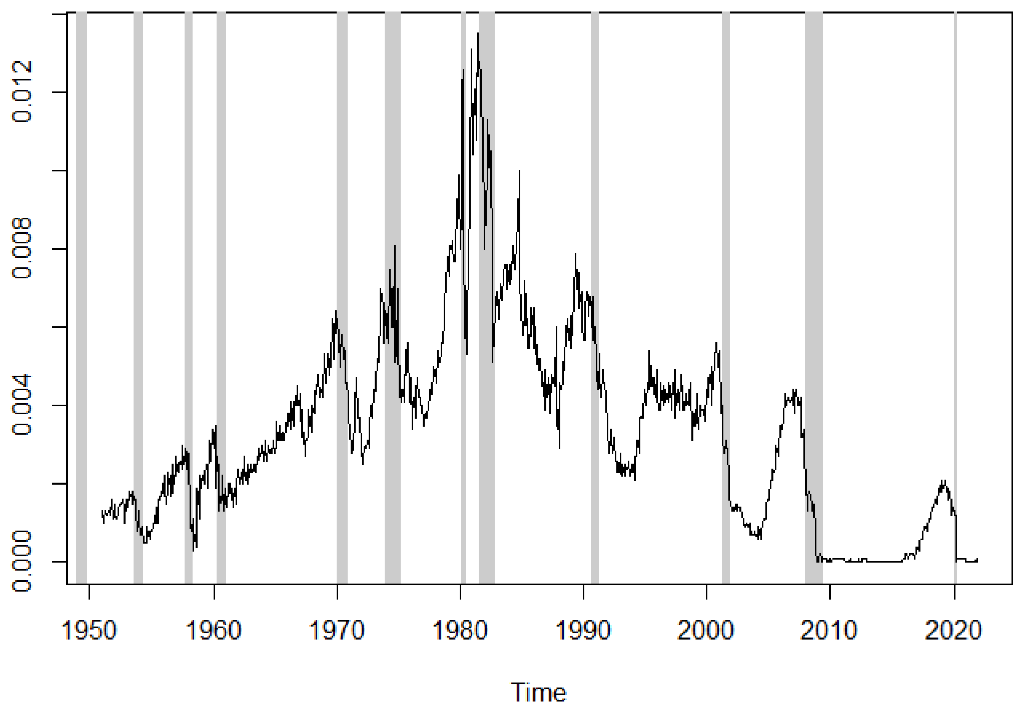

Figure 2. Risk-free interest rate from January 1951 to December 2021. The grey shades present the NBER's recession periods.

In Figure 2, grey shades represent the NBER's recession periods. Figure 2 shows that the risk-free interest rate has had an increasing trend up to the 1980s, after which the trend has been largely decreasing. It is also observed that during recession periods, interest rates have mostly dropped sharply, which is entirely reasonable as the demand for money decreases in recessionary conditions. In Figure 3, the cumulative S&P 500 equity premiums are presented from January 1951 to December 2021

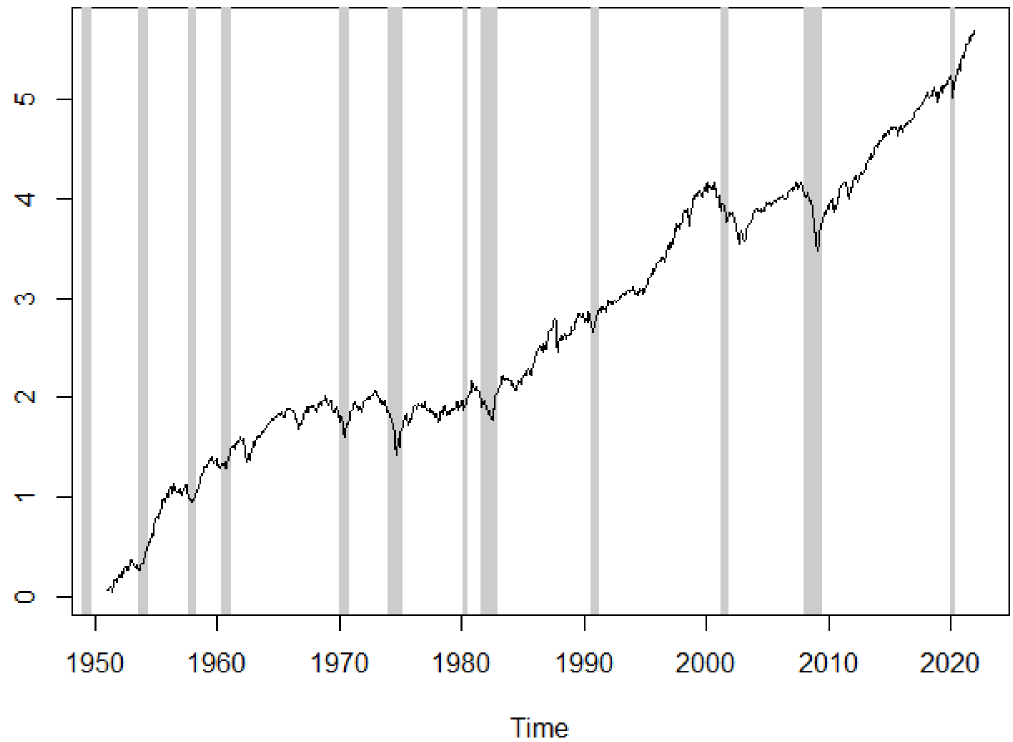

Figure 3. Cumulative S&P 500 equity premium displayed from January 1951 to December 2021. The grey shades present NBER's recession periods.

The grey shades in Figure 3 presents the NBER's recession periods. Now, the graph clearly shows the historically rising trend of the equity premium, although the 1970 energy crisis appears as a period of lower equity premiums. The decade of high stock returns after the financial crisis is also clearly visible. From the picture, it can be seen that equity premiums have been mainly negative during recession periods.

5.1.1 Technical indicators

The technical indicators used in this study correspond to the popular trend-following technical indicators employed by Neely et al. (2014). A total of 14 technical indicators are constructed, which are then combined and used with various forecasting models to predict the equity premium.

The first technical indicator is the Moving-Average ($\boldsymbol{MA}$) indicator, which implies a buy ($S_{i,t} = 1$) or sell ($S_{i,t} = 0$) signal when comparing the longer and shorter-term moving averages. Thus,

$$S_{i,t} = \begin{cases} 1, \text{if} MA_{s,t} \ {}^{3}MA_{l,t}, \\ 0, \ \text{if} \ MA_{s,t} \ < \ MA_{l,t}, \end{cases} \tag{15}$$

where

$$MA_{j,t} = \frac{1}{j}\sum_{i=0}^{j-1} P_{t-i}, \tag{16}$$

where $j = s, l,$ and P_t is the level of the S&P 500 index, and l is the length of long moving-average window and s is the length of short moving-average window. Hence, if the short moving average is higher than the long moving-average, the $\boldsymbol{MA(s, l)}$ indicator takes on a value of 1. Correspondingly, if the short moving-average is lower than the long moving average, the indicator takes on a value of 0. In this study, I examine three short periods (*s*=1, *s*=2, *s*=3) and two long periods (*l*=12, *l*=9). These choices correspond to the selections made by Neely et al. (2014). Thus, six binary variables predicting the equity premium are generated from the moving-average indicators.

Momentum is a well-studied phenomenon in asset pricing, and Neely et al. (2014) specify it such that if the stock market index (S&P 500) is higher at time *t* than at time *t-m*, the momentum indicator takes a value of 1, indicating relatively high expected return and a buy signal. Otherwise, it takes a value of 0, which, like the moving-average case, indicates a sell signal. Mathematically expressed

$$S_{i,t} = \begin{cases} 1, \ \text{if} \ P_t \geq \ P_{t-m}, \\ 0, \ \text{if} \ P_t \ < \ P_{t-m}. \end{cases} \tag{17}$$

Momentum indicator is denoted by ***MOM(m)***, where *m* is the length of the "look-back". Following Neely et al. (2014) I consider three and four quarter look-backs, i.e. ***MOM(9)*** and ***MOM(12)***.

The final indicator takes into account trading volume along with previous prices. Following Neely et al. (2014) again, I define

$$OBV_t = \sum_{i=1}^{t} VOL_i D_i, \tag{18}$$

where VOL_i is the trading volume[3] on period i and D_i is defined as follows:

[3] Trading volume data up until December 2011 has been downloaded from the website of David E. Rapach (Rapach, 2014). For the period from January 2012 onwards, the data has been gathered from Yahoo Finance (Yahoo Finance, n.d.).

$$D_i = \begin{cases} 1, & \text{if } P_i - P_{i-1} \geq 0, \\ -1, & \text{if } P_i - P_{i-1} < 0. \end{cases} \tag{19}$$

To form the trading signal, following Neely et al (2014), I compute the moving-average of OBV_t as follows

$$MA_{j,t}^{OBV} = \frac{1}{j}\sum_{i=0}^{j-1} OBV_{t-i}. \tag{20}$$

Now the trading signals for volume-based indicator, denoted by ***VOL(s,l)*** is constructed as

$$S_{i,t}^{OBV} = \begin{cases} 1, \text{if} MA_{s,t}^{OBV} \, {}^{3} MA_{l,t}^{OBV}, \\ 0, \ \text{if } MA_{s,t}^{OBV} < MA_{l,t}^{OBV}. \end{cases} \tag{21}$$

Consistent with the moving-average indicator, I examine the indicator with the choices *s=1, s=2, s=3 and l=12, l=9.*

5.1.2 Macroeconomic predictors

I introduce all the macroeconomic predictors employed by Goyal and Welch (2008), which encompass both stock market-based and macroeconomic-based metrics.

In their study, as well as in other literature examined in this thesis, dividends are composed of the 12-month trailing sum of S&P 500 index dividends. Data prior to 1987 is sourced from Robert Shiller's website, while data from 1988 to 2005 is sourced from S&P Corporation. Two commonly used dividend derivatives are employed in the studies: The Dividend Price Ratio (***DP***), which is the difference between the natural logarithm of dividends and the natural logarithm of prices, and the Dividend Yield (***DY***), which is the difference between the logarithm of dividends and the logarithm of lagged prices. (See, e.g., Ball, 1978; Campbell & Shiller, 1988b; Campbell & Viceira, 2002; Campbell & Yogo, 2006; Fama & French, 1988, Cochrane, 1997; Hodrick, 1992; Lewellen, 2004; Menzly, Santos & Veronesi, 2004; Rozeff, 1984; Shiller, 1984.)

Earnings are the twelve-month trailing sum of S&P 500 index earnings. Similarly, data prior to 1987 is sourced from Robert Shiller's website, while earnings from 1988 onwards are Goyal and Welch's own estimates based on quarterly earnings interpolation. The used variables are the Earnings Price Ratio (***EP***), which is the difference between the logarithm of earnings and the logarithm of prices, and the Dividend Payout Ratio, which is the difference between the logarithm of dividends and the logarithm of earnings (***DE***). (See, e.g., Campbell & Shiller, 1988a, 1988b; Lamont, 1998.)

The Stock Variance (***SVAR***) is the sum of squared daily returns of the S&P 500 (Guo, 2006). Goyal and Welch sourced the data from CRSP.

In the studies, the relative valuation of stocks with high and low CAPM-implied betas, known as the cross-sectional beta premium (***CSP***), is also used as an explanatory variable. This is proposed by Polk et al. (2006). Goyal and Welch obtained this data from Samuel Thompson, and it consists of observations starting from March 1937.

As book value, Goyal and Welch (2008) use book market values from Value Line's website (Long-Term Perspective Chart of the Dow Jones Industrial Average). As a predictor variable they use widely known Book-to-Market Ratio (***BM***), which is the ratio of book value to market value for the Dow Jones Industrial Average. For the months of March through December, the book value at the end of the previous year is divided by the price at the end of the current month. For the months of January and February, the book value at the end of two years ago is divided by the price at the end of the current month. (See, e.g, Kothari & Shanken, 1997; Pontiff & Schall, 1998.)

Goyal and Welch (2008) use two measures of corporate issuing activity. The first, Net Equity Expansion (***NTIS***) is the ratio of 12-month sums of net issues by NYSE listed stocks divided by the total end-of-year market capitalization of NYSE stock. Net equity issuing activity, consisting of IPOs, SEOs, stock repurchases, less dividends, is calculated as

$$Net\ issue_t = Mcap_t - Mcap_{t-1}(1 + vwretx_t), \tag{22}$$

where $Mcap$ is the total market capitalization, and $vwretx$ is the value weighted return excluding dividends on the NYSE index. The second measure, Percent Equity Issuing (***EQIS***) proposed by Baker and Wurgler (2000), is calculated as equity issuing activity divided by total issuing activity. Authors provided this data to Goyal and Welch, expect for 2005, which they added themselves.

Goyal and Welch use Treasury-bill rates as predictive variables. Rates from 1920 to 1933, the Treasury-bill (***TBL***) rates are sourced from the NBER Macrohistory database, specifically the U.S. Yields On Short-Term United States Securities, Three-Six Month

Treasury Notes and Certificates, Three Month Treasury series. From 1934 to 2005, the rates are obtained from the Federal Reserve Bank at St. Louis' economic research data base and specifically the 3-Month Treasury Bill: Secondary Market Rate. (See, e.g., Campbell, 1987; Hodrick, 1992.)

Goyal and Welch (2008) use three Long Term Yield based variable as predictors. They use U.S. Yield on Long-Term United States Bonds series in the NBER's Macrohistory database as their long-term government bond (***LTY***). They got the data from Ibbotson's Stocks, Bonds, Bills and Inflation Yearbook. The same source provided them with the Long-Term Rate of Returns (***LTR***). Their third Long Term Yield based variable is the Term Spread (***TMS***) which is the difference between long term yield on government bonds and the Treasury-bill. (See, e.g., Campbell, 1987; Fama & French, 1989.)

Goyal and Welch (2008) obtain long-term corporate bond returns from the same source, Ibbotson's Stocks, Bonds, Bills and Inflation Yearbook. As a predictive variable, they use the Default Return Spread (***DFR***) which is the difference between long-term corporate bond and long-term government bond returns. They get Corporate Bond Yields on AAA and BAA-rated bonds from FRED and use their difference (the return on BAA bonds minus the return on AAA bonds), the Default Yield Spread (***DFY***), as a predictive variable. (See, e.g., Keim & Stambaugh, 1986; Fama & French, 1989.)

As inflation (***INFL***), they use the Consumer Price Index for All Urban Consumers, which is retrieved from the Bureau of Labor Statistics. Because monthly inflation is released only in the following month, Goyal and Welch use it as a lagged predictor. (See, e.g., Lintner, 1975; Fama & Schwert, 1977; Fama, 1981; Campbell & Vuolteenaho, 2004.)

Second to last, they use Investment to Capital ratio (***IK***) suggested by Cochrane (1991), who also provided them with the data, as a predictor variable. The investment capital is aggregate (private non-residential fixed) investment divided by aggregate capital for the whole economy.

In addition to the variables presented above, Goyal and Welch (2008) use variable that measures the ratio of consumption, wealth, and income (***cay***). The variable is suggested by Lettau and Ludvigson (2001). Goyal and Welch (2008) modify the variable in a way that it does not use look-ahead data. They estimate aggregate consumption as follows:

$$c_t = \alpha + \beta_a a_t + \beta_y y_t + \sum_{i=-k}^{k} b_{a,i}\, \Delta a_{t-i} + \sum_{i=-k}^{k} b_{y,i}\, \Delta y_{t-i} + \varepsilon_t, \quad (23)$$

where y is the aggregate income, a is the aggregate wealth and $t = k + 1, \ldots, s - k$.

Furthermore, $\boldsymbol{cay}$ is estimated using equation

$$\widehat{caypt}_t = c_t - \widehat{\beta_a} a_t - \widehat{\beta_y} y_t, \tag{24}$$

where $t = k + 1, \ldots, T - k$.

5.1.3 Descriptive statistics

In my analysis forecasting models are examined only in an out-of-sample sense (see Section 6.1.1), where forecasting is carried out using the expanding window approach. At each step, only the next period (t+1) is predicted. The entire dataset consists of a total of 852 monthly observations, starting in January 1951 and ending in December 2021. In the expanding window approach (see Section 6.1.2), the first in-sample (estimation) sample consists of observations from the beginning of the dataset to January 1965, thus comprising 169 observations. The following 9 years, or 108 observations, serve as the so-called holdout[4] out-of-sample, during which expanding window forecasting is carried out, but which is not used for forecast evaluation. The remaining part of the dataset, from January 1975 to December 2021, is the forecast evaluation period, which consists of 575 monthly observations.

[4] Following Sock & Watson (2004), and Rapach & Zhou (2013) the holdout out-of-sample is necessary for calculating DMSFE, which is required for computing the optimally weighted forecast combination.

Before proceeding with the empirical forecasting phase, it is prudent to take a closer look at the variables employed in this study. Table 1 presents the number of buy and sell signals for different technical indicators.

Table 1. The amount of buy and sell signals for each technical indicator

Indicator	Buy signals	Sell signals
MA(1 9)	413	162
MA(1,12)	431	144
MA(2 9)	419	156
MA(2,12)	429	146
MA(3 9)	422	153
MA(3 12)	430	145
MOM(9)	430	145
MOM(12)	439	136
VOL(1,9)	408	167
VOL(1,12)	419	156
VOL(2,9)	408	167
VOL(2,12)	420	155
VOL(3,9)	412	163
VOL(3,12)	416	159

From Table 1, it can be observed that the number of buy and sell signals during the forecast evaluation period for different technical indicators are quite similar. The ***MOM(12)*** indicator generates the most buy signals, with a total of 439 buy signals. Conversely, the most sell signals are implied by the ***VOL(2, 9)*** and ***VOL(1, 9)*** indicators.

Table 2 further examines the characterization of technical indicators by presenting their (sample) correlation matrix.

Table 2. Correlation matrix of technical indicators

	MA(1,9)	***MA(1,12)***	***MA(2,9)***	***MA(2,12)***	***MA(3,9)***	***MA(3,12)***	***MOM(9)***	***MOM(12)***	***VOL(1,9)***	***VOL(1,12)***	***VOL(2,9)***	***VOL(2,12)***	***VOL(3,9)***	***VOL(3,12)***
MA(1 9)	1													
MA(1,12)	0.86	1												
MA(2 9)	0.82	0.84	1											
MA(2,12)	0.78	0.89	0.88	1										
MA(3 9)	0.74	0.80	0.88	0.87	1									
MA(3 12)	0.69	0.79	0.81	0.89	0.87	1								
MOM(9)	0.74	0.85	0.76	0.83	0.75	0.83	1							
MOM(12)	0.61	0.74	0.66	0.76	0.67	0.76	0.78	1						
VOL(1,9)	0.64	0.62	0.60	0.58	0.54	0.51	0.56	0.47	1					
VOL(1,12)	0.63	0.64	0.62	0.64	0.58	0.60	0.61	0.54	0.87	1				
VOL(2,9)	0.62	0.62	0.65	0.65	0.61	0.60	0.59	0.50	0.78	0.84	1			
VOL(2,12)	0.61	0.63	0.62	0.65	0.61	0.64	0.63	0.56	0.75	0.85	0.86	1		
VOL(3,9)	0.58	0.60	0.62	0.63	0.63	0.65	0.59	0.51	0.69	0.76	0.83	0.85	1	
VOL(3,12)	0.58	0.63	0.63	0.67	0.64	0.68	0.65	0.58	0.68	0.78	0.80	0.91	0.87	1

From Table 2, it can be observed that there is generally a high positive correlation between the technical indicators. Moving Average (MA) indicators with different combinations of short and long windows show a strong positive correlation with each other, ranging from 0.69 to 0.89. Momentum (MOM) indicators with 9 and 12-month windows also show strong positive correlation with each other (0.78). Volatility (VOL) indicators with different combinations of short and long windows show strong positive correlation with each other, ranging from 0.68 to 0.91. The correlation between differently constructed indicators is also strong. None of the correlations are below 0.50, and all are positive. Based on the correlation matrix of the technical indicators, one could assume that the equity premium forecasts formed using these indicators are quite similar.

In Figure 4, the time series of the ***MA(2, 12)*** indicator is presented for the forecast evaluation period.

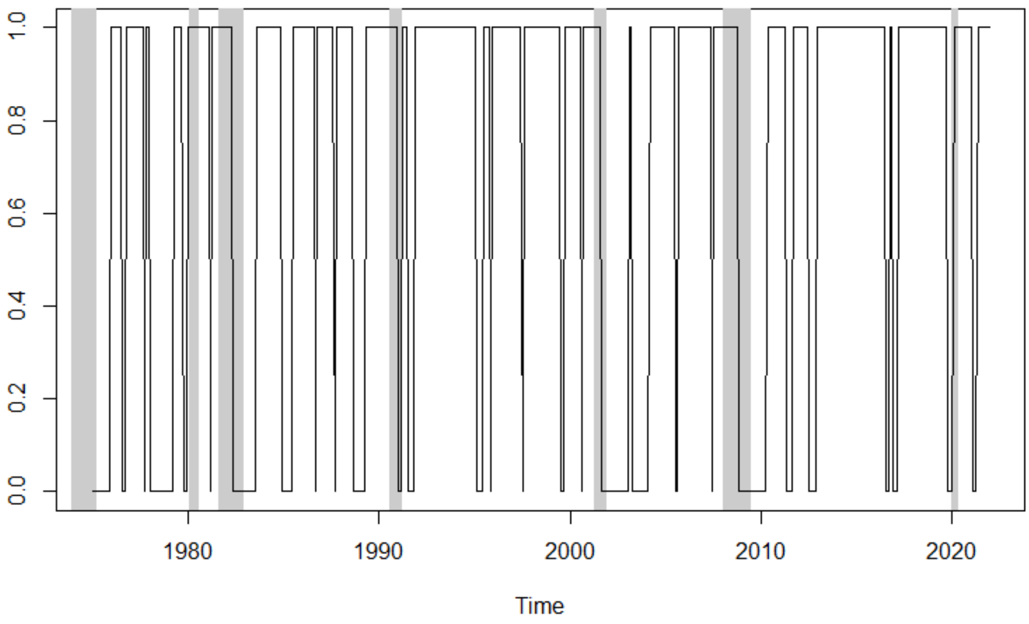

Figure 4. ***MA(2, 12)*** technical indicator. Forecast evaluation period. Grey shades presents the NBER recession periods.

In Figure 4, the behaviour of the technical indicator is illustrated using the ***MA(2, 12)*** indicator. The indicator implies a buy signal for the majority of the time when its value is 1. When it implies a sell signal, its value is 0. It can be observed that the indicator often implies a sell signal at the beginning of recession periods.

Table 3 presents the descriptive statistics of the macroeconomic predictors used in the empirical analysis, including sample mean (Mean), standard deviation (SD), minimum (Min), maximum (Max), skewness (Skew.), kurtosis (Kurt.), autocorrelation (lagged by one period) (Auto. cor.), and the sample size (N).

Table 3. Descriptive statistics. Macroeconomic predictors. December 1951 to December 2021

The table presents the descriptive statistics for 14 macroeconomic predictors and the dependent variable (equity premium). The equity premium in the table is represented as its logarithmic transformation in percentage terms. Additionally, SVAR, TBL, LTY, LTR, TMS, DFY, DFR, and CPI are also expressed in percentages.

	Mean	**SD**	**Min**	**Max**	**Skew.**	**Kurt.**	**Auto. cor.**
y	0.58	4.20	-24.76	14.73	-0.66	2.31	0.04
DP	-3.56	0.42	-4.52	-2.60	-0.12	-0.77	0.99
DY	-3.55	0.42	-4.53	-2.61	-0.12	-0.75	0.99
EP	-2.82	0.42	-4.84	-1.90	-0.62	2.44	0.99
DE	-0.73	0.29	-1.24	1.38	2.60	16.02	0.99
SVAR	0.20	0.47	0.01	7.32	10.92	146.00	0.40
BM	0.50	0.25	0.12	1.21	0.65	-0.33	0.99
NTIS	0.01	0.02	-0.06	0.05	-0.77	0.23	0.98
TBL	4.10	3.10	0.01	16.30	0.91	1.10	0.99
LTY	5.76	2.85	0.62	14.82	0.76	0.16	0.99
LTR	0.53	2.77	-11.24	15.23	0.50	2.96	0.05
TMS	1.66	1.38	-3.65	4.55	-0.06	-0.12	0.96
DFY	0.96	0.43	0.32	3.38	1.85	4.93	0.97
DFR	0.03	1.44	-9.76	7.37	-0.66	7.85	-0.06
INFL	0.29	0.36	-1.92	1.81	0.13	2.62	0.55

As illustrated in Table 3, equity premium (***y***) has an average value of 0.58 and a standard deviation of 4.20. The distribution of the equity premium exhibits a negative skewness of -0.66 and a kurtosis of 2.31. The autocorrelation of 0.04 reveals weak serial dependence in the data.

The Dividend-Price Ratio (***DP***) and Dividend Yield (***DY***) have similar characteristics, with means of -3.56 and -3.55, and standard deviations of 0.42 for both variables. The distributions of ***DP*** and ***DY*** are nearly symmetric, as evidenced by their skewness values of -0.12. The high autocorrelation of 0.99 for both variables denote a strong persistence in the series over time. The Earnings-Price Ratio (***EP***) demonstrates a mean of -2.82, a standard deviation of 0.42, and a negatively skewed distribution with a skewness of -0.62.

Similar to ***DP*** and ***DY***, ***EP*** also exhibits a high autocorrelation of 0.99. The Dividend-Earnings Ratio (***DE***) has a mean of -0.73, a standard deviation of 0.29, and a range from -1.24 to 1.38. The distribution of ***DE*** is positively skewed with a skewness of 2.60 and a high kurtosis of 16.02. The autocorrelation of 0.99 indicates strong persistence in the series.

The Stock Variance (***SVAR***) has a mean of 0.20 and standard deviation of 0.47. and a range from 0.01 to 7.32. It has a highly positive skewness of 10.92 and a high kurtosis of 146.00. Note that variance is always non-negative by definition. The autocorrelation for ***SVAR*** is 0.40, which is relatively lower than other variables in the dataset.

The Book-to-Market Ratio (***BM***) has a mean value of 0.50 and a standard deviation of 0.25, indicating a moderately dispersed distribution. The range of values for ***BM*** spans from 0.12 to 1.21, with a skewness of 0.65 and a kurtosis of -0.33. The autocorrelation of BM is notably high at 0.99, indicating strong persistence in the series.

Net Equity Expansion (***NTIS***) presents a mean value of 0.01 and a standard deviation of 0.02. The distribution ranges between -0.06 and 0.05 and is characterized by a negative skewness of -0.77 and a kurtosis of 0.23. The autocorrelation for ***NTIS*** is high, with a value of 0.98.

The Long-Term Yield (***LTY***) has a mean of 5.76 and a standard deviation of 2.85. Its values range from 0.62 to 14.82. ***LTY*** exhibits a high autocorrelation of 0.99. The Long-Term Return (***LTR***) features a mean of 0.53 and a standard deviation of 2.77. The distribution ranges from -11.24 to 15.23. The autocorrelation for ***LTR*** is relatively low at 0.05, implying weak serial dependence in the data. The Term Spread (***TMS***) has a mean of 1.66, a standard deviation of 1.38, and values ranging from -3.65 to 4.55. The autocorrelation for ***TMS*** is high at 0.96, indicating a strong persistence in the series.

The Default Yield Spread (***DFY***) has a mean of 0.96 and a standard deviation of 0.43, with values ranging from 0.32 to 3.38. The autocorrelation for ***DFY*** is 0.97, suggesting strong persistence in the series. The Default Return Spread (***DFR***) has a mean of 0.03 and a standard deviation of 1.44. The autocorrelation for DFR is -0.06, revealing weak negative serial dependence in the data.

Finally, the Inflation Rate (***INFL***) has a mean value of 0.29 and a standard deviation of 0.36. The autocorrelation for ***INFL*** is 0.55, denoting moderate persistence in the series.

Table 4 presents the descriptive statistics of the same macroeconomic predictors and equity premium (log) as previously discussed, focusing on the mean and standard deviation across three different samples: 1951–2021, 1951–2011, and 2012–2021. The sample from 1951 to 2011 is the sample Neely et al. (2014) used in their study.

Table 4. Mean and standard deviation of macroeconomic predictors and the (log) equity premium across three different samples

	Mean			**Std. Dev.**		
	1951-2021	1951-2011	2012-2021	1951-2021	1951-2011	2012-2021
y	0.58	0.47	1.24	4.20	4.26	3.77
DP	-3.56	-3.49	-3.96	0.42	0.42	0.13
DY	-3.55	-3.49	-3.95	0.42	0.42	0.12
EP	-2.82	-2.78	-3.09	0.42	0.44	0.20
DE	-0.73	-0.71	-0.87	0.29	0.30	0.17
SVAR	0.20	0.20	0.22	0.47	0.42	0.68
BM	0.50	0.54	0.29	0.25	0.25	0.05
NTIS	0.01	0.02	-0.01	0.02	0.02	0.01
TBL	4.10	4.68	0.59	3.10	2.95	0.79
LTY	5.76	6.32	2.36	2.85	2.67	0.75
LTR	0.53	0.55	0.37	2.77	2.76	2.85
TMS	1.66	1.64	1.77	1.38	1.43	1.01
DFY	0.96	0.96	0.94	0.43	0.45	0.24
DFR	0.03	0.01	0.17	1.44	1.39	1.76
INFL	0.29	0.30	0.17	0.36	0.36	0.32

From the Table 4, it can be observed that the Equity Premium (log) exhibits an increase in the mean value from 0.47 during 1951-2011 to 1.24 in the more recent period of 2012-2021. The overall mean value for the entire sample (1951-2021) is 0.58. The standard deviation also shows a decrease from 4.26 in the first period to 3.77 in the second period, while the overall standard deviation is 4.20.

For the ***DP*** variable, the mean value decreases from -3.49 during 1951-2011 to -3.96 in 2012-2021, while the standard deviation decreases from 0.42 to 0.13 in the same periods. Similarly, the ***DY*** variable exhibits a decrease in the mean value from -3.49 in 1951-2011 to -3.95 in 2012-2021, with a corresponding decrease in standard deviation from 0.42 to 0.12. The mean value of ***EP*** decreases slightly from -2.78 in 1951-2011 to -3.09 in 2012-2021, and the standard deviation increases marginally from 0.44 to 0.20 over the same periods. This may indicate minor changes in the earnings-price ratios during the recent period. The mean of ***DE*** decreases from -0.71 in 1951-2011 to -0.87 in 2012-2021, and the standard deviation decreases from 0.30 to 0.17. These modest changes could reflect variations in the debt-equity ratios among firms in different periods.

The mean of ***BM*** value declines from 0.54 in 1951-2011 to 0.29 in 2012-2021, and the standard deviation remains stable at 0.25. This indicates a decrease in the average book-to-market ratio in the recent period.

The mean of ***TBL*** decreases notably from 4.68 in 1951-2011 to 0.59 in 2012-2021, and the standard deviation decreases from 2.95 to 0.79. This indicates a substantial decline in Treasury Bill rates during the recent period, possibly reflecting changes in monetary policy. The mean of ***LTY*** decreases from 6.32 in 1951-2011 to 2.36 in 2012-2021, and the standard deviation decreases from 2.67 to 0.75. This suggests a significant decline in long-term government bond yields during the recent period, potentially due to changes in monetary policy or economic conditions. Similarly, the mean of ***LTR*** decreases slightly from 0.55 in 1951-2011 to 0.37 in 2012-2021, and the standard deviation remains stable at 2.77. This implies a minor change in long-term government bond returns during the recent period. The mean of ***TMS*** increases marginally from 1.64 in 1951-2011 to 1.77 in 2012-2021, and the standard deviation decreases from 1.43 to 1.01. This indicates a slight increase in the term spread during the recent period. The mean of ***DFY*** remains stable at 0.96 in 1951-2011 and 0.94 in 2012-2021, while the standard deviation decreases from 0.45 to 0.24. This suggests a modest change in default yield spread during the recent period. Furthermore, the mean of ***DFR*** increases from 0.01 in 1951-2011 to 0.17 in 2012-2021, and the standard deviation increases from 1.39 to 1.76.

The mean of ***INFL*** decreases slightly from 0.30 in 1951-2011 to 0.17 in 2012-2021, while the standard deviation decreases marginally from 0.36 to 0.32. This suggests a modest decline in consumer price inflation during the recent period, potentially due to changes in monetary policy or other macroeconomic factors. Table 5 presents the correlations between different macroeconomic predictors and the equity premium.

Table 5. Correlation coefficients between the equity premium and the macroeconomic predictors.

	y	***DP***	***DY***	***EP***	***DE***	***SVAR***	***BM***	***NTIS***	***TBL***	***LTY***	***LTR***	***TMS***	***DFY***	***DFR***	***INFL***
y	1														
DP	-0.06	1													
DY	0.04	0.99	1												
EP	-0.06	0.77	0.76	1											
DE	0.01	0.33	0.33	-0.35	1										
SVAR	-0.32	-0.07	-0.10	-0.16	0.14	1									
BM	-0.07	0.89	0.89	0.82	0.10	-0.09	1								
NTIS	-0.03	0.29	0.28	0.24	0.07	-0.20	0.31	1							
TBL	-0.11	0.45	0.45	0.53	-0.12	-0.06	0.58	0.11	1						
LTY	-0.08	0.40	0.39	0.43	-0.05	-0.02	0.51	0.07	0.90	1					
LTR	0.05	0.00	0.01	0.02	-0.03	0.16	0.00	-0.07	0.04	0.04	1				
TMS	0.07	-0.20	-0.19	-0.30	0.16	0.10	-0.25	-0.09	-0.40	0.05	-0.01	1			
DFY	0.00	0.19	0.19	0.03	0.23	0.29	0.29	-0.35	0.32	0.47	0.14	0.25	1		
DFR	0.27	-0.01	0.02	-0.08	0.11	-0.23	-0.01	0.01	-0.05	-0.01	-0.46	0.08	0.07	1	
INFL	-0.08	0.23	0.23	0.34	-0.17	-0.07	0.37	0.10	0.44	0.37	-0.05	-0.21	0.08	-0.05	1

From Table 5, it can be observed that ***DP***, ***DY***, and ***EP*** display a high positive correlation with each other, indicating that these variables tend to move in the same direction. The correlation between ***DP*** and ***DY*** is almost 1, which can be explained by the fact that these variables are theoretically very similar. This high correlation indicates that they share common information and trends, which makes them nearly interchangeable in the context of predicting the equity premium. Their correlations with the equity premium are weak and mostly negative, suggesting that they might not be strong predictors of the equity premium. ***DE***, on the other hand, has a weak positive correlation with the equity premium and a moderate positive correlation with ***DP*** and ***DY***, while having a negative correlation with ***EP***.

SVAR demonstrates a moderate negative correlation with the equity premium, indicating that it might have some predictive power. It also has weak negative correlations with ***DP*** and ***DY***, and a weak positive correlation with ***DE***. The relationship between ***SVAR*** and other variables appears to be less straightforward.

BM exhibits strong positive correlations with ***DP*** and ***DY***, and a moderately positive correlation with ***EP***, implying that it shares common trends with these variables. Its correlation with the equity premium is weakly negative, suggesting limited predictive capacity.

NTIS shows weak positive correlations with ***DP***, ***DY***, and ***EP***, and a weak negative correlation with the equity premium. Its relationship with other variables appears to be relatively weak and may not be particularly informative for predicting the equity premium.

TBL and ***LTY*** display moderate positive correlations with ***DP***, ***DY***, and ***EP***, and weak negative correlations with the equity premium. They also have a strong positive correlation with each other. These patterns suggest that these variables might share similar trends, but their predictive power for the equity premium appears to be limited. ***LTR***'s correlations with the majority of the other variables are weak or close to zero, implying that its relationship with these variables is relatively independent. Its weak positive correlation with the equity premium suggests that it may not have a strong predictive capacity. ***TMS*** has weak to moderate negative correlations with ***DP***, ***DY***, and ***EP***, and a weak positive correlation with the equity premium. Its correlations with other variables are mostly weak, but its negative correlations with some of the variables may indicate a potential for predicting the equity premium. ***DFY*** displays weak to moderate positive correlations with several variables, but its relationship with the equity premium is very weak. Its correlations suggest a potential connection with the other variables, but its predictive power for the equity premium seems limited. ***DFR*** has a moderate positive correlation with the equity premium, indicating that it might have some predictive power. It also has weak correlations with most of the other variables, suggesting that its relationship with these variables is relatively independent.

Lastly, ***INFL*** has weak negative correlations with the equity premium, ***DP***, and ***DY***, and a moderate positive correlation with ***EP***. Its correlations with the other variables are generally weak, which might indicate that its predictive capacity for the equity premium is limited.

While some variables display moderate to strong correlations with each other, their correlations with the equity premium are mostly weak, suggesting that their predictive power for the equity premium may be limited.

5.2 Literature on Equity Premium Forecasting

In their famous article, Goyal and Welch (2008) carry out a rigorous examination of the empirical performance of equity premium prediction models. Their research aims to assess the out-of-sample predictive power of numerous variables that have been proposed in the literature.

Goyal and Welch (2008) utilize multiple regression models to estimate the equity premium based on the lagged values of the predictive variables. They evaluate the forecasting performance of these models using various statistical metrics, including mean squared forecast error (MSFE, see Section 6.1.3 for more on MSFE) and out-of-sample R-squared (R2). From this point forward denoted as R2. Their out-of-sample R2 is defined as

$$R2 = 1 - \frac{\text{MSFE}_N}{\text{MSFE}_{HA}}, \tag{25}$$

where MSFE_{HA} denotes the mean squared forecast error obtained using the naïve model, i.e. the historical average forecast and MSFE_N denotes the mean squared forecast error obtained using one of their own predictive models. This specification is used throughout this study and is predominantly employed in the other research articles I present as well.

In their study, Goyal and Welch utilize single-variable linear prediction models. They use each of the variables introduced in Section 5.1 individually as predictors, and in addition, they employ a multivariate regression model that includes all the variables from Section 5.1 as predictive variables. They refer to this model as the 'Kitchen Sink' model (as will I in this thesis).

The main findings of their study are quite sobering. Most of the proposed models exhibit limited success in predicting equity premiums, with many of them unable to outperform the historical average. Furthermore, the authors find no robust evidence to support the use of any specific predictive variable. They also show that the predictive power of these models declines significantly when accounting for data snooping and small-sample biases.

These results have important implications for both academics and practitioners, suggesting that the existing literature on equity premium prediction may have overstated the true forecasting capabilities of these models. Goyal and Welch's research not only challenges the effectiveness of widely used equity premium prediction models but also underscores the need for further investigation into developing more robust forecasting techniques in finance. Their work serves as a cautionary note for those who rely on these models for predicting equity premiums and making investment decisions, emphasizing the importance of considering potential biases and overfitting when evaluating model performance.

Campbell and Thompson (2008) delve into the out-of-sample performance of various models for predicting excess stock returns, aiming to determine if any model can outperform the simple historical average forecast. The authors focus on models incorporating predictor variables which are similar to the variables considered in the Goyal and Welch (2008) study. Campbell and Thompson (2008) impose economically sensible restrictions on the coefficients and the prediction. The constraint on the regression coefficients means that they allow the coefficients to take values that matches their theoretically expected sign (historical) and the constraint on the prediction is implemented so that the forecast cannot be negative. From this point forward, this constraint is referred to as the CT restriction. This approach differs from that of Goyal and Welch, who do not impose any restrictions on regression coefficients or the prediction. The main finding of Campbell and Thompson (2008) study is that when these economically sensible restrictions are imposed, some models can indeed outperform the historical average in predicting excess stock returns out of sample. Notably, the dividend-price ratio and the term spread are identified as relevant predictors. These results suggest that, under certain conditions, carefully specified forecasting models with economically sensible restrictions can provide superior out-of-sample performance compared to the historical average.

Pettenuzzo et al. (2014) also examine the impact of economic constraints on the forecasting of stock returns. The authors propose a new methodology that takes into account these constraints, arguing that the traditional approach to stock return forecasting may result in biased forecasts due to the presence of economic constraints. The authors employ a Bayesian framework and incorporate economic constraints through a shrinkage-like estimator. They consider non-negative equity premium and conditional Sharpe ratio[5] restrictions. They test their methodology using the same variables as Goyal and Welch (2008). Their findings suggest that the proposed methodology yields significant improvements in out-of-sample forecasting performance when compared to the traditional unconstrained approach.

Pan et al. (2020) propose a predictor-constrained approach to forecasting stock returns. The authors address the issue of estimation uncertainty in models that use a large number of predictors, which can lead to poor out-of-sample performance. By applying constraints on the predictors, they aim to reduce estimation uncertainty and improve forecast accuracy. They evaluate the out-of-sample performance of their predictor-constrained approach using U.S. stock return data and compare it to other commonly used forecasting methods. They use the same predictors as Goyal and Welch (2008) once again, but implement a nonlinear change to the predictors, which can be represented as follows. Let x_{ti} be the value of predictor i at time t. They modify the predictive variable as follows

$$\overline{x_{ti}} = \begin{cases} x_{ti}, & \text{if } \; x_{ti} > \max\left(x_{(t-1),i}, x_{(t-2),i}, \dots, x_{(t-b),i}\right) \; or \; x_{ti} < \min\left(x_{(t-1),i}, x_{(t-2),i}, \dots, x_{(t-b),i}\right) \\ 0 & \text{otherwise,} \end{cases} \tag{26}$$

[5] The Sharpe ratio is a widely used financial metric that measures the risk-adjusted performance of an investment. It is calculated by dividing the equity premium by the market returns standard deviation. A higher Sharpe ratio indicates a better risk-adjusted performance, as it means that the investment is generating more return per unit of risk. Conversely, a lower Sharpe ratio suggests that the investment is not as efficient in terms of risk-adjusted performance. (Cochrane, 2005, 20-21.)

where b is the "look-back" period. They use these truncated variables in univariate regression models, where they predict the equity premium. They also consider forecast combinations, suggested by Rapach et al. (2010) and they use the same variables as Goyal and Welch (2008). Their findings reveal that the predictor-constrained approach significantly enhances out-of-sample forecast accuracy compared to traditional unconstrained models. Moreover, they find this approach to be at least as accurate as those proposed by Campbell and Thompson (2008) and Pettenuzzo et al. (2014), but often demonstrates even greater accuracy. They find that their methodology is significantly more accurate than Goyal and Welch (2008) unrestricted forecasts.

Ferreira and Santa-Clara (2011) explore a novel approach to forecasting stock market returns by focusing on the components of aggregate market returns rather than the aggregate market return itself. The authors argue that forecasting individual components separately and then aggregating the forecasts can lead to more accurate predictions of overall stock market returns. The study decomposes aggregate equity premium into two components: dividend yield and capital gains. The authors then develop forecasting models for each component based on relevant predictor variables from macroeconomic and financial market data. After obtaining individual component forecasts, they combine these forecasts to generate a prediction for the aggregate stock market return. Ferreira and Santa-Clara (2011) find that their component-based approach outperforms univariate models suggested by Goyal and Welch (2008). The out-of-sample performance of their component-based method is shown to be significantly better than that of models based on historical averages as well as the univariate models suggested by Goyal and Welch (2008). Their findings suggest that considering the underlying factors driving market returns and forecasting them separately can lead to more accurate predictions of overall market performance. They provide a new perspective on forecasting stock market returns and highlights the potential benefits of adopting a component-based approach.

Rapach et al. (2010) investigate the out-of-sample predictability of the equity premium by combining forecasts. The authors propose a combination forecasting approach that incorporates multiple predictors and assesses their individual and combined predictive ability. Forecast combinations are constructed by first using univariate forecasting models and then combining them. Their findings reveal that combination forecasts can significantly improve out-of-sample equity premium forecasts compared to the benchmark models. The results also suggest that the predictive power of the combined models is closely related to the real economy, as the models perform better during periods of economic expansion than during recessions. Rapach et al. (2010) consider both equal-weighted combinations, where each individual predictor contributes equally to the combined forecast, and optimally weighted combinations, where the weights are determined based on the individual predictor's past forecasting performance. In their analysis, the authors find that both equal-weighted and optimally weighted combination methods yield significant improvements in out-of-sample forecasting performance compared to the benchmark models.

Ludvigson and Ng (2007) examine the empirical risk-return relationship by employing a factor analysis approach. The authors propose a novel method for equity premium prediction based on a diffusion indices that captures the common fluctuations in a large set of macroeconomic and financial variables. The authors argue that conventional asset pricing models often fail to account for the complex interdependencies between the numerous potential return predictors. To address this issue, they develop a latent factor model that accommodates the co-movements among the predictors, allowing them to extract a small number of relevant factors using principal component analysis. These estimated factors are then utilized as regressors in the predictive regression model. The findings reveal that using diffusion indices significantly outperforms traditional models in terms of in-sample predictive power and out-of-sample forecasting performance. The estimated factors are found to have a strong explanatory power for the equity premium, both statistically and economically. Moreover, the results are robust to different sample periods and variable selections, suggesting the effectiveness of the factor analysis approach in capturing the empirical risk-return relation. Diffusion indices approach is further studied by, for example, Kelly and Pruitt (2013).

Guidolin and Timmermann (2007) investigate asset allocation strategies under a multivariate regime-switching framework. The authors propose a model that accounts for changes in the joint distribution of asset returns, allowing for a more flexible and realistic representation of the dynamics in financial markets. The multivariate regime-switching model is characterized by different states, each representing a specific set of economic conditions, such as periods of high or low growth and volatility. The model allows for the possibility that asset returns and their covariances change across different regimes, which can have a significant impact on optimal asset allocation. The authors employ a Bayesian approach to estimate the model parameters and compute optimal portfolio weights under various investment horizons and risk preferences. The findings reveal that the multivariate regime-switching model leads to substantial improvements in out-of-sample portfolio performance compared to traditional models that assume constant parameters. The results indicate that investors should adjust their portfolio allocations in response to changes in the underlying economic regimes, as ignoring such changes can lead to suboptimal investment decisions. Additionally, the study shows that the benefits of incorporating regime-switching dynamics are more pronounced for longer investment horizons and investors with higher risk aversion. Henkel et al. (2011) and Dangl and Halling (2012) also consider equity premium forecasting under regime-switching framework.

Neely et al. (2014) examine the role of technical indicators in equity premium forecasts. They use the technical indicators defined in Section 5.1.1 for univariate equity premium forecasts using both linear regression models and principal component regression. They compare the results obtained from these models to the univariate linear regression results of macroeconomic predictors (see Section 5.1.2), as well as the principal component regressions (see Section 2.1) formed from these variables. Their findings suggest that the combination of technical indicators and macroeconomic predictors yields superior out-of-sample forecasts compared to models using only macroeconomic predictors. Moreover, the results indicate that technical indicators play a significant role in improving forecast accuracy, providing support for the use of technical analysis in predicting the equity risk premium.

6 Forecasting Equity Premium: Methods and Models

6.1 Forecasting Evaluation

In this section I introduce the forecast evaluation methods used in my thesis. Sections 6.1.1, 6.1.2, and 6.1.3 are largely based on the research by Diebold and Mariano (1995), Stock and Watson (2008), and Hastie et al. (2009, 23-27). Section 6.1.4 is based on the research by Clark and West (2006), and Section 6.1.5 is based on Gallant (1987, 137-139).

6.1.1 Out-of-sample forecasting

Out-of-sample forecasting pertains to scenarios where attempts are made to emulate real-life forecasting situations in which future observations remain unknown. In such a real-world context, an agent possesses observations only up to time t and aims to predict time $t+h$, where h represents the forecast horizon. In this thesis, my primary focus will be on forecasts with $h = 1$. It is important to note that the time frequency may range, say, from daily to annual observations. Typically, macroeconomic data is most frequently available on a monthly basis, while stock prices are available, for example, at daily level.

Out-of-sample forecasting can be executed through various approaches, such as dividing the data into training and testing sets. In this case, the prediction model is "trained" using the training data, and its predictive capacity in the out-of-sample context is assessed on the test data. Dynamic out-of-sample forecasting methods can also be employed, which are commonly utilized in macro and financial econometrics. The most prevalent dynamic out-of-sample forecast models include rolling and expanding window methods. These methods differ from the simple split of training and testing data as the estimation window is not constant.

6.1.2 The expanding window

The expanding window approach is an out-of-sample forecasting method that aims to simulate a real-life forecasting situation. Imagine that the data consists of T observations. In an expanding window approach, a subsample is typically selected and used for estimation before evaluating the forecasting accuracy. Let us say that we first estimate the parameters of the models using first P observations $(1 \leq P < T)$. These estimates are then used for forecasting, before expanding the estimation window. Next, the estimation data is extended to consist of $P + 1$ observations and again estimation is performed on the now extended data. The procedure is continued until $P + k = T$, where k is the number of out-of-sample observations. The results of the estimations at all time points are collected and can be used, for example, to determine the predictive accuracy of a method. Figure 5 visualizes the expanding window out-of-sample forecasting method.

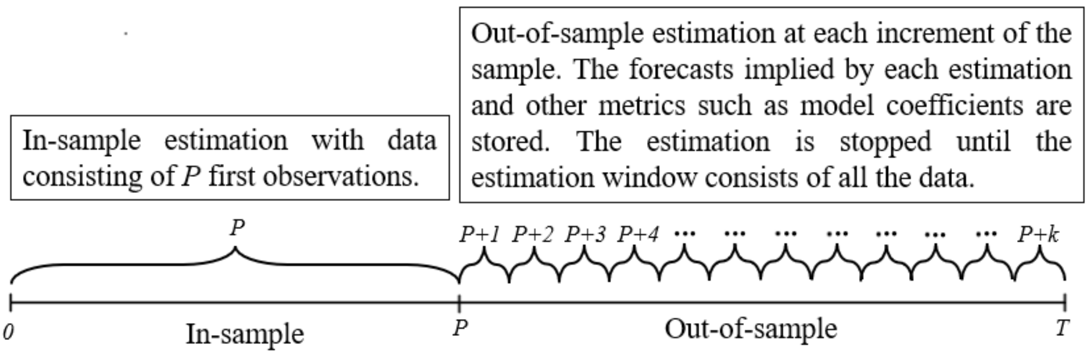

Figure 5. Expanding window out-of-sample forecasting methodology

In Figure 5, it is shown how the expanding window method divides the data into in-sample and out-of-sample portions, with each estimation step expanding the window. Typically, P is chosen to be a small fraction of the total observation set, in which case the out-sample fraction k can be very large. Expanding window method can be computationally very heavy, due to multiple estimation runs, especially if the models used for prediction are complex or computationally intensive. For example, when forecasting monthly stock returns, the in-sample sample length could be 15 years and the out-sample 50 years. The in-sample would then consist of 180 observations and the out-sample of 600 observations. Implementing this with an expanding window approach is not yet very difficult, but if the data used consisted of daily returns, the implementation is clearly more cumbersome.

The expanding window is a plausible idea for a real-life forecasting method in the sense that data is not constant but grows over time. It provides a realistic evaluation of the model's ability to generalize to new data, which is important for assessing the performance of a model in a real-world setting. It also takes account the uncertainty associated with the model and estimation, as well as the possible instability of the model.

The rolling window is an alternative to the expanding window. It differs from the expanding window in that the size of the estimation window is fixed. In other words, while the expanding window increases the window size with each estimation step, the rolling window method removes the oldest observation from the beginning of the time series after each estimation and adds the newest observation to the end of the time series. Among these two methods, the expanding window better represents a real-life forecasting situation, where the length of the time series grows with new observations. For this reason, and following, for example Neely et al. (2014) and Pan et al. (2020), I implement the expanding window method for generating and evaluating my own forecasts instead of the rolling window.

6.1.3 Mean Squared Forecast Error

Mean Squared Forecast Error (MSFE) is a commonly used measure of the accuracy of a forecasting model. It is a measure of the average squared difference between the forecasted values and the actual observed values. The MSFE provides a way to quantify the deviation between the predicted values and the actual values and is widely used to evaluate the performance of different models.

The MSFE is calculated as the average of the squared differences between the forecasted values and the actual values, over a specific forecasting window. The MSFE is expressed as

$$MSFE = \frac{1}{k}\sum_{t=k+1}^{P+k}(f_t - y_t)^2, \quad (27)$$

where k is the size of the forecast window, f_t is the forecast at time t, and y_t is the actual observed value at time t.

A lower MSFE indicates a better predictive accuracy of the model. The MSFE can be used to compare the performance of different forecasting models and to select the best-performing model for a given problem. It is important to note that the MSFE is sensitive to outliers, so it is sometimes necessary to use other measures of accuracy, such as Mean Absolute Error (MAE) or Mean Absolute Percentage Error (MAPE), to get a more robust assessment of model performance. Overall, the MSFE is a valuable tool for evaluating the accuracy of forecasting models and for comparing the performance of different models.

6.1.4 Clark and West test

The Clark and West test is a statistical test used to compare the accuracy of two or more forecasting models. It was developed by Stephen Clark and James West in 2004 as an alternative to the Diebold and Mariano (1995) test, which was the most commonly used test for comparing the accuracy of forecasting models at the time.

The main motivation behind the development of the Clark and West test was to address some of the limitations of the Diebold-Mariano test, particularly when comparing nested models. The Clark and West test is based on the idea of comparing the mean squared forecast error (MSFE) between two models. The test statistic is based on the differences in MSFE between the models, and it is calculated as the ratio of the difference in MSFE between the models to a correction term that accounts for the uncertainty in the MSFE estimates.

The correction term in the Clark and West test statistic is designed to account for the serial correlation in the forecast errors, which can lead to incorrect inference about the differences in accuracy between the models if it is not taken into account. The Clark and West test statistic also adjusts for differences in the persistence of the forecast errors between the models, which can affect the accuracy of the MSFE estimates if not considered.

The Clark and West test is implemented by first calculating the MSFE for each model over a given sample period, and then using the MSFE estimates to calculate the Clark and West test statistic. The test statistic is then compared to critical values from a standard normal distribution to determine whether there is evidence to reject the null hypothesis that the models have the same accuracy. The Clark and West (2006) test can be implemented mathematically as follows:

1. Construct the forecast errors for each model for each time period.
2. Compute the mean squared forecast error (MSFE) for each model.
3. Calculate the difference in MSFE between each pair of models.
4. Test the null hypothesis that the difference in MSFE between each pair of models is equal to zero.

5. If the null hypothesis is rejected, then the models are deemed to have significantly different accuracy.

To perform the test, a t-statistic is calculated for each difference in MSFE, using the following formula:

$$t = \frac{d_i}{\sqrt{Var(d_i)}}, \tag{28}$$

where d_i is the difference in MSFE between two models and $Var(d_i)$ is the variance of the difference in MSFE.

The t-statistic is then compared to a critical value from the t-distribution with degrees of freedom equal to the number of time periods minus two. If the calculated t-statistic is greater than the critical value, then the null hypothesis is rejected and the difference in MSFE between the two models is considered statistically significant.

The Clark and West test is a robust method for comparing the accuracy of nested forecasting models that takes into account the serial correlation in the forecast errors and adjusts for differences in the persistence of the forecast errors between the models. It provides more accurate inferences about the differences in accuracy between nested models than the Diebold-Mariano test, and it has become a widely used alternative to the Diebold-Mariano test in the forecasting literature.

It's important to note that the Clark and West test assumes that the forecast errors are normally distributed and have constant variance over time. If these assumptions are not met, the results of the test may be biased and should be interpreted with caution.

6.1.5 Newey-West adjusted heteroscedastic t-statistics

To account for potential autocorrelation and heteroscedasticity of the error terms when estimating parameters for linear regression models, I use Newey-West adjusted heteroscedastic t-statistics, to evaluate the statistical significance of different predictive variables.

Consider a predictive linear regression model

$$Y_{t+1} = X_t'\beta + \varepsilon_{t+1}, \tag{29}$$

where Y_{t+1} is the dependent variable, X_t is a p-dimensional vector of predictive variables (generally including a constant term), β is a p-dimensional vector of parameters, and ε_{t+1} is the error term. The Newey-West estimator of the long-run covariance matrix is given by:

$$\Omega_{NW} = \sum_{j=-q}^{q} \omega_j \, E\left[\varepsilon_t \varepsilon_{t-j}\right], \tag{30}$$

where q is the maximum lag length, and ω_j are the weights assigned to the covariances of the residuals at different lags. A popular weighting scheme is the Bartlett kernel, which is defined as:

$$\omega_j = 1 - |j|/(q+1), \tag{31}$$

for $j = 0, ..., q$. After estimating the long-run covariance matrix Ω_{NW}, the Newey-West standard errors for the estimated parameters can be computed. The Newey-West covariance matrix of the estimated parameters, $V_{NW}(\beta)$, is given by:

$$V_{NW}(\beta) = (X'X)^{-1}(X'\Omega_{NW}X)(X'X)^{-1}, \tag{32}$$

where X is the $T \times p$ matrix of predictive variables. Finally, the Newey-West adjusted heteroscedastic t-statistics for each parameter are calculated as follows:

$$t_{NW}(i) = \hat{\beta}_i / \sqrt{\left(V_{NW}(\hat{\beta})_{ii}\right)}, \tag{33}$$

where $\hat{\beta}_i$ is the i-th estimated parameter, and $V_{NW}(\hat{\beta})_{ii}$ is the i-th diagonal element of the Newey-West covariance matrix. Employing the Newey-West adjusted heteroscedastic t-statistics ensures more accurate inferences about the parameters of our linear regression models, accounting for potential autocorrelation and heteroskedasticity in the error terms.

6.2 Forecasting Models

Before generating forecasts, the forecasting models used in this study are presented. These models are based on those proposed in the articles reviewed in the literature review section, with some modifications. Macroeconomic predictor-based models are directly adapted from the previously suggested models in the literature, while the use of technical indicators has been expanded to various models. Forecasts are generated by implementing the Campbell and Thompson (2008) and Pan et al. (2020) (From this point forward, Pan restriction) models both with and without constraints.

6.2.1 Bivariate regression models

Univariate forecasts are formed using single-predictor linear regression models. These are constructed using all macroeconomic predictors as well as all individual technical indicators individually.

$$\hat{y}_{t+1} = \hat{\alpha}_{i,t} + \hat{\beta}_{i,t} x_{i,t} \tag{34}$$

where $\hat{\alpha}_{i,t}$ and $\hat{\beta}_{i,t}$ are ordinary least squares estimates of bivariate predictive regression model for predictor x_i at time t and $\hat{y}_{t+1}$ is the equity premium forecast. Both, Pan and CT restrictions are implemented forecasts constructed using macroeconomic predictors. Forecasts that use individual technical indicators are referred to as TECH forecasts. Pan constraints are not implemented for TECH variables, as the use of Pan constraints would not be meaningful for binary variables. Similarly, CT constraints are not utilized in forecasts produced by technical indicators, as the CT constraints do not alter the forecasts at all (given that the parameter estimates have correct signs). Technical indicators cannot take negative values, and it is highly unlikely that forecasts using technical indicators would imply negative equity premium forecasts.

6.2.2 Multiple-predictor regression models

Following the literature, I use two multiple regression models to forecast equity premium using both predictor sets, technical indicators and macroeconomic predictors. Again, CT and Pan restrictions are implemented to forecasts produced by macroeconomic predictors.

Multiple regression models include the Kitchen Sink forecast model used by Goyal and Welch (2008), which has been slightly modified. In their specification, it includes 14 macroeconomic predictors, but my specification does not include the variables ***DE*** and ***TMS***, as these variables can be formed as linear combinations of other predictors, making the use of a linear model problematic. Thus, the model is the form

$$\hat{y}_{t+1} = \hat{\alpha}_t + \boldsymbol{x}'_t \widehat{\boldsymbol{\beta}}_t, \tag{35}$$

where $\hat{\alpha}_t$ *and* $\widehat{\boldsymbol{\beta}}_t$ are ordinary least squares estimates of predictive multiple regression model for predictor set $\boldsymbol{x}_t$ at time *t*.

Second multiple regression model is one in which, at each estimation step, a maximum of three predictive variables are selected from the same 12 variables as in the previous multiple regression model, using the Bayesian information criterion. Thus, following Cremers (2002) the model takes the form

$$\hat{y}_{t+1} = \hat{\alpha}_{m,t} + \boldsymbol{x'}_{m,t}\hat{\boldsymbol{\beta}}_{m,t}, \tag{36}$$

where *m* refers to the predictor set with the lowest BIC at time *t*. For example, m can be 1, when the BIC choose only the first variable as predictor, or m can be (1,6,12), when according to the BIC the best model to predict equity premium consist of the first, the sixth and the twelfth predictor of the corresponding predictor space $\boldsymbol{x_t}$. Coefficients $\hat{\alpha}_{m,t}$ *and* $\hat{\boldsymbol{\beta}}_{m,t}$ are ordinary least squares estimates of predictive multiple regression model for the predictor vector $\boldsymbol{x}_{m,t}$ at time *t*.

6.2.3 Forecast combinations

Following Rapach et al (2010), I examine the forecast combination of different bivariate forecasts. Forecast combinations are implemented separately for forecasts produced with technical indicators and for forecasts produced with macroeconomic predictors. I consider two types of forecast combinations. The equally weighted forecast combination, which is referred to as POOL-AVERAGE, and the optimally weighted forecast combination, which is referred to as POOL-DMSFE (discount MSFE). For forecasts produced by bivariate models, pooled forecasts take the form

$$\hat{y}_{t+1}^{pool} = \sum_{i=1}^{M} \omega_{i,t}\hat{y}_{i,t+1}\,, \tag{37}$$

where $\hat{y}_{i,t+1}$ is the forecast produced with bivariate regression model *i, M* is the total number of bivariate forecasts produced using either macroeconomic predictors or technical indicators, and $\omega_{i,t}$ the weight of the forecast *i*. Following, Stock and Watson (2004) and Rapach et al. (2010)

$$\omega_{i,t} = \frac{\frac{1}{\phi_{i,t}}}{\sum_{m=1}^{M}\frac{1}{\phi_{m,t}}}, \tag{38}$$

where

$$\phi_{i,t} = \sum_{s=P_0}^{t-1} \theta^{t-1-s}\left(y_{s+1} - \hat{y}_{i,s+1}\right)^2, \tag{39}$$

where θ is a discount factor and $P_0 + 1$ is the start of the holdout out-of-sample period. Following Rapach et al. (2010) I use $\theta = 0.75$ and $\theta = 1$ as discount factors. Setting the discount factor to $\theta = 1$ is a special case of POOL-DMSFE resulting equally weighted forecast combination, i.e., POOL-AVERAGE model, which can be written as

$$\hat{y}_{t+1}^{pool} = \frac{1}{M}\sum_{i=1}^{M} \hat{y}_{i,t+1}. \tag{40}$$

6.2.4 Diffusion indices

In diffusion indices forecasts, which are based on the principal components of predictor spaces, the predictors must first be standardized by subtracting their variable-specific means and dividing by their variable-specific standard deviations (Ludvigson & Ng, 2007). After this, principal components are formed from the set of predictors, and the first principal component (Rapach et al., 2013) is used in the predictive principal component regression. Again, this is carried out separately for technical indicators and macroeconomic predictors.

6.2.5 Sum-of-the-parts

Following Ferreira and Santa-Clara (2011), I generate forecasts using their sum-of-the-parts method. The sum-of-the-parts method does not use technical indicators, but only focuses on DP (dividend-price ratio) and earnings. Sum-of-the-parts model is form

$$\hat{y}_{t+1}^{SOP} = \bar{e}_{t,20} + log(\mathrm{DP} + 1) - r_{t+1}^{\mathrm{f}}, \tag{41}$$

where $\hat{y}_{t+1}^{SOP}$ is the equity premium forecast produced using the sum-of-the-parts model, $\bar{e}_{t,20}$ is the 20-year moving-average of log earnings growth at time *t* and r_t^{f} is the risk-free rate.

6.2.6 Technical indicator combinations

Lastly, I present a new, but fairly simple, way to combine technical indicators. In this method, technical indicators are used in a multiple-predictor models in such a way that the equity premium is predicted together by ***MA(s, l)***, ***VOL(s, l)***, and ***MOM(l)***. Thus, the combinations are formed by combining all three different indicators and using the same s and l lengths in all of them. A total of six sets of these indicator groups are generated, and I refer to them as ***COMB(s, l)***. Table 6 shows the forecast sets formed by different technical indicators are presented.

Table 6. Technical indicator combination sets

COMB(1, 9)	***{MA(1, 9), MOM(9), VOL(1, 9)}***
COMB(1, 12)	***{MA(1, 12), MOM(12), VOL(1, 12)}***
COMB(2, 9)	***{MA(2, 9), MOM(9), VOL(2, 9)}***
COMB(2, 12)	***{MA(2, 12), MOM(12), VOL(2, 12)}***
COMB(3, 9)	***{MA(3, 9), MOM(9), VOL(3, 9)}***
COMB(3, 12)	***{MA(3, 12), MOM(12), VOL(3, 12)}***

7 Results

In this section, I present the forecast results in both statistical (MSFE) and economic (CER gains) terms. Of the 17 variables used by Goyal and Welch (2008), I employ 14 variables[6] in my own analysis. Technical indicators are also presented. Thus, I will use the total of 28 individual predictors which are presented in Table 7.

Table 7. All individual predictor variables

Variable	Definition
DP	Dividend Price Ratio
DY	Dividend Yield
EP	Earnings Price Ratio
DE	Dividend Payout Ratio
SVAR	Stock Variance
BM	Book-to-Market Ratio
NTIS	Net Equity Expansion
TBL	Treasury-bill Rate
LTY	Long-term Government Bond
LTR	Long-term Rate of Returns
TMS	Term-spread
DFY	Default Yield Spread
DFR	Default Return Spread
INFL	Consumer Price Inlation
MA(1, 9)	Moving-Average indicator with *s*=1 and *l*=9
MA(1, 12)	Moving-Average indicator with *s*=1 and *l*=12
MA(2, 9)	Moving-Average indicator with *s*=2 and *l*=9
MA(2, 12)	Moving-Average indicator with *s*=2 and *l*=12
MA(3, 9)	Moving-Average indicator with *s*=3 and *l*=9
MA(3, 12)	Moving-Average indicator with *s*=3 and *l*=12
MOM(9)	Momentum Indicator with *l*=9
MOM(12)	Momentum Indicator with *l*=12
VOL(1, 9)	Volume based indicator with *s*=1 and *l*=9
VOL(1, 12)	Volume based indicator with *s*=1 and *l*=12
VOL(2, 9)	Volume based indicator with *s*=2 and *l*=9
VOL(2, 12)	Volume based indicator with *s*=2 and *l*=12
VOL(3, 9)	Volume based indicator with *s*=3 and *l*=9
VOL(3, 12)	Volume based indicator with *s*=3 and *l*=12

[6] Amit Goyal's updated dataset on his website does not include the variables ***EQIS***, ***cayp***, or ***IK***; therefore, I will not use them in the empirical part of my thesis.

In addition to the variables presented in Table 7, I also examine the predictive power of six combined technical indicators shown in Table 6. The variables are commonly used in all presented forecasting models, except for the Sum-of-the-parts model, which is only applied in the cases of the unrestricted and CT-constrained models and solely with macroeconomic predictors. The predictive power of the variables and different models are examined through both statistical and economic significance.

7.1 Statistical significance

In this section, the predictive ability of forecasting models and variables is examined in a statistical sense. Following Goyal and Welch (2008), as well as many others, predictive ability is examined through the so-called R2 metric, which is of the form of (25). I present the R2 values separately for all bivariate forecasting model variables in the entire sample, as well as in different macroeconomic states of the world, which are determined according to NBER's recession and expansion periods. Forecasts made with other models are presented in a comprehensive manner, but like the bivariate forecasting models, predicting in different business cycle regimes (macroeconomic states) is examined separately.

Table 8. R2 statistics for all different bivariate models. Full forecast evaluation period.

R2 statistics for all different bivariate forecast models during the entire forecast evaluation period. Note that the R2 is the out-of-sample R2 presented in equation (25), not the traditional R-squared. In the table, * indicates statistical significance at least at the 10% risk level, ** correspondingly at least at the 5% risk level, and *** at least at the 1% risk level.

R2 Overall					
Predictor	Unconstrained	CT	PAN	TECH	
DP	-0.98	-0.76	-0.07	***MA(1, 9)***	0.13
DY	-1.02	-0.79	-0.01	***MA(1, 12)***	0.37
EP	-0.93	-0.45	-0.15	***MA(2, 9)***	0.12
DE	-1.40	-0.52	-0.22	***MA(2, 12)***	0.53*
SVAR	-1.84	-1.84	-2.82	***MA(3, 9)***	-0.20
BM	-1.31	-1.11	-0.24	***MA(3, 12)***	-0.14
NTIS	-1.01	-0.95	-0.13	***MOM(9)***	-0.06
TBL	-1.10	-0.03	-0.30	***MOM(12)***	-0.03
LTY	-0.96	0.31*	-1.25	***VOL(1, 9)***	0.21
LTR	-0.34*	-0.31	-0.38	***VOL(1, 12)***	0.41*
TMS	-1.29	-1.47	-0.16	***VOL(2, 9)***	0.26
DFY	-0.55	-0.47	-0.46	***VOL(2, 12)***	0.30
DFR	-0.66	-0.11	-1.13	***VOL(3, 9)***	-0.06
INFL	-0.68	-0.34	-0.22	***VOL(3, 12)***	0.51*

Table 8 presents R2 values for different forecasting models during the entire forecast evaluation period. In Table 8, it is evident that none of the unconstrained models have positive R2 value and only one of them is statistically significant, suggesting a poor fit or insignificant results. As for the models with CT restrictions, none of the models show statistical significance, with most models exhibiting negative R2 values. Similarly, models with PAN restrictions also have negative R2 values, implying that they are not outperforming the historical average forecast.

When considering the moving-average indicators, it is noteworthy that the ***MA(2, 12)*** indicator has a positive R2 value of 0.53 and is statistically significant. Moreover, the ***VOL(1, 12)*** and ***VOL(3, 12)*** models also have statistically significant positive R2 values of 0.41 and 0.51, respectively. However, the other forecasts produced using technical indicators do not demonstrate statistical significance, indicating that they may not provide reliable predictions for the equity premium. It can be observed that only among the technical indicators are the only forecasts that are statistically significantly more accurate than the historical average throughout the entire forecast evaluation period.

Table 9 presents the R2 statistics for all different bivariate forecast models during the NBER's expansion periods in forecast evaluation period.

Table 9. R2 statistics for all different bivariate models. Business cycle expansion periods.

R2 statistics for all different bivariate forecast models during the NBER's expansion periods in forecast evaluation period. Note that the R2 is the out-of-sample R2 presented in equation (25), not the traditional R-squared. In the table, * indicates statistical significance at least at the 10% risk level, ** correspondingly at least at the 5% risk level, and *** at least at the 1% risk level.

R2 Expansion					
Predictor	Unconstrained	CT	PAN	TECH	
DP	-1.97	-1.64	-0.28	***MA(1, 9)***	-0.76
DY	-2.33	-1.97	0.19	***MA(1, 12)***	-0.70
EP	-0.88	-0.68	-0.41	***MA(2, 9)***	-0.65
DE	-2.05	-1.27	0.38	***MA(2, 12)***	-0.27
SVAR	0.04	0.04	1.01	***MA(3, 9)***	-0.95
BM	-0.95	-0.87	-0.02	***MA(3, 12)***	-0.47
NTIS	-0.01*	0.03*	-0.02	***MOM(9)***	-0.45
TBL	-1.42	0.22*	0.03	***MOM(12)***	-0.45
LTY	-1.22	0.41*	-3.57	***VOL(1, 9)***	-0.58
LTR	-2.12	-1.26	-0.83	***VOL(1, 12)***	-0.29
TMS	-3.04	-3.07	-0.26*	***VOL(2, 9)***	-0.15
DFY	-0.53	-0.53	-0.53	***VOL(2, 12)***	0.16
DFR	0.07	0.13	0.16	***VOL(3, 9)***	-0.32
INFL	-0.38	-0.05	-0.30	***VOL(3, 12)***	0.08

From Table 9 it can be observed that the macroeconomic predictors with or without any restrictions generally exhibit negative R2 values, indicating a weaker predictive power than the historical average model. Notably, ***NTIS*** and ***TBL*** show a slightly positive R2 value when CT restrictions are implemented, and ***TMS*** exhibits a marginally positive R2 value with PAN restrictions.

All Moving-Average based technical indicators display negative R2 values, which implies weaker forecasting performance than the historical average model during economic expansions. Similarly, the momentum models ***MOM(9)*** and ***MOM(12)*** exhibit negative R2 values, signifying a lack of predictive power in expansionary phases. Volume based models present mixed results. While ***VOL(1, 9)***, ***VOL(1, 12)***, ***VOL(2, 9)***, and ***VOL(3, 9)*** show negative R2 values, ***VOL(2, 12)*** and ***VOL(3, 12)*** report positive R2 values.

In summary, most forecasting models exhibit weak performance during economic expansions, with only a few models presenting limited predictive power. This highlights the difficulty in accurately forecasting stock returns during periods of normal economic growth.

Table 10 presents R2 statistics for all different bivariate forecast models during the NBER's recession periods in forecast evaluation period.

Table 10. R2 statistics for all bivariate forecasts. Business cycle recession periods.

R2 statistics for all different bivariate forecast models during the NBER's recession periods in forecast evaluation period. Note that the R2 is the out-of-sample R2 presented in equation (25), not the traditional R-squared. In the table, * indicates statistical significance at least at the 10% risk level, ** correspondingly at least at the 5% risk level, and *** at least at the 1% risk level.

R2 Recession					
Predictor	Unconstrained	CT	PAN	TECH	
DP	1.23	1.22	0.40	***MA(1, 9)***	2.16**
DY	1.94**	1.88**	-0.46	***MA(1, 12)***	2.77**
EP	-1.03	0.06	0.44	***MA(2, 9)***	1.86*
DE	0.08	1.17	-1.57	***MA(2, 12)***	2.36*
SVAR	-6.08	-6.08	-11.47	***MA(3, 9)***	1.50
BM	-2.12	-1.64	-0.72	***MA(3, 12)***	0.59
NTIS	-3.27	-3.16	-0.38	***MOM(9)***	0.85
TBL	-0.38	-0.61	-1.06	***MOM(12)***	0.91
LTY	-0.37	0.09	3.98*	***VOL(1, 9)***	2.00**
LTR	3.65**	1.82*	0.63	***VOL(1, 12)***	2.00*
TMS	2.66**	2.14**	0.06	***VOL(2, 9)***	1.18
DFY	-0.58	-0.32	-0.33	***VOL(2, 12)***	0.60
DFR	-2.28	-0.64	-4.04	***VOL(3, 9)***	0.54
INFL	-1.36	-0.98	-0.05	***VOL(3, 12)***	1.49

From Table 10 it can be observed that during recession periods, the unconstrained models, and models with CT, and PAN restrictions display mixed results. Positive R2 values are observed for predictors ***DP***, ***DY***, ***DE***, ***LTY***, ***LTR***, and ***TMS***. In particular, the unconstrained and CT models have statistically significant positive R2 values for ***DY***

(1.94** and 1.88** respectively) and TMS (2.66** and 2.14** respectively). The ***LTR*** predictor exhibits a statistically significant positive R2 value for the unconstrained model (3.65**), while the CT model has the highest significant R2 value for variable ***LTY*** (3.98*).

The moving-average technical indicator models, such as ***MA(1, 9)***, ***MA(1, 12)***, ***MA(2, 9)***, and ***MA(2, 12)***, show generally positive R2 values during recession periods. Among them, ***MA(1, 9)*** and ***MA(1, 12)*** exhibit statistically significant R2 values of 2.16** and 2.77**, respectively. Similarly***, MA(2, 9)*** and ***MA(2, 12)*** have significant R2 values of 1.86* and 2.36*, respectively. Volume based models present mixed results, with ***VOL(1, 9)*** and ***VOL(1, 12)*** showing statistically significant R2 values of 2.00** and 2.00*, respectively. It is worth noting that, unlike with macroeconomic predictors, all technical indicators have a positive R2, which indicates their superiority compared to the historical average. This is observed in only a few macroeconomic predictors.

Table 11 presents R2 statistics for multiple-predictor forecast models during full forecast evaluation period (containing both expansion and recession periods).

Table 11. R2 for different forecasting models. Full forecast evaluation period.

Panel A presents R2 values for full forecast evaluation period for multiple-predictor models with macroeconomic predictors (second, third and fourth column) and technical indicators (fifth column). Note that the R2 is the out-of-sample R2 presented in equation (25), not the traditional R-squared. Panel B presents R2 values for the same period but using TECH COMB forecasts. * indicates statistically significantly lower MSFE than historical average forecast according to Clarke and West (2006) test at least at the 10% risk level, ** correspondingly at least at the 5% risk level, and *** at least at the 1% risk level.

R2 Overall						
Panel A	Unconstrained	CT	PAN	TECH	Panel B	TECH COMB
Kitchen sink	-6.54	-1.06**	-6.97	-1.19	***Comb(1,9)***	-0.17
BIC	-7.49	-3.37	-3.45	0.17	***Comb(1,12)***	0.24
POOL-AVG	0.30	0.41*	0.13	0.30	***Comb(2,9)***	-0.26
POOL-DMSFE	0.36	0.36*	0.15	0.30	***Comb(2,12)***	0.19
Diffusion indices	0.07	0.12	-0.08	0.31	***Comb(3,9)***	-0.84
Sum-of-the-parts	0.55**	0.63**			***Comb(3,12)***	0.27

Table 11 shows mixed results for different forecasting models. The Kitchen Sink and BIC models perform significantly worse, and the Sum-of-the-parts model appears to work best. Pooled models also perform well, producing exclusively positive R2 values throughout the entire evaluation period. From the table, it can also be observed that forecasts implemented with CT constraints produce statistically significantly smaller MSFEs than the historical average model. In particular, the Sum-of-the-parts model appears to produce the smallest MSFEs at the lowest risk level.

Technical indicators demonstrates mixed results, with positive R2 values for BIC (0.17), POOL-AVG (0.30), POOL-DMSFE (0.30), and the diffusion indices (0.31). However, negative R2 values are observed for the kitchen sink (-1.19) model.

Regarding the TECH comb models, ***Comb(1, 12)***, ***Comb(2, 12)*** and ***Comb(3, 12)*** show positive R2 values of 0.24, 0.19, and 0.27 respectively, while ***Comb(1, 9)*** and ***Comb(2, 9)*** present negative R2 values of -0.17 and -0.26, respectively. Additionally, ***Comb(3, 9)*** has a significantly negative R2 value of -0.84, and shows a positive R2 value of 0.27.

Table 12 presents R2 statistics for different forecasting models on expanding periods.

Table 12. R2 for different forecasting models. Business cycle expansion periods.

Panel A presents R2 values for expansion periods for multiple-predictor models with macroeconomic predictors (second, third and fourth column) and technical indicators (fifth column). Panel B presents R2 values for the same period but using TECH COMB forecasts. Note that the R2 is the out-of-sample R2 presented in equation (25), not the traditional R-squared. * indicates statistically significantly lower MSFE than historical average forecast according to Clark and West (2006) test at least at the 10% risk level, ** correspondingly at least at the 5% risk level, and *** at least at the 1% risk level.

R2 Expansion						
Panel A	Unconstrained	CT	PAN	TECH	Panel B	TECH COMB
Kitchen sink	-1.91**	0.48***	-3.66	-3.31	***Comb(1,9)***	-1.43
BIC	-6.33	-3.66	-3.21	-0.70	***Comb(1,12)***	-0.87
POOL-AVG	0.01	0.02	0.26	-0.26	***Comb(2,9)***	-1.22
POOL-DMSFE	-0.11	-0.08	0.23	-0.27	***Comb(2,12)***	-0.61
Diffusion indices	-0.51	-0.51	-0.25	-0.42	***Comb(3,9)***	-1.74
Sum-of-the-parts	0.17	0.19			***Comb(3,12)***	-0.18

Table 12 shows that all forecasts produced using technical indicators generate negative R2 statistics, implying that during expansion periods, multidimensional models formed with technical indicators are not effective in predicting equity premium. The weakness of technical indicators in predicting equity premium during expansions is further supported by the fact that in Table 7, only the ***VOL(3, 12)*** indicator had better predictive ability than the historical average model, albeit very slightly.

Multiple-predictor models implemented with macroeconomic predictors are also mostly weak, although among these, the Sum-of-the-parts model without and with constraints, as well as all POOL-AVG model implemented with macroeconomic indicators, produce positive R2 values.

Table 13 presents R2 statistics for different forecasting models on recession periods.

Table 13. R2 for different forecasting models. Business cycle recession periods.

Panel A presents R2 values for recession periods for multiple-predictor models with macroeconomic predictors (second, third and fourth column) and technical indicators (fifth column). Panel B presents R2 values for the same period but using TECH COMB forecasts. Note that the R2 is the out-of-sample R2 presented in equation (25), not the traditional R-squared. * indicates statistically significantly lower MSFE than historical average forecast according to Clark and West (2006) test at least at the 10% risk level, ** correspondingly at least at the 5% risk level, and *** at least at the 1% risk level.

R2 Recession						
Panel A	Unconstrained	CT	PAN	TECH	Panel B	TECH COMB
Kitchen sink	-16.97	-4.52	-14.43	3.60**	***Comb(1,9)***	2.68**
BIC	-10.12	-2.73	-3.99	2.12*	***Comb(1,12)***	2.77**
POOL-AVG	0.97	1.26*	-0.14	1.55*	***Comb(2,9)***	1.91*
POOL-DMSFE	1.43	1.37*	-0.05	1.58*	***Comb(2,12)***	2.01*
Diffusion indices	1.39**	1.54**	0.31	1.93*	***Comb(3,9)***	1.20
Sum-of-the-parts	1.40*	1.63**			***Comb(3,12)***	1.30

Table 13 shows that during recession periods, the Kitchen sink and BIC models using macroeconomic predictors produce significantly less accurate forecasts in terms of MSFE than other comparison models. The TECH model consistently demonstrates positive R2 values, and the TECH comb models show positive R2 values for all combinations, suggesting that these models may be more suitable for forecasting stock returns during recession periods. When examining the predictive power of technical indicators compared to macroeconomic predictors using the same models, it is observed that technical indicators outperform macroeconomic predictors in all models (except, of course, the sum-of-the-parts model, which is not suitable for technical indicators). It can also be seen that TECH comb models generally perform better than forecasts formed with macroeconomic predictors.

From Table 13, it can also be observed that for all forecasts formed using technical indicators, with the exception of ***Comb(3, 9)*** and ***Comb(3, 12)***, the Clark and West (2006) test indicates statistically significantly smaller MSFEs than the historical average model at least at the 10% significance level. It is also observed that in all cases except for forecasts produced by BIC and Kitchen sink models, the forecasts implemented with CT constraints have statistically significantly smaller MSFEs than the historical average model.

7.2 Economic significance

Examining economic significance, following for example Campbell and Thompson (2008), Ferreira and Santa-Clara (2011), Neely et al. (2014), and Pan et al. (2020), the Certainty Equivalent Returns (*CERs*, see Section 3.5) are considered for a mean-variance agent who optimally allocates their assets between equities and risk-free interest rates based on the equity premium forecasts produced using examined models. At the end of each period *t*, the agent invests a proportion w_t of their assets in equities and a proportion $1 - w_t$ in risk-free rate. Optimal w_t is given by

$$w_t = \frac{1}{\gamma}\frac{\hat{y}_{t+1,i}}{\hat{\sigma}_{t+1,i}}, \tag{43}$$

where $\hat{y}_{t+1,i}$ is the forecast produced with model *i,* $\hat{\sigma}_{t+1,i}$ is a forecast of its variance and γ is the coefficient of investors relative risk aversion. In accordance with Neely et al. (2014), w_t is set between 0 and 1.5, allowing for 50% leverage. The investor is assumed to use a moving window of the past five years of monthly stock returns to determine the variance of the equity premium. The portfolio return for the following period (*t*+1) is

$$R_{p,t+1} = w_t y_{t+1} + r_{t+1}^f, \tag{44}$$

where y_{t+1} is the realized equity premium and r_{t+1}^f is the realized risk-free rate on period *t*+1. Thus, the *CER* of the portfolio is

$$CER_p = \hat{\mu}_{p,t} - \frac{1}{2}\gamma\hat{\sigma}_{t+1,i}, \tag{45}$$

where $\hat{\mu}_{p,i}$ is the sample mean and $\hat{\sigma}_{t+1,i}$ the sample variance of the investor's portfolio during forecast evaluation period.

As the measure of economic significance of different models and variables I calculate *CER* gains, the differences between *CER*s obtained using different forecasting models presented in Section 6.2 and *CER*s obtained using historical average model. *CER* gain for model *i* is denoted as Δ_i. *CER* gains are multiplied by 1200 so that they can be intepreted as annual percentages.

Table 14 presents the CER gains for different macroeconomic predictors and technical indicators over the entire forecast evaluation period.

Table 14. CER gains for bivariate forecasts. Full forecast evaluation sample.

Panel A presents CER gains for bivariate forecasts produced by macroeconomic predictors within the full forecast evaluation period (columns 2, 3, 5). Panel B presents CER gains for bivariate forecasts produced by technical indicators (column 6) for full forecast sample.

Δ Overall					
Panel A	Unrestricted	CT	PAN	Panel B	TECH
DP	-1.53	-1.53	0.17	***MA(1,9)***	1.24
DY	-1.22	-1.22	0.38	***MA(1,12)***	2.15
EP	0.16	0.16	-0.11	***MA(2,9)***	1.38
DE	-0.80	-0.73	-0.08	***MA(2,12)***	2.33
SVAR	0.79	-0.01	0.07	***MA(3,9)***	0.95
BM	-1.54	-1.37	0.12	***MA(3,12)***	0.80
NTIS	-0.80	-0.79	-0.15	***MOM(9)***	0.99
TBL	0.36	0.36	-0.30	***MOM(12)***	0.93
LTY	0.37	0.37	-0.36	***VOL(1,9)***	1.17
LTR	0.20	0.20	-0.04	***VOL(1,12)***	1.77
TMS	0.37	0.37	1.66	***VOL(2,9)***	1.02
DFY	-0.75	-0.67	-0.58	***VOL(2,12)***	1.06
DFR	0.27	0.88	-0.14	***VOL(3,9)***	0.46
INFL	-0.12	-0.12	-0.17	***VOL(3,12)***	1.61

From Table 14 it can be observed that technical indicators consistently produce positive CER gains, which implies their clear superiority over the historical average. Among them, the ***MA(1, 12)*** indicator yields the highest economic benefit relative to the historical mean model. Table 14 shows that the CT constraints do not alter the unrestricted Goyal and Welch (2008) setup in any way. Similarly, the PAN constraints do not guarantee high economic benefits when examining the entire forecast evaluation period in. Among the forecasts implemented with PAN constraints, only TMS exhibits relatively high CER gains. In the unrestricted models, and also with the CT constrains and PAN constrains, variable-specific negative economic benefits are observed in relation to the historical average model.

Table 15 presents the CER gains for different macroeconomic predictors and technical indicators for NBER expansion periods.

Table 15. CER gains for bivariate forecasts. Business cycle expansion periods.

Panel A presents CER gains for bivariate forecasts produced by macroeconomic predictors within expansion periods (columns 2, 3, 5). Panel B presents CER gains for bivariate forecasts produced by technical indicators (column 6) within expansion periods.

Δ Expansion					
Panel A	Unrestricted	CT	PAN	Panel B	TECH
DP	-1.53	-1.53	0.17	***MA(1,9)***	-0.68
DY	-1.22	-1.22	0.38	***MA(1,12)***	-0.56
EP	0.16	0.16	-0.11	***MA(2,9)***	-0.54
DE	-0.80	-0.73	-0.08	***MA(2,12)***	0.11
SVAR	0.79	-0.01	0.07	***MA(3,9)***	-0.90
BM	-1.54	-1.37	0.12	***MA(3,12)***	-0.48
NTIS	-0.80	-0.79	-0.15	***MOM(9)***	-0.34
TBL	0.36	0.36	-0.30	***MOM(12)***	-0.38
LTY	0.37	0.37	-0.36	***VOL(1,9)***	-0.60
LTR	0.20	0.20	-0.04	***VOL(1,12)***	-0.33
TMS	0.37	0.37	1.66	***VOL(2,9)***	-0.34
DFY	-0.75	-0.67	-0.58	***VOL(2,12)***	-0.01
DFR	0.27	0.88	-0.14	***VOL(3,9)***	-0.41
INFL	-0.12	-0.12	-0.17	***VOL(3,12)***	-0.10

Table 15 reveals similar findings to those in Table 9. Variables that produce superior forecasts than the historical average during expansion periods are scarce. Technical indicators, with one exception, produce negative CER gains. Forecasts generated using macroeconomic predictors are also predominantly weak, although there are several low but positive CER gains among them. Constraints do not provide any noticeable benefit in comparison to unrestricted bivariate forecasts.

Table 16 presents the CER gains for different macroeconomic predictors and technical indicators for NBER recession periods.

Table 16. CER gains for bivariate forecasts. Business cycle recession periods.

Panel A presents CER gains for bivariate forecasts produced by macroeconomic predictors within recession periods (columns 2, 3, 5). Panel B presents CER gains for bivariate forecasts produced by technical indicators (column 6) within recession periods.

Δ Recession					
Panel A	Unrestricted	CT	PAN	Panel B	TECH
DP	2.78	2.78	4.67	***MA(1,9)***	13.04
DY	6.50	6.50	0.52	***MA(1,12)***	18.85
EP	4.35	4.35	2.43	***MA(2,9)***	13.20
DE	3.71	4.22	-2.81	***MA(2,12)***	15.95
SVAR	8.75	0.00	2.23	***MA(3,9)***	12.35
BM	-4.41	-3.73	0.78	***MA(3,12)***	8.73
NTIS	-7.75	-7.73	-1.00	***MOM(9)***	9.27
TBL	-1.37	-1.37	-2.75	***MOM(12)***	9.10
LTY	0.29	0.29	3.70	***VOL(1,9)***	12.04
LTR	6.57	6.57	1.99	***VOL(1,12)***	14.69
TMS	5.87	5.87	7.15	***VOL(2,9)***	9.48
DFY	-3.75	-3.18	0.24	***VOL(2,12)***	7.72
DFR	0.44	3.68	-3.87	***VOL(3,9)***	5.91
INFL	-0.20	-0.20	0.60	***VOL(3,12)***	12.12

From the Table 16, it is evident that the economic benefits generated by technical indicators are substantially higher than those of other individual explanatory variables during recession periods. In particular, ***MA(1, 12)*** and ***MA(2, 12)*** are very high. Positive and relatively high economic benefits are found in all model families, but negative values are also present among macroeconomic predictors. The observations support the argument that stock market forecasting heavily depends on the prevailing macroeconomic environment.

Table 17 presents CER gains for multiple forecast models during full forecast evaluation period.

Table 17. CER gains for different forecasting models. Full forecast evaluation period.

Panel A presents CER gains for full forecast evaluation period for multiple-predictor models with macroeconomic predictors (second, third and fourth column) and technical indicators (fifth column). Panel B presents CER gains for the same period using TECH COMB forecasts (sixth column).

Δ Overall						
Panel A	Unrestricted	CT	PAN	TECH	Panel B	COMB
Kitchen sink	1.20	1.20	-1.49	1.34	***Comb(1,9)***	1.23
BIC	-1.95	-1.95	-1.13	1.94	***Comb(1,12)***	2.03
POOL-AVG	-0.03	-0.03	0.19	1.35	***Comb(2,9)***	1.03
POOL-DMSFE	-0.11	-0.11	0.18	1.36	***Comb(2,12)***	2.19
Diffusion indices	-0.65	-0.65	0.09	1.72	***Comb(3,9)***	0.42
Sum-of-the-parts	0.59	0.59			***Comb(3,12)***	1.58

Table 17, it is once again observed that forecasts formed using technical indicators, this time with any model specification, generate large positive economic benefits compared to the historical average. The same cannot be said for all other models. Surprisingly, the Kitchen Sink model performs quite well, even though it was statistically weak. The results, however, supports the need for using technical indicators in equity premium forecasts, not only as individual explanatory variables in bivariate forecasts, but with larger models as well. It is also observed that the models generating the highest *CER* gains are combinations of technical indicators where s=1 and l=12, and s=2 and l=12.

Table 18 presents the CER gains for multiple-predictor models for NBER expansion periods.

Table 18. CER gains for different forecasting models. Business cycle expansion periods. Panel A presents CER gains for expansion periods for multiple-predictor models with macroeconomic predictors (second, third and fourth column) and technical indicators (fifth column). Panel B presents CER gains for the same period using TECH COMB forecasts (sixth column).

Δ Expansion						
Panel A	Unrestricted	CT	PAN	TECH	Panel B	COMB
Kitchen_sink	1.76	1.76	-1.11	-1.12	***Comb(1,9)***	-1.11
BIC	-1.48	-1.48	0.16	-0.26	***Comb(1,12)***	-0.50
POOL-AVG	0.79	0.79	0.32	-0.32	***Comb(2,9)***	-0.95
POOL-DMSFE	0.87	0.87	0.32	-0.32	***Comb(2,12)***	0.25
Diffusion indices	0.43	0.43	0.88	-0.31	***Comb(3,9)***	-1.43
Sum-of-the-parts	1.48	1.48			***Comb(3,12)***	-0.27

From Table 18, it can be observed that the CER gains for forecasts implemented using macroeconomic predictors are now predominantly positive, unlike those for technical indicators or their combinations. It appears that by using macroeconomic predictors and some of the presented forecasting models, higher economic benefits than the historical average can be achieved. Among the technical indicators, only ***Comb(2, 12)*** generates positive CER gains, but these are significantly than those produced by macroeconomic predictors. Of these, only the BIC model produces lower CER gains than the best-performing technical indicator.

Table 19 presents the CER gains for multiple-predictor models during NBER recession periods.

Table 19. CER gains for different forecasting models. Business cycle recession periods.

Panel A presents CER gains for recession periods for multiple-predictor models with macroeconomic predictors (second, third and fourth column) and technical indicators (fifth column). Panel B presents CER gains for the same period using TECH COMB forecasts (sixth column).

Δ Recession						
Panel A	Unrestricted	CT	PAN	TECH	Panel B	COMB
Kitchen_sink	4.97	4.97	1.22	16.45	***Comb(1,9)***	15.69
BIC	1.19	1.19	8.00	15.48	***Comb(1,12)***	17.71
POOL-AVG	5.80	5.80	1.20	11.65	***Comb(2,9)***	13.25
POOL-DMSFE	6.88	6.88	1.25	11.76	***Comb(2,12)***	14.20
Diffusion indices	7.04	7.04	5.66	14.23	***Comb(3,9)***	11.76
Sum-of-the-parts	6.94	6.94			***Comb(3,12)***	12.89

In Table 19, the CER gains for various multiple-predictor models using macroeconomic predictors, technical indicators, and combinations of technical indicators are presented. From the table, it can be observed that technical indicators clearly produce the highest CER gains. Of these, the ***Comb(1, 12)*** model yields the highest. Models using macroeconomic predictors also generate exclusively positive economic benefits, albeit significantly weaker than technical indicators, further indicating that technical indicators perform better than macroeconomic indicators when predicting recession periods.

7.3 Discussion

In bivariate forecasts, the regression coefficients of technical indicators are generally more stable and have smaller absolute values than those of macroeconomic predictors, which may partly explain the superiority of technical indicators over macroeconomic predictors. Figure 6 displays the time series of recursive expanding window regression coefficients for the ***MA(2, 12)*** indicator and ***DP***.

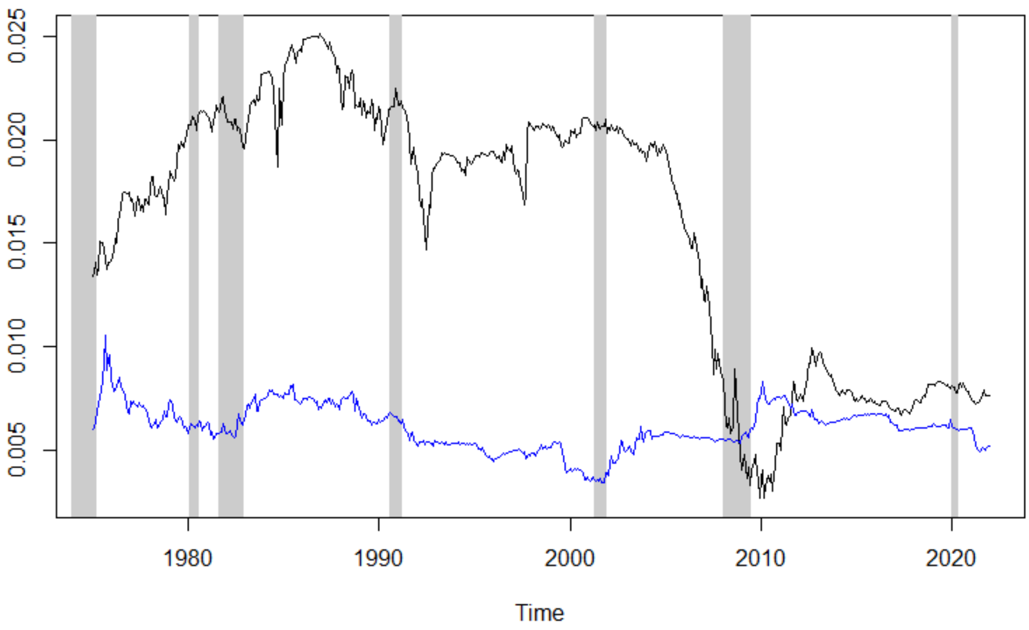

Figure 6. ***MA(2, 12)*** indicator (blue line) and ***DP*** (black line) regression coefficients. Forecast evaluation period. Grey shades present NBER recession periods.

In Figure 6, it can be observed that the ***DP*** regression coefficient (black line) varies much more broadly than the one for the ***MA(2, 12)*** technical indicator (blue line). The ***DP*** coefficient is relatively large before the financial crisis, after which it drops significantly but then rises sharply again in the 2010s. In contrast, the ***MA(2, 12)*** indicator's regression coefficient fluctuates relatively stable on both sides of its average coefficient estimate. Similar conclusions can be drawn from the regression coefficients of other bivariate model variables (See Appendix 2. Bivariate Regression Coefficients). The regression coefficients of technical indicators are of a much lower magnitude, resulting in forecasts that do not deviate nearly as much from those produced by the historical average model as those made with macroeconomic predictors.

In Figure 7, the time series of forecasts made using macroeconomic predictors with the POOL-DMSFE is presented.

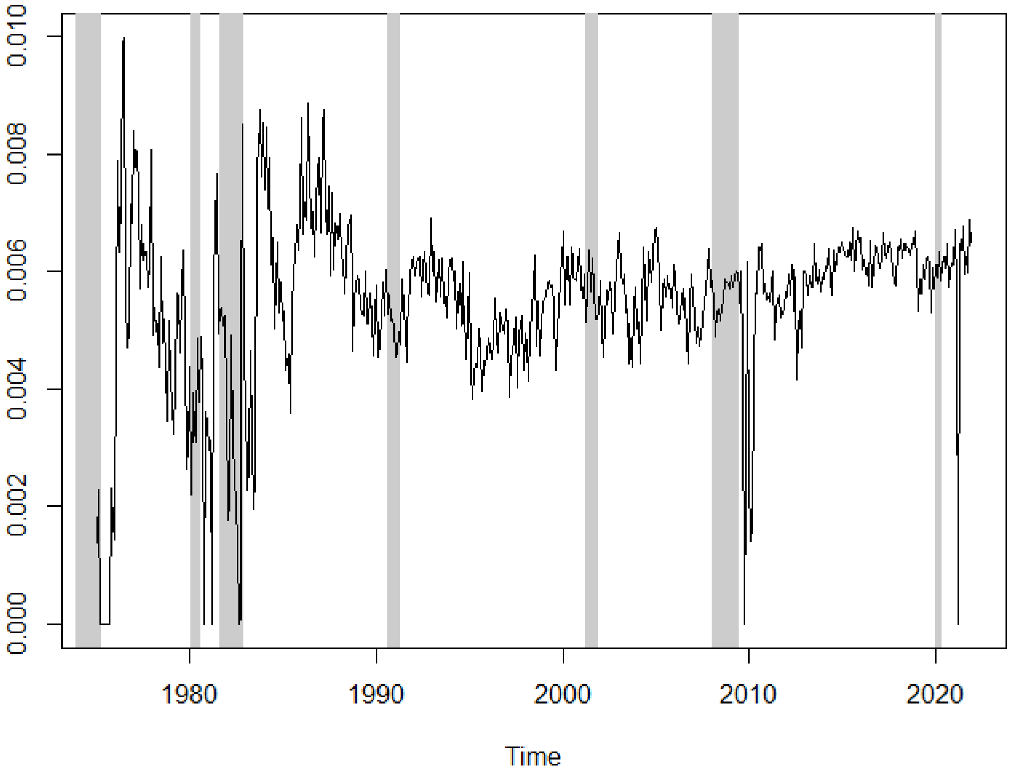

Figure 7. POOL-DMSFE with macroeconomic predictors equity premium forecast (CT restriction implemented) time series. Forecast evaluation period. Grey shades represent NBER recession periods.

In Figure 7, it can be observed that the POOL-DMSFE forecasts made with macroeconomic predictors have a very noisy nature. The variation in forecasts is strongest at the beginning of the forecast evaluation period. As we move into the 90s, the forecasts concentrate quite tightly around 0.005 on both sides. The impact of NBER recession periods on the forecasts is not apparent.

In Figure 8, the time series of forecasts made using technical indicators with the POOL-DMSFE is presented.

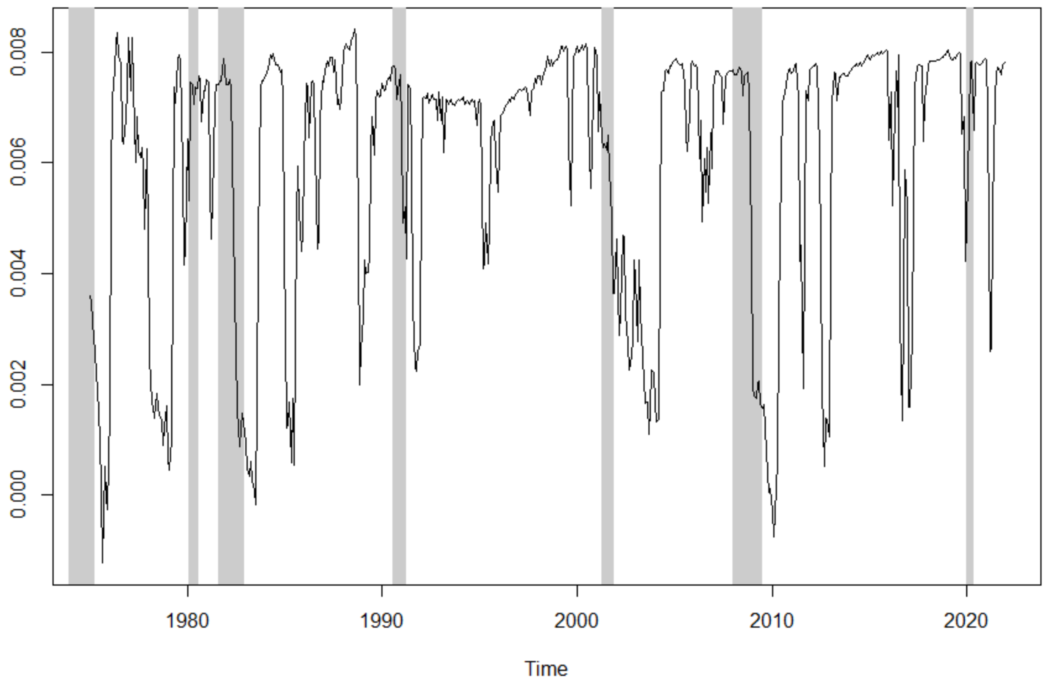

Figure 8. POOL-DMSFE with technical indicators equity premium forecast time series. Forecast evaluation period. Grey shades represent NBER recession periods.

From Figure 8, it can be observed how the nature of technical indicators is reflected in the forecasts made using them. The minimum and maximum values of the forecasts do not differ much from the forecasts made with macroeconomic predictors, but their nature is clearly less noisy. Unlike with macroeconomic predictors, the forecasts are noticeably more persistent. High values are often followed by several high forecasts, whereas forecasts made with macroeconomic predictors tend to oscillate more frequently on both sides of the average forecast. It can be observed that during NBER recession periods, the forecasts often decrease significantly.

Although technical indicators seem to produce better forecasts than macroeconomic predictors, it does not necessarily mean that they provide better information about future returns. The superiority of technical indicators may be due to their more stable nature compared to macroeconomic predictors. Low regression coefficients for technical indicators imply that the constant term plays a significant role in the forecasts produced by them, which cannot deviate much from the historical average forecast. Technical indicators, therefore, fine-tune the historical average forecasts with a small, stable information addition. The forecasts implied by technical indicators are quite similar to each other, which is not surprising based on the Table 2 showing strong mutual correlations. This is also evident in the regression coefficients (see Appendix 2. Bivariate Regression Coefficients), which appear to follow a very similar pattern.

Results show that combining technical indicators using various models does not seem to yield substantial added value. Conversely, employing macroeconomic predictors in conjunction with more comprehensive models often results in appreciable benefits. This can be attributed to the inherent stability of technical indicators, their coefficients, and consequently, the forecasts generated, which is in stark contrast to the situation with macroeconomic predictors. Rapach and Zhou (2013) propose that by adopting appropriate model specifications that factor in uncertainty and instability, it is possible to derive considerably more precise forecasts. This approach is indispensable for macroeconomic predictors, unlike the case with technical indicators.

The conclusions of the economic significance analysis do not change when using other risk aversion parameters instead of γ=5. The conclusions regarding statistical significance or economic significance remain the same when examining different expanding window sizes or different sample sizes.

8 Conclusions

In this thesis, I have conducted a review of recent equity premium forecasting and introduced the background theory and concepts relevant to this forecasting. The forecasting methods and evaluation methods presented are widely applicable to other economic applications. In particular, the methods used provide an overview of out-of-sample forecasting methodology in financial econometric literature.

Equity premium forecasts support existing literature in many ways. It is observed that forecasting during economic expansions is particularly challenging. Well-established macroeconomic predictors in the literature are unable to produce statistically accurate forecasts or forecasts from which investors could benefit economically during expansion periods. Technical indicators also yield significantly weaker economic benefits in expansions compared to the historical average model. However, investors could benefit economically from macroeconomic predictors by using them together in broader forecasting models, even during economic expansions. On the other hand, forecasting during recessions is not as hard as forecasting during expansions. Technical indicators provide the best forecasts both statistically and economically during recession periods and during full forecast evaluation period. Combining technical indicators or including them in broader models does not seem to make a significant difference due to their stable nature.

The results of the study suggests that an optimal approach could involve the deployment of multiple-predictor models that use macroeconomic predictors during periods of economic expansion, and individual technical indicators during recessions. This contribution to the discourse on equity premium forecasting advocates for a state-dependent forecasting methodology. One could use different models or predictors depending on the current economic state, a recession or expansion. By doing so, it might be possible to enhance the forecasting accuracy and economic benefits derived from the predictions. This state-dependent forecasting methodology could potentially provide more reliable and robust results for investors and policymakers who need to make informed decisions based on the expected equity premium. Further research could focus on developing and testing state-dependent forecasting models and identifying the most suitable predictors for each economic state.

It should indeed be noted from the forecast results that the historical average of the equity premium is positive, while in a recession, equity premium is generally negative. Therefore, instead using historical average as benchmark model during both recession and expansion periods, it might be appropriate to compare the equity premium separately in recessions and expansions using their state-dependent historical averages as benchmark models. This approach could provide a more sensible benchmark for comparing different forecasting models.

Referencess

Baker, M., & Wurgler, J. (2000). The Equity Share in New Issues and Aggregate Stock Returns. *The Journal of Finance, 55*(5), 2219-2257.

Baker, M., & Wurgler, J. (2006). Investor Sentiment and the Cross-Section of Stock Returns. *The Journal of Finance, 61*(4), 1645-1680.

Baker, M., & Wurgler, J. (2007). Investor Sentiment in the Stock Market. *Journal of Economic Perspectives, 21*(2).

Ball, R. (1978). Anomalies in relationships between securities' yields and yield-surrogates. *Journal of Financial Economics, 6*(2-3), 103-126.

Blume, L., Easley, D., & O'Hara, M. (1994). Market Statistics and Technical Analysis: The Role of Volume. *The Journal of Finance, 49*(1), 153-181.

Brockwell, P. J., & Davis, R. A. (2016). *Introduction to Time Series and Forecasting.* Springer.

Brown, D. P., & Jennings, R. H. (1989). On Technical Analysis. *The Review of Financial Studies, 2*(4), 527-551.

Cambpell, J. Y., & Shiller, R. J. (1988a). Stock Prices, Earnings, and Expected Dividends. *The Journal of Finance, 43*(3), 661-676.

Cambpell, J. Y., & Shiller, R. J. (1988b). The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors. *The Review of Financial Studies, 1*(3), 195-228.

Campbell, J. Y. (1987). Stock returns and the term structure. *Journal of Financial Economics, 18*(2), 373-399.

Campbell, J. Y., & Thompson, S. B. (2008). Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? *The Review of Financial Studies, 21*(4), 1509-1531.

Campbell, J. Y., & Viceira, L. M. (2002). *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors.* Oxford University Press.

Campbell, J. Y., & Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of Financial Economics, 81*(1), 27-60.

Cochrane, J. H. (1991). Production-Based Asset Pricing and the Link Between Stock Returns and Economic Fluctuations. *The Journal of Finance, 46*(1), 209-237.

Cochrane, J. H. (1997). Where is the market going? Uncertain facts and novel theories. *Federal Reserve Bank of Chicago - Economic Perspectives, 21*(6), 3-37.

Cochrane, J. H. (2005). *Asset Pricing* (Revised ed.). Princeton University Press.

Cootner, P. (1964). *The Random Character of Stock Market Prices.* The MIT Press.

Dangl, T., & Halling, M. (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics, 106*(1), 157-181.

De Long, B. J., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise Trader Risk in Financial Markets. *Journal of Political Economy, 98*(4), 703-738.

Diebold, F., & Mariano, R. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics, 13*(3), 253-263.

Doan, B., & Lan, C. (2022). Stock price movements: Evidence from global equity markets. *Journal of Empirical Finance, 69*, 123-143.

Edmans, A., Goldstein, I., & Jiang, W. (2015). Feedback Effects, Asymmetric Trading, and the Limits to Arbitrage. *American Economic Review, 105*(12), 3766-3797.

Efron, B., Burman, P., Denby, L., Landwehr, J. M., Mallows, C. L., SHen, X., . . . Zhang, C. (2004). The Estimation of Prediction Error: Covariance Penalties and Cross-Validation. *Journal of the American Statistical Association, 99*(467), 619-642.

Fama, E. F. (1965). The Behavior of Stock-Market Prices. *The Journal of Business, 38*(1), 34-105.

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *25*(2), 383-417.

Fama, E. F. (1981). Stock Returns, Real Activity, Inflation, and Money. *The American Economic Review, 71*(4), 545-565.

Fama, E. F., & French, K. R. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics, 22*(1), 3-25.

Fama, E. F., & French, R. K. (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics, 25*(1), 23-49.

Fama, E. F., & Schwert, G. W. (1977). Asset returns and inflation. *Journal of Financial Economics, 5*(2), 115-146.

Ferreira, M. A., & Santa-Clara, P. (2011). Forecasting stock market returns: The sum of the parts is more than the whole. *Journal of Financial Economics, 100*(3), 514-537.

Franke, J., Härdle, W. K., & Hafner, C. M. (2019). *Statistics of Financial Markets: An Introduction.* Springer.

Friedman, J., Hastie, T., & Tibshirani, R. (2017). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer.

Gallant, R. A. (1987). *Nonlinear Statistical Models.* Wiley.

Goyal, A. (n.d.). Retrieved September 1, 2022, from https://sites.google.com/view/agoyal145

Goyal, A., & Welch, I. (2008). A Comprehensive Look at The Empirical Performance of Equity Premium Prediction. *The Review of Financial Studies, 21*(4), 1455–1508.

Greene, W. H. (2003). *Econometric Analysis.* Prentice Hall.

Grundy, B. D., & McNichols, M. (1989). Trade and the Revelation of Information through Prices and Direct Disclosure. *The Review of Financial Studies, 2*(4), 495-526.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies, 33*(5), 2223-2273.

Guidolin, M., & Timmermann, A. (2007). Asset allocation under multivariate regime switching. *Journal of Economic Dynamics and Control, 31*(11), 3503-3544.

Guo, H. (2006). On the Out-of-Sample Predictability of Stock Market Returns. *The Journal of Business, 79*(2), 645-670.

Henkel, S., Nardari, F., & Spencer, M. J. (2011). Time-varying short-horizon predictability. *Journal of Financial Economics, 99*(3), 560-580.

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.

Hong, H., & Stein, J. C. (1999). A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets. *The Journal of Finance, 54*(6), 2143-2184.

Keim, D. B., & Stambaugh, R. F. (1986). Predicting returns in the stock and bond markets. *Journal of Financial Economics, 17*(2), 357-390.

Kelly, B., & Pruitt, S. (2013). Market Expectations in the Cross-Section of Present Values. *The Journal of Finance, 68*, 1721-1756.

Keynes, J. M. (1936). *The General Theory of Employment, Interest and Money.* Macmillan.

Konishi, S., & Kitagawa, G. (2008). *Information Criteria and Statistical Modeling.* Springer.

Kothari, S. P., & Shanken, J. (1997). Book-to-market, dividend yield, and expected market returns: A time-series analysis. *Journal of Financial Economics, 44*(2), 169-203.

Lamont, O. (1998). Earnings and Expected Returns. *The Journal of Finance, 53*(5), 1563-1587.

Lamoureux, C. G., & Zhou, G. (2015). Temporary Components of Stock Returns: What Do the Data Tell Us. *The Review of Financial Studies, 9*(4), 1033-1059.

Lettau, M., & Ludvigson, S. (2001). Consumption, Aggregate Wealth, and Expected Stock Returns. *The Journal of Finance, 56*(3), 815-849.

Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics, 74*(2), 209-235.

Lintner, J. (1965). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *The Review of Economics and Statistics, 47*(1), 13-37.

Lintner, J. (1975). Inflation and security returns. *The Journal of Finance, 30*(2), 259-280.

Ludvigson, S., & Ng, S. (2007). The empirical risk-return relation: A factor analysis approach. *Journal of Financial Economics, 83*(1), 171-222.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis.* Springer.

Malkiel, B. G. (2003). The Efficient Market Hypothesis and Its Critics. *Journal of Economic Perspectives, 17*(1), 59-82.

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance, 7*(1), 77-91.

Mele, A. (2007). Asymmetric stock market volatility and the cyclical. *2*.

Menzly, L., Veronesi, P., & Santos, T. (2004). Understanding Predictability. *Journal of Political Economy, 112*(1), 1-47.

Moskowitz, T. J., Ooi, Y. H., & Pedersen, L. H. (2012). Time Series Momentum. *Journal of Financial Economics, 104*(2), 228-250.

Mossin, J. (1966). Equilibrium in a Capital Asset Market. *Econometrica, 34*(4), 768-783 .

Neely, C. J., Rapach, D., Tu, J., & Zhou, G. (2014). Forecasting the Equity Risk Premium: The Role of Technical Indicators. *Management Science, 60*, 1772-1791.

Pan, Z., Pettenuzzo, D., & Yudong, W. (2020). Forecasting stock returns: A predictor-constrained approach. *Journal of Empirical Finance, 55*(C), 200-217.

Pesaran, M. H., & Timmermann, A. (1995). Predictability of Stock Returns: Robustness and Economic Significance. *The Journal of Finance, 50*(4), 1201-1228.

Pettenuzzo, D., Timmermann, A., & Valkanov, R. (2014). Forecasting stock returns under economic constraints. *Journal of Financial Economics, 114*(3), 517-533.

Polk, C., Thompson, S., & Vuolteenaho, T. (2006). Cross-sectional forecasts of the equity premium. *Journal of Financial Economics, 81*(1), 101-141.

Pontiff, J., & Schall, L. D. (1998). Book-to-market ratios as predictors of market returns. *Journal of Financial Economics, 49*(2), 141-160.

Rapach, D. E. (2014). *Dave Rapach*. Retrieved September 1, 2022, from https://sites.google.com/slu.edu/daverapach/publications?authuser=0

Rapach, D. E., & Zhou, G. (2013). Chapter 6 - Forecasting Stock Returns. *Handbook of Economic Forecasting, 2*(Part A), 328-383.

Rapach, D. E., Zhou, G., & Strauss, J. K. (2010). Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. *The Review of Financial Studies, 23*(2), 821-862.

Rozeff, M. S. (1984). Dividend Yields are Equity Risk Premiums. *Journal of Portfolio Management, 11*(1), 68-75.

Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk. *The Journal of Finance, 19*(3), 425-442.

Shiller, R. J. (1984). Stock prices and social dynamics. *Brookings Papers on Economic Activity, 2*, 457-498.

Soros, G. (2015). *The Alchemy of Finance.* Wiley.

Stock, J. H., & Watson, M. W. (2004). Combination forecasts of output growth in a seven-country dataset. *Journal of Forecasting, 23*, 405-430.

Stock, J. H., & Watson, M. W. (2008). *Phillips Curve Inflation Forecasts* (Vol. 14322). NBER Workin Paper Series.

Treynor, J. L., & Ferguson, R. (1985). In Defense of Technical Analysis. *The Journal of Finance, 40*(3), 757-773.

Vespa, G., & Vives, X. (2012). Dynamic Trading and Asset Prices: Keynes vs. Hayek. *The Review of Economic Studies, 79*(2), 539-580.

Welch, I. (2017). *Corporate Finance* (4th ed.). Ivo Welch.

Yahoo Finance. (n.d.). Retrieved September 1, 2022, from https://finance.yahoo.com/quote/%5EGSPC/history/

Appendices

Appendix 1. Basic Concepts

Random Variable

Let $(\Omega, \mathrm{F}, \mathrm{P})$ be a probability space where Ω is the set of all possible elementary events, F is the sigma-algebra of all events, and P is the probability measure defined on F. A random variable Y is defined as a real-valued function on Ω, such that for every real number c, the set $\mathrm{A_c} = \{\omega \in \Omega | \mathrm{Y}(\omega) \leq \mathrm{c}\}$ is in F. The function $F: R \to [0,1]$, where $F(\mathrm{c}) = P(\mathrm{A}_c)$, is called the cumulative distribution function (CDF) of the random variable Y. (Lütkepohl 2006, 2-3)

Expected Value of a Random Variable

The expected value, also known as the mathematical expectation or the mean, is a fundamental concept in probability and statistics. It is defined as the sum of the product of each outcome of a random variable and its corresponding probability. In mathematical terms, for a discrete random variable X with a probability mass function $p(x)$, the expected value can be calculated as:

$$E[X] = \sum x\, p(x). \tag{1}$$

For a continuous random variable with a probability density function $f(x)$, the expected value is calculated as:

$$E[X] = \int x\, f(x) dx. \tag{2}$$

(Franke et al. 2015. 40-41)

Variance and Covariance of a Random Variable

Variance is a statistical measure that describes the dispersion or spread of a set of data points around the mean or average value. Mathematically, the variance of a random variable X is defined as the expected value of the squared deviation of X from its mean, and is represented as $Var(X)$. Variance can be formulated as

$$Var(X) = E[(X - \mathrm{\mu})^2], \tag{3}$$

where μ represents the mean of X. (Franke et al. 2015, 40-41)

Covariance, on the other hand, is a measure of the joint variability of two random variables X and Y. It describes the extent to which the two variables change together. Positive covariance indicates that the variables tend to increase or decrease together, while negative covariance indicates that one variable tends to increase as the other decreases. Covariance can be presented as

$$Cov(X,Y) = E\left[(X - \mu_x)(Y - \mu_y)\right], \tag{4}$$

where μ_x and μ_y represent the means of X and Y, respectively.

Both variance and covariance are important concepts in statistics, finance, and other fields where data analysis and risk management are important. They provide a way to quantify and analyze the relationships between different variables and to assess the risk of investment portfolios. (Franke et al. 2015. 42-43)

Autocorrelation

Autocorrelation, also referred as serial correlation, is a statistical concept that measures the linear dependence between the values of a time series and its lagged values. In other words, autocorrelation refers to the extent to which the value of a time series at time t is correlated with its value at time *t-k*, where *k* is the lag.

Mathematically, autocorrelation is defined as the correlation coefficient between two sets of values, one being the original time series and the other being a lagged version of the same series. The autocorrelation coefficient at lag k, denoted as *ρ(k)*, can be constructed as:

$$\rho(k) = Cov(X_t, X_{t-k})/Var(X_t). \tag{5}$$

$Cov(X_t, X_{t-k})$ represents the covariance between the values at time *t* and *t-k*, and $Var(X_t)$ represents the variance of the original time series. If a time series has positive autocorrelation, it means that its values are positively related to their lagged values, and if a time series has negative autocorrelation, it means that its values are negatively related to their lagged values. (Brockwell and Davis 2016, 13-14)

Random Walk

Random walk is a mathematical concept that models a process where an object moves from one position to another over time based on random chance. The concept is widely used in finance, physics, and other fields to describe a variety of physical and abstract systems.

Mathematically, a random walk can be modelled as a sequence of random variables, where the value of each variable depends on the value of the previous one. For example, let $X_1, X_2, X_3, \dots, X_t$ be a sequence of random variables representing the movement of an object at discrete points in time. The value of X_t at time t can be defined as the sum of the previous values:

$$X_t = X_{t-1} + \varepsilon_t, \tag{3}$$

where ε_t is a random error term representing the deviation of the movement from the previous step. The random error terms can be modelled as independent and identically distributed (i.i.d.) random variables, such as normal or uniform distributions.

The cumulative sum of the random walk, also known as the random walk process, can be defined as:

$$S_t = X_1 + X_2 + \cdots + X_t \tag{4}$$

Random walk models are often used to simulate the behaviour of financial systems, such as the movement of stock prices, and to analyze the distribution of the cumulative sum over time. (Franke et al. 2015, 54-55)

Appendix 2. Bivariate Regression Coefficients

In this appendix, the time series of regression coefficients for different variables are presented for bivariate forecasts during the forecast evaluation period.

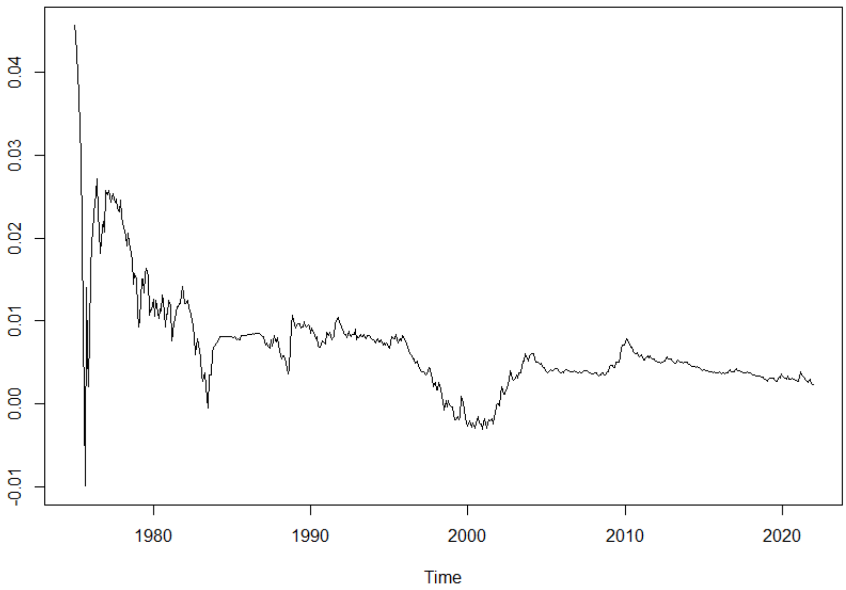

Figure A. 1. Regression coefficient of ***BM***

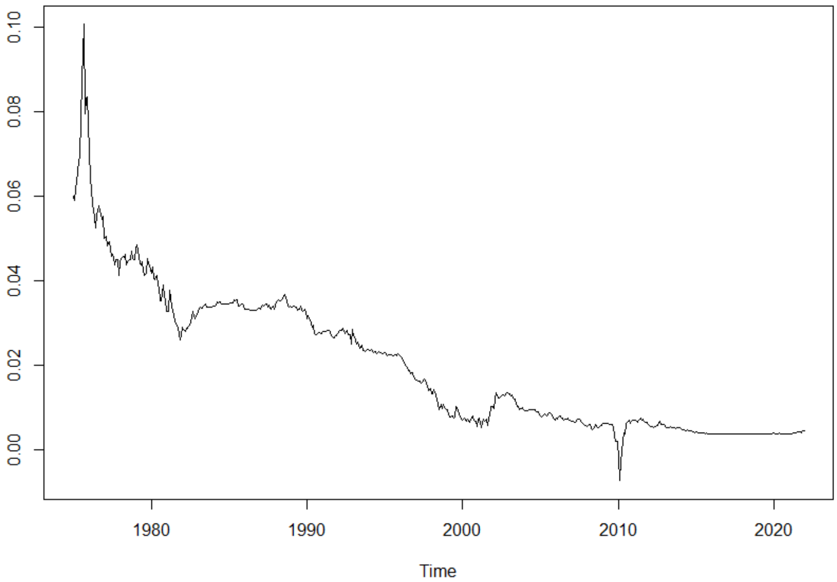

Figure A. 2. Regression coefficient of ***DE***

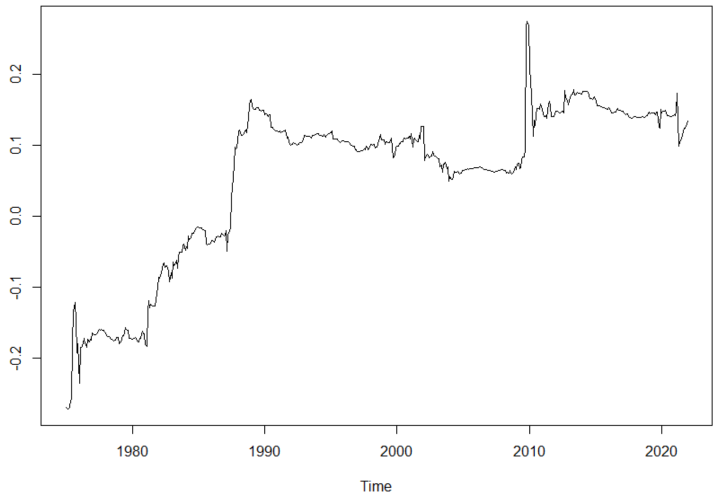

Figure A. 3. Regression coefficient of ***DFR***

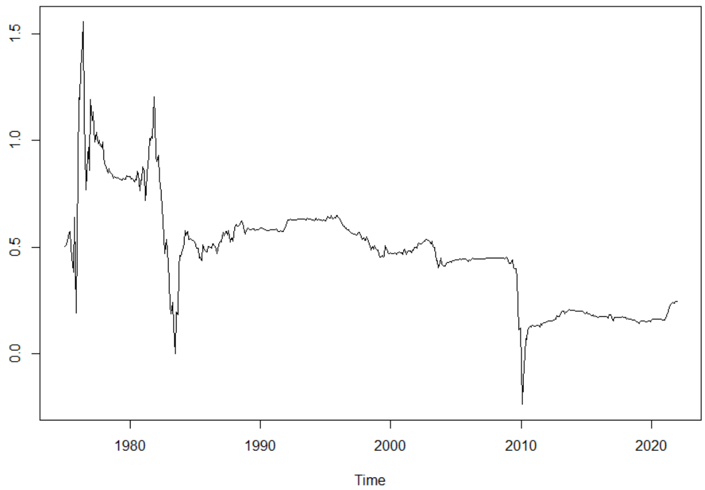

Figure A. 4. Regression coefficient of ***DFY***

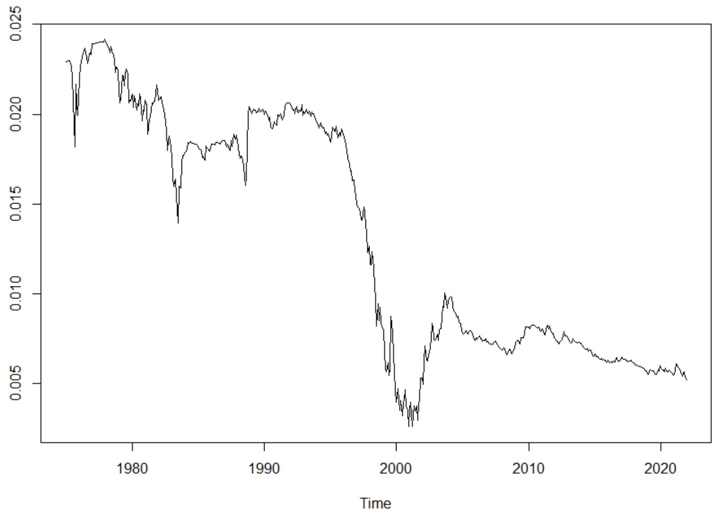

Figure A. 5. Regression coefficient of ***DP***

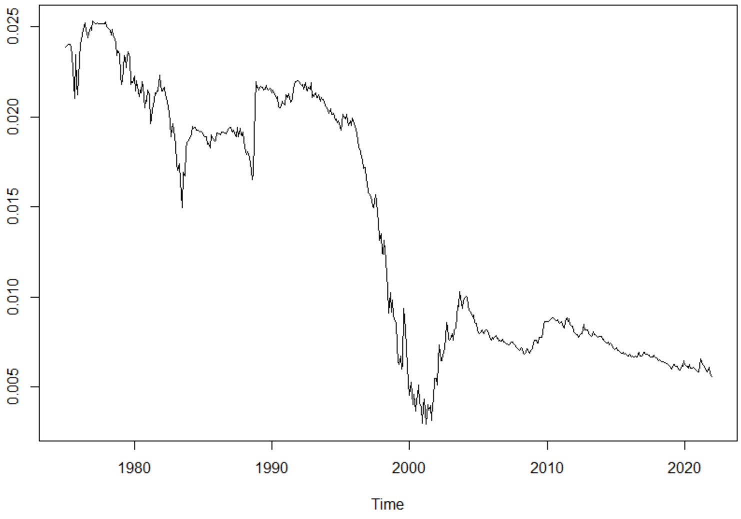

Figure A. 6. Regression coefficient of ***DY***

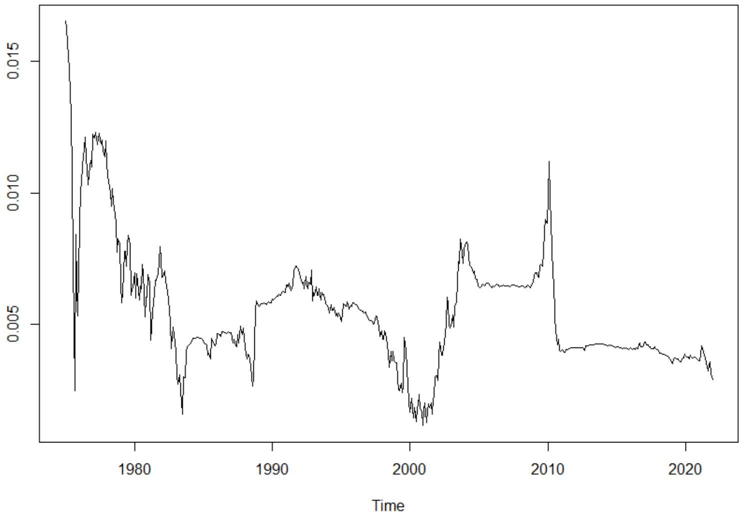

Figure A. 7. Regression coefficient of ***EP***

Figure A. 8. Regression coefficient of ***INFL***

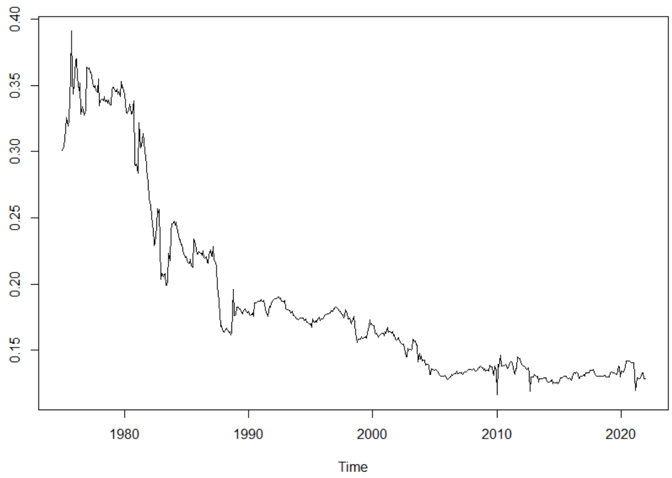

Figure A. 9. Regression coefficient of ***LTR***

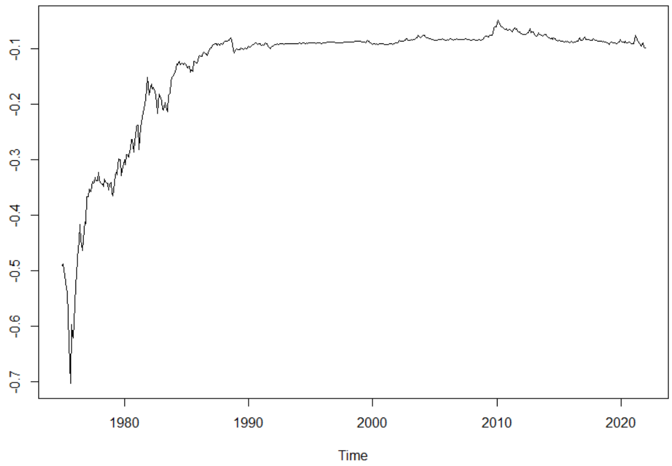

Figure A. 10. Regression coefficient of ***LTY***

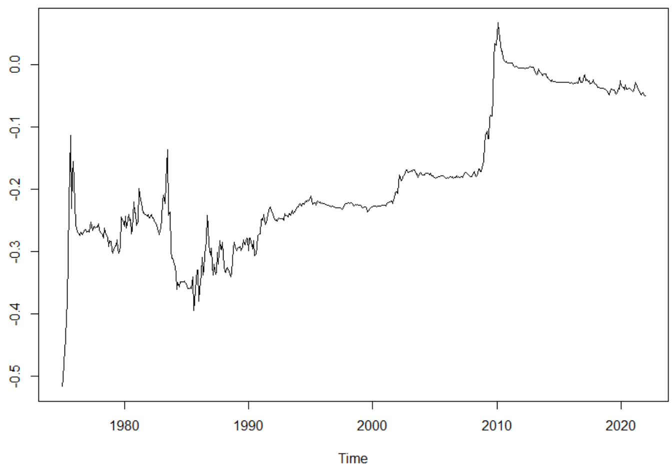

Figure A. 11. Regression coefficient of ***NTIS***

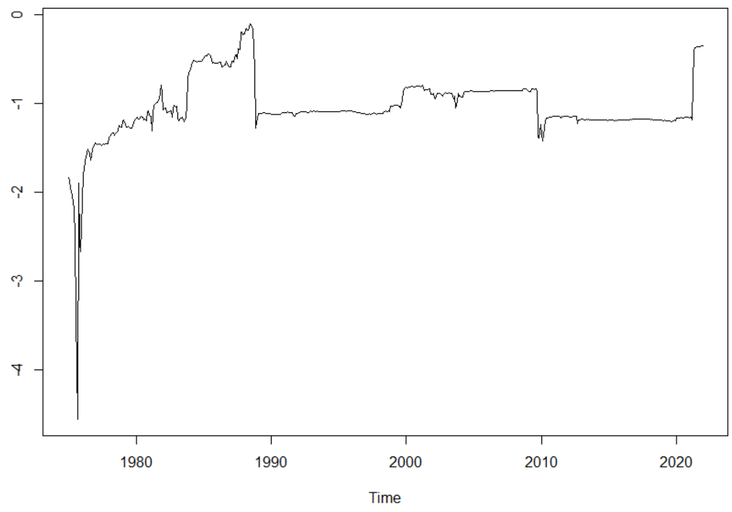

Figure A. 12. Regression coefficient of ***SVAR***

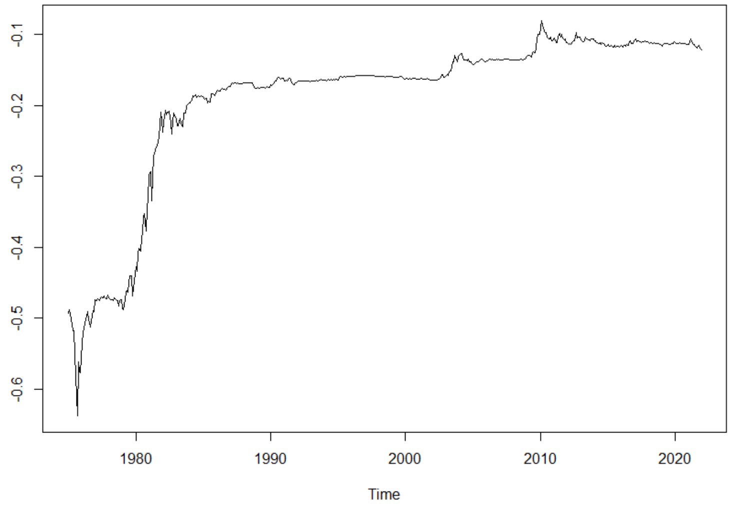

Figure A. 13. Regression coefficient of ***TBL***

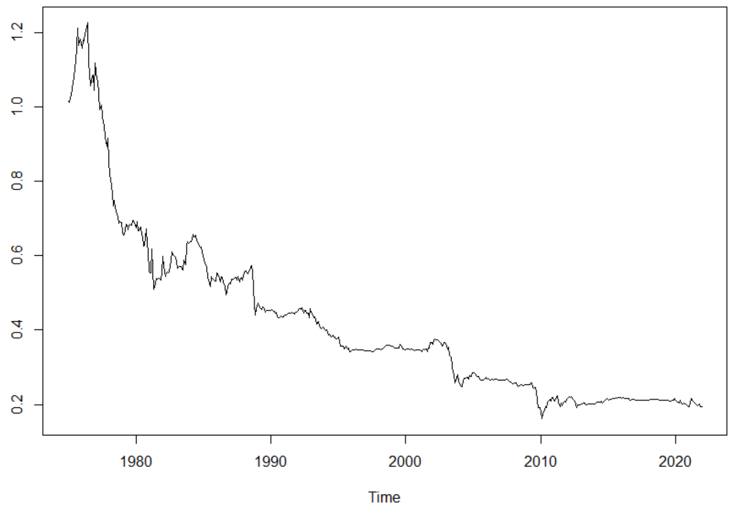

Figure A. 14. Regression coefficient of ***TMS***

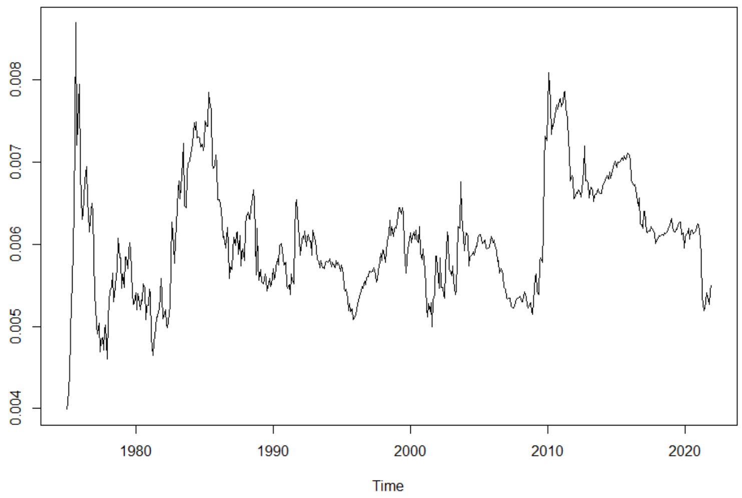

Figure A. 15. Regression coefficient of ***VOL(1, 9)***

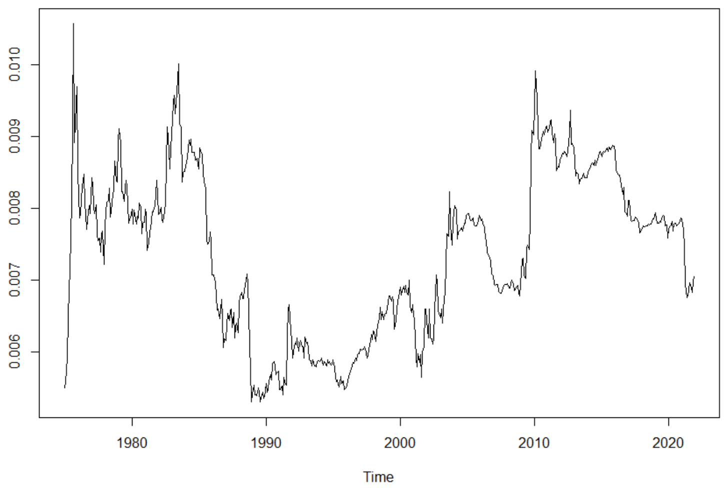

Figure A. 16. Regression coefficient of ***VOL(1, 12)***

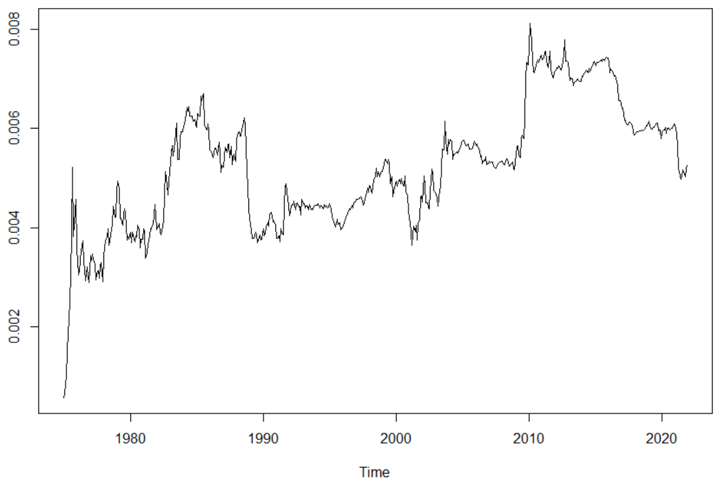

Figure A. 17. Regression coefficient of ***VOL(2, 9)***

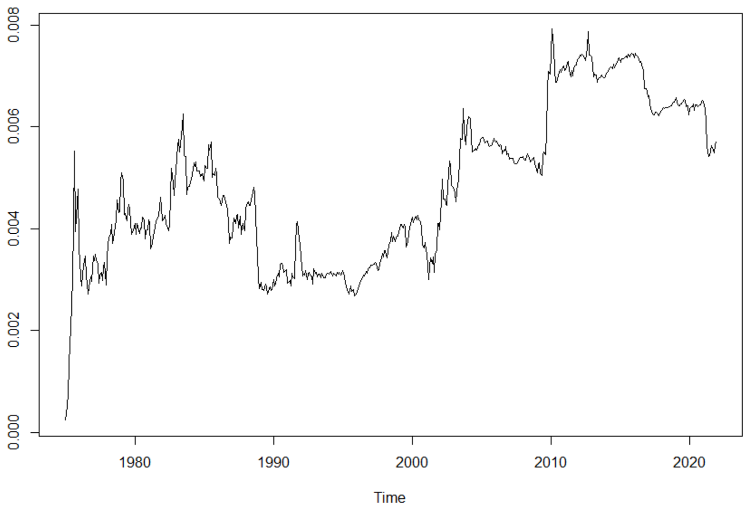

Figure A. 18. Regression coefficient of ***VOL(2, 12)***

Figure A. 19 Regression coefficient of ***VOL(3, 9)***

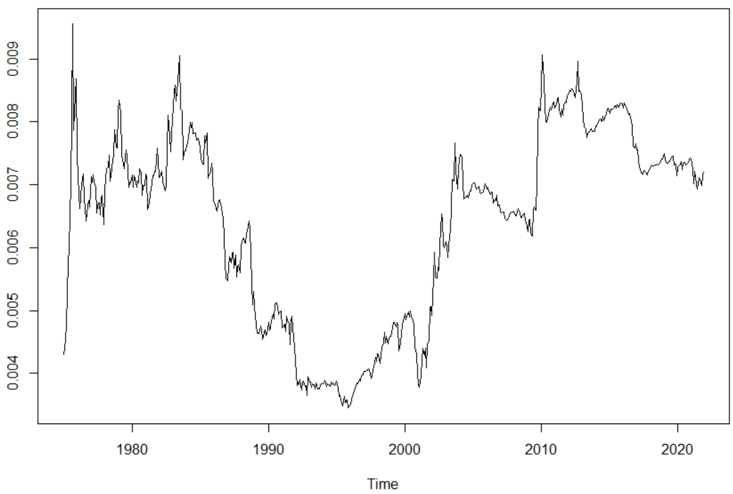

Figure A. 20. Regression coefficient of ***VOL(3, 12)***

Figure A. 21. Regression coefficient of ***MA(1, 9)***

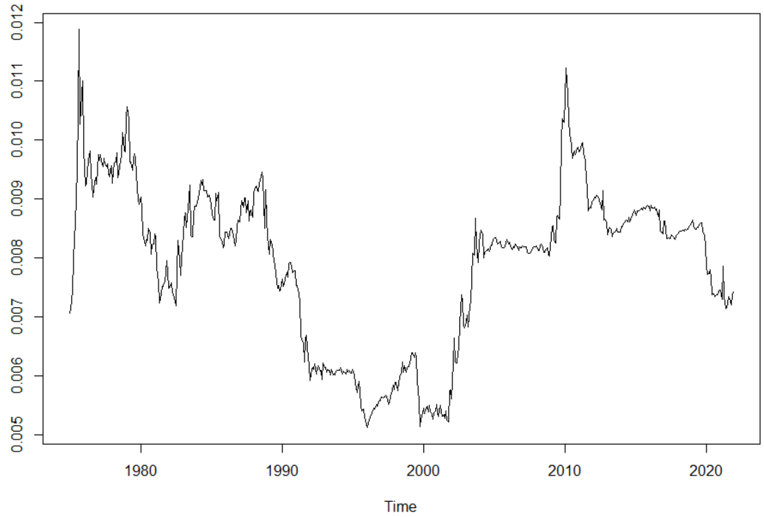

Figure A. 22. Regression coefficient of ***MA(1, 12)***

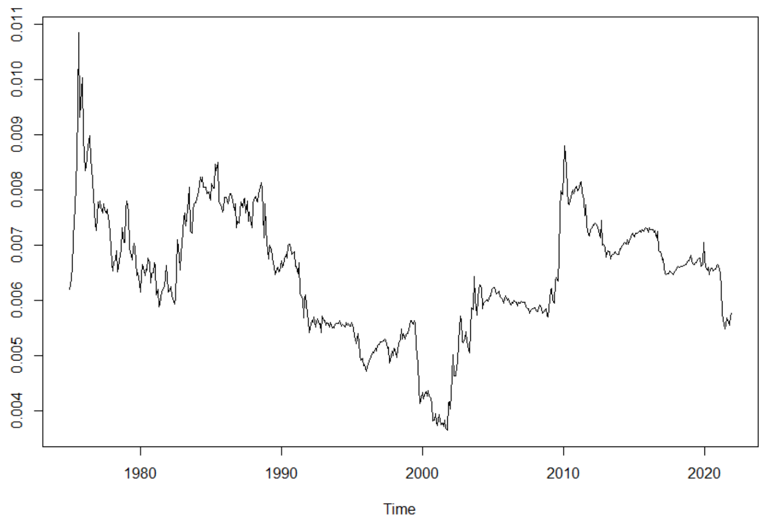

Figure A. 23. Regression coefficient of ***MA(2, 9)***

Figure A. 24. Regression coefficient of ***MA(2, 12)***

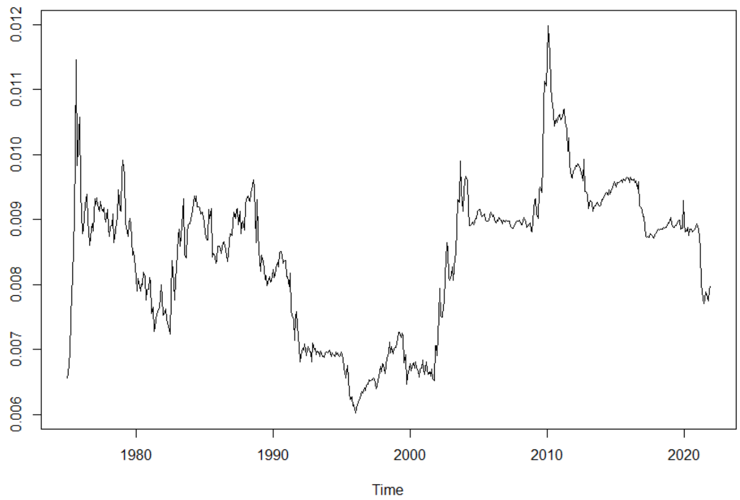

Figure A. 25. Regression coefficient of ***MA(3, 9)***

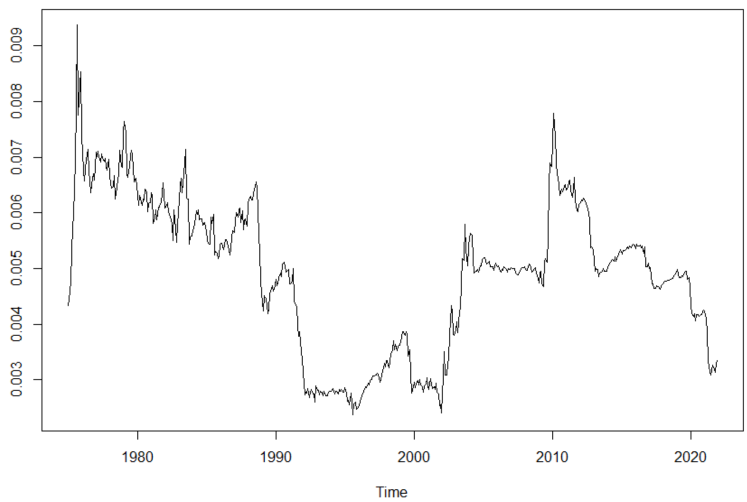

Figure A. 26. Regression coefficient of ***MA(3, 12)***

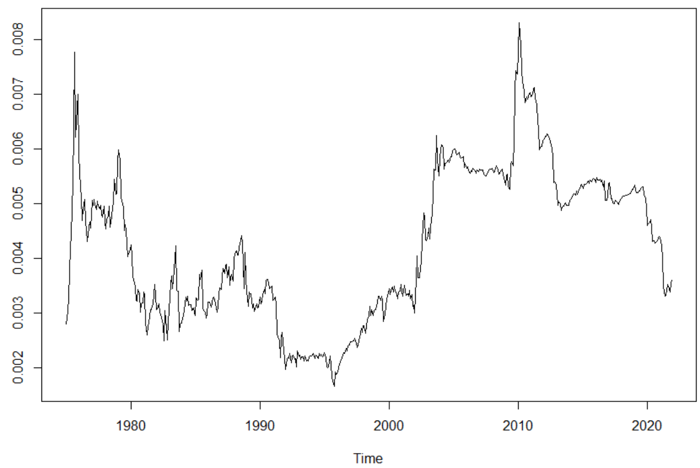

Figure A. 27. Regression coefficient of ***MOM(9)***

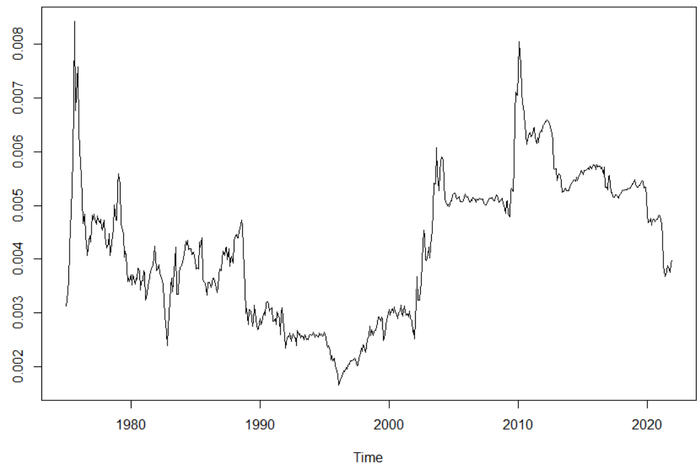

Figure A. 28. Regression coefficient of ***MOM(12)***

Appendix 3. Equity Premium Forecasts with Multiple-Predictor Models

In this appendix, the time series of equity premium forecasts for multiple-predictor models are presented for macroeconomic predictors with CT constraints and technical indicators during full forecast evaluation period.

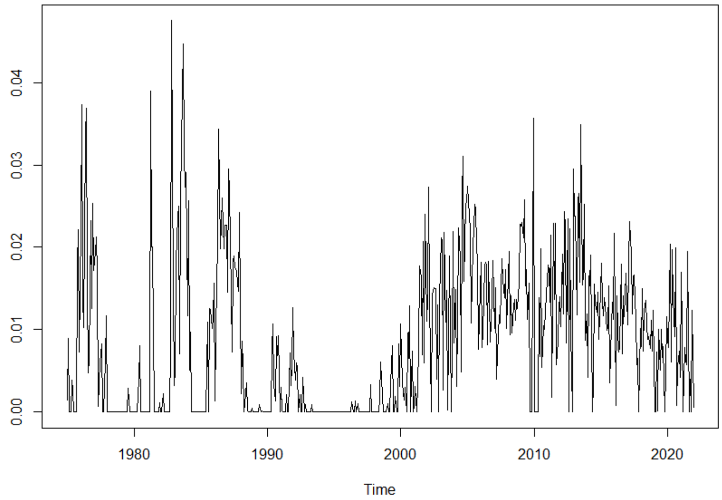

Figure A. 29. Kitchen sink forecast with macroeconomic predictors (CT restricted)

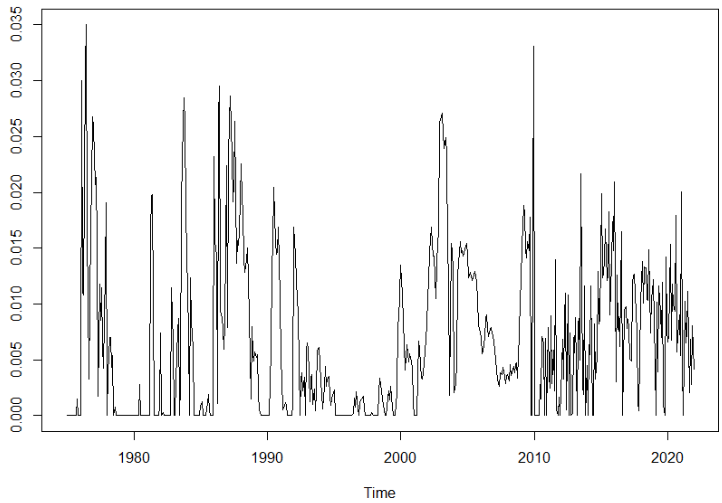

Figure A. 30. BIC forecast with macroeconomic predictors (CT restricted)

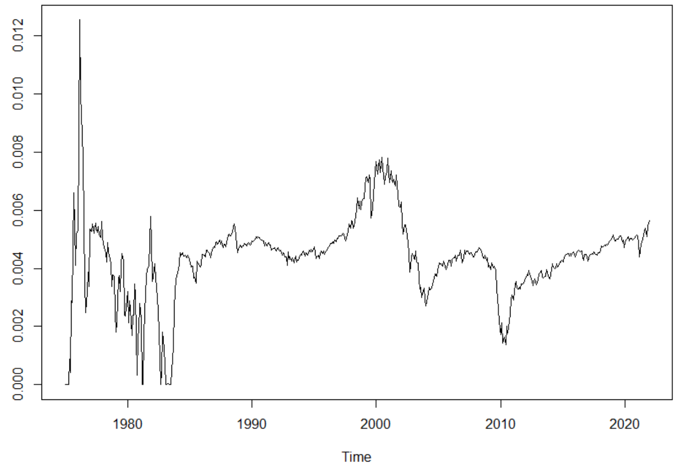

Figure A. 31. Sum-of-the-parts forecast with macroeconomic predictors (CT restricted)

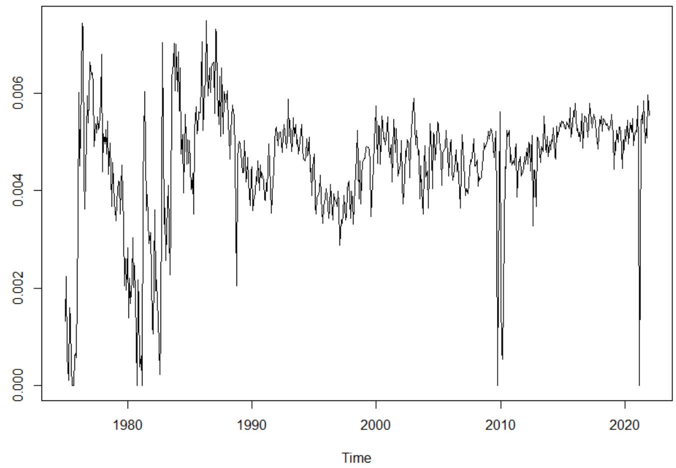

Figure A. 32. POOL-AVG forecast with macroeconomic predictors (CT restricted)

pool avg

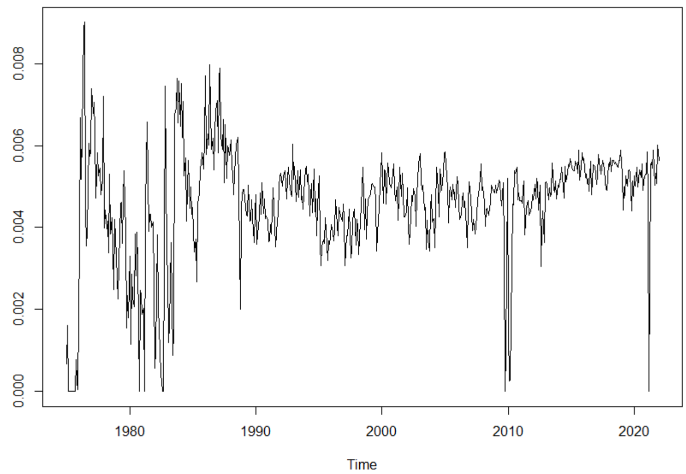

Figure A. 33. POOL-DMSFE forecast with macroeconomic predictors (CT restricted)

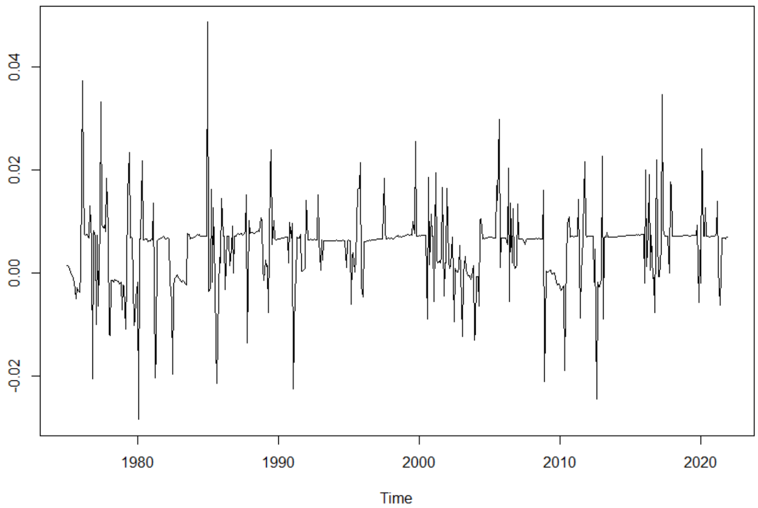

Figure A. 34. Kitchen sink forecast with technical indicators

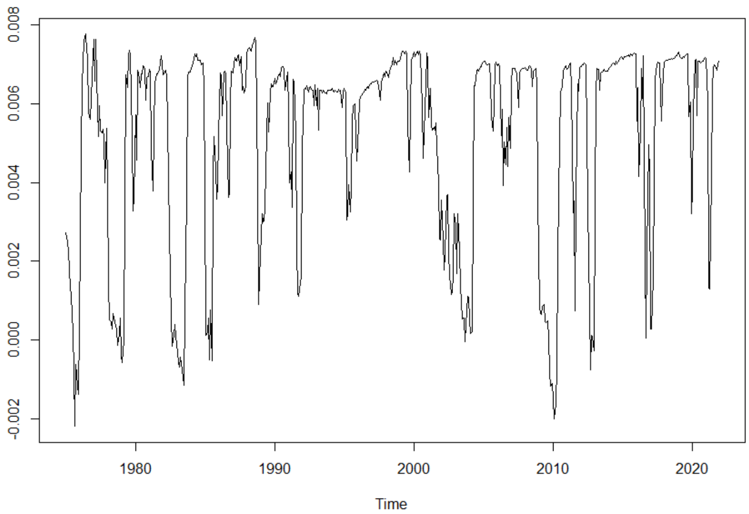

Figure A. 35. POOL-AVG forecast with technical indicators

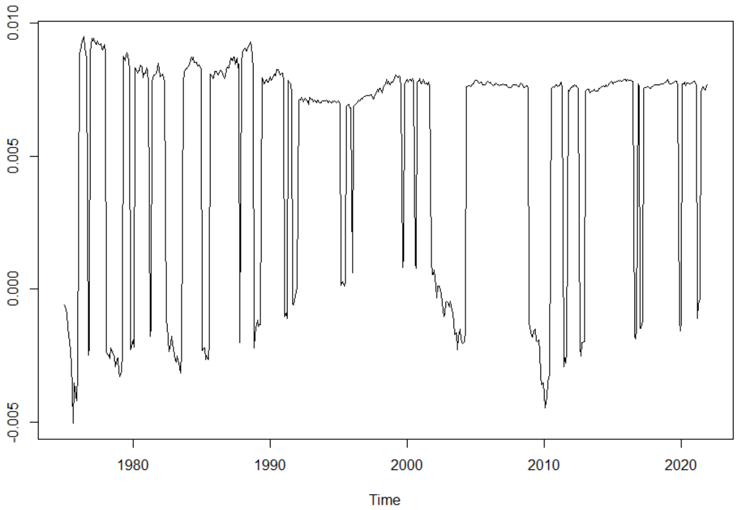

Figure A. 36. BIC forecast with technical indicators

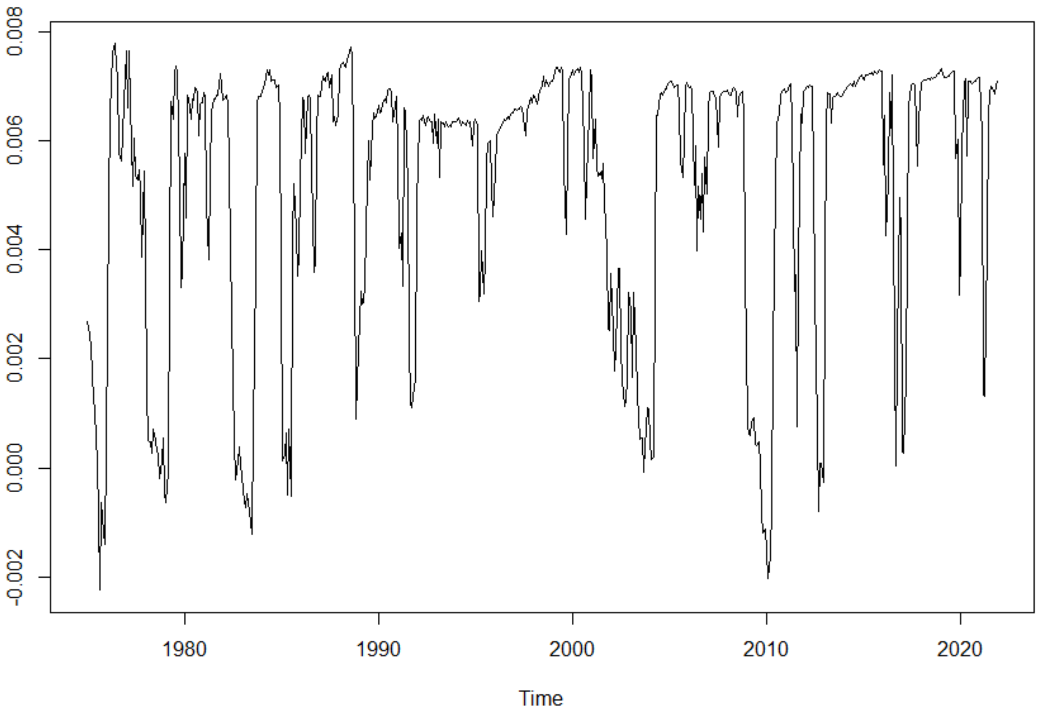

Figure A. 37. POOL-DMSFE forecast with technical indicators