ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Master Thesis MSc Econometrics and Management Science

Specialization: Quantitative Finance

Forecasting the Equity Premium in a Bayesian Setting

Extracting Predictive Information and Accounting for Regime-Dependency and Stochastic Volatility

Author: Evers, M.M.A.

Student ID Number: 406626

Supervisor: Scholtus, K.

Second Assessor: Prof. Dr. Franses, P.H.B.F.

6 September 2019

Abstract

The subject of this Master thesis is the predictability of the equity risk premium. It aims to increase the predictability by summarizing information from a large set of variables into a small number of factors. The variables include macroeconomic variables, technical indicators, and measures of investor sentiment. Technical indicators provide most predictive power at short forecast horizons, whereas sentiment measures provide complementary value at longer forecast horizons. A market timing study confirms that this predictability also translates into economic value when implemented as a trading strategy. Next, time-varying volatility and regime-dependent specifications of the equity premium's volatility process increase the predictability further, especially during the post-2008 period. Time-varying volatility performs especially well at short forecast horizons, whereas regime-dependent models yield additional improvements at longer forecast horizons. The thesis also tests the value of assuming regime-dependent relations between the predictors and the equity premium. This does not generate additional improvements. The results support the hypothesis that the countercyclical pattern of equity premium predictability is at least partially driven by the time-varying nature of the volatility process. It does not support the hypothesis that it is driven by a regime-dependent relation between the predictors and the risk premium. The research takes estimation uncertainty of all models into account by making use of Bayesian MCMC methods.

Key words: equity risk premium predictability, out-of-sample forecasts, asset allocation, business cycle, regime-dependency, stochastic volatility

Contents

1 Introduction

The Master thesis aims to improve the predictability of the equity risk premium in two ways. The first is making use of a large set of predictors of varying categories in an efficient manner. The second is accounting for the countercyclical nature of equity risk premium predictability by assuming regime-dependent parameters or time-varying volatility in the forecasting model.

Making accurate forecasts of the equity risk premium is a problem of key interest in the areas of asset management, market timing, and investment decision making in general. An increase in forecast accuracy allows market participants to improve their financial planning, reduce the uncertainty of future returns, and reduce the risk of market exposure. Although it is clear that market participants highly desire the improvement of predictability, the existence of this predictability is still a topic of debate. Initially Welch and Goyal (2007) conclude that the equity risk premium is generally not predictable in a linear setting. However, subsequent research shows that the right variable selection and different estimation approaches can increase the predictability.

One case for which predictability of the equity risk premium is present is when a relationship exists between current variables and the future risk premium. The literature yields a substantial number of these variables, corresponding to different categories. The category that is most often used includes financial variables, such as valuation ratios, the interest rate, and the credit spread (Ang & Bekaert, 2006; Campbell & Thompson, 2007). Both Bai (2010) and Çakmaklı and van Dijk (2016) expand on these findings by adding macroeconomic variables. They show that although earlier research does not yield positive out-of-sample forecasting performance when using macroeconomic variables, an effective remedy is the use of a much larger set of possible predictors. By summarizing the information into a limited number of factors via principal component analysis (PCA), they reduce the number of variables and improve forecasting performance. The gains result from the fact that the larger set of variables leads to much more consistent forecasting performance across different periods. In contrast, individual predictors often showcase varying degrees of forecasting power at different periods.

Whereas the two aforementioned categories focus on the state of the economy and its companies, one can also look at variables measuring the behaviour of stock prices and investors. A prime example is Neely et al. (2014), who condense a range of technical index price indicators into a single PCA factor. They show that the addition of these variables yields positive forecasting performance, both in and out-of-sample. A related approach is the use of investor sentiment proxies to predict the equity premium. Brown and Cliff (2004) show positive forecasting results for several proxies of investor sentiment when implementing this approach. Later, Baker and Wurgler (2007) expand the method by constructing a sentiment index that combines different proxies into a PCA factor. Inspired by the methods used by Kelly and Pruitt (2013), Huang et al. (2015) further improve the sentiment index of Baker and Wurgler (2007). By constructing the factors using the partial least squares (PLS) technique of Kelly and Pruitt (2013) instead of PCA, the out-of-sample forecasting ability of the sentiment index significantly improves. Aiming to replicate this improvement for other categories of factors as well, this thesis uses PLS to construct factors.

Across categories, it is also the case that the relative forecast accuracy of the predictive models displays a countercyclical pattern. It consists of higher relative accuracy during recessions, and lower relative accuracy during expansions.[1,2] The existence of the countercyclical equity premium predictability in linear models raises the question what the source of this phenomenon is. A possible explanation is that the relationship between the different categories of predictors and equity returns is in reality not linear but regime-dependent. A model that distinguishes between different stages of the business cycle, such as a Markov switching (MS) model, can account for this. Improved forecasting performance over linear models would suggest that regime-dependence is an important source of the countercyclical nature. Indeed, this is what Sander (2018) and Henkel et al. (2011) find, among others.

Another possible source of the countercyclical behaviour of equity risk premium predictability lies in the nature of the equity return's volatility process. According to Cont (2001), the clustering nature of volatility and the heavy-tailed nature of the returns are major empirical stylized facts of asset returns. Note that stock market volatility also closely follows the state of the business cycle, showing prolonged elevated levels during economic recessions (Hamilton & Lin, 1996). Such a pattern is similar to the countercyclical nature of equity risk premium predictability, which suggests that it could be an important factor driving the phenomenon. Standard models do not capture either the clustering or heavy-tailed nature of returns. This means that one needs to opt for methods such as GARCH or stochastic volatility (SV) to account for these properties.

This thesis incorporates both explanations of the countercyclical predictability of the equity risk premium. Section 3 specifies models that incorporate regime-dependence and models that

[1] This thesis refers to the phenomenon of countercyclical relative forecasting accuracy of models that make use of predictors compared to uninformative models that base the forecasts on the historical average as simply *countercyclical predictability* or *countercyclical pattern*.

[2] Henkel et al. (2011), Bai (2010), Huang et al. (2015) and Neely et al. (2014) all report the countercyclical pattern.

incorporate time-varying volatility and volatility clustering. In order to fully take into account the uncertainty of these models, a Bayesian estimation approach is appropriate. This approach also facilitates the estimation of stochastic volatility models, which are difficult to track in a frequentist setting (Jacquier et al., 1994). The thesis evaluates the models using a pseudo-out-of-sample forecasting study. The main goal is to compare the different models and establish which specification is able to yield the highest predictive accuracy. The accuracy of the different models also provides an initial interpretation of what the main source of the countercyclical predictability of the equity risk premium is.

The approach of the research starts by performing pseudo-out-of-sample studies of different linear models that predict the risk premium. The models differ according to the categories of variables used as predictors, and the amount of desired explanatory variables. The studies use an expanding window procedure. At each point in time of the procedure, the PLS method recreates factors of the desired categories, and a Bayesian MCMC sampler reestimates the distributions of model parameters, including the forecast step. Summaries of the predictive performances across the entire sample allow for the evaluation of each model's predictive accuracy. In an attempt to improve the accuracy of well-performing variations further, the procedure is repeated by replacing the linear specification with non-linear specifications that either take into account regime-dependency or time-varying volatility. Markov switching, GARCH, and stochastic volatility models allow for the existence of these phenomena in the forecasting procedure. Finally, the research includes a market timing study that translates the 1-month ahead forecasts into monthly updated trading strategies.

This thesis contributes to the literature in several ways. First of all, earlier papers show that the use of factors generated from specific categories of variables has value when forecasting the equity risk premium. This thesis assesses the predictive value when putting multiple of these categories together. It also builds on previous findings by assessing the accuracy across multiple forecast horizons and by creating the factors using a more advanced dimension reduction technique called PLS. Secondly, this thesis specifically tackles the common finding of countercyclical relative predictability. Models that incorporate time-varying volatility or regime-dependent predictive relations assume two major explanations of the phenomenon. When incorporating predictive factors into Markov switching and stochastic volatility models, the aim is to successfully account for the phenomenon and thereby yield better and more consistent forecast performance across the business cycle. Finally, by using Bayesian estimation methods to make forecasts, the research also takes into account the full parameter uncertainty and the accuracy of the entire predictive density.

The main findings are the following. First, it is clear that factors generated from technical indicators lead to the best forecast performance at short time horizons. Sentiment factors provide some complementary value, as they yield positive predictive accuracy at the forecast horizon of 12 months. There is no evidence found of positive forecasting performance generated by macroeconomic factors. The predictability measures at the 1-month forecast horizon all translate into similar economic value when implemented as a trading strategy. Moreover, alternative specifications of the risk premium's volatility process also lead to additional predictability. The strong performance at the 1-month forecast horizon mostly occurs during the post-2008 period. At short forecast horizons, stochastic volatility models are the best option. However, at longer forecast horizons, Markov switching models that assume regime-dependent volatility and a regime-dependent intercept generate the best results. Both methods lead to more consistent results across the business cycle. Hence, this suggests that the time-varying nature of volatility is at least partly responsible for the presence of countercyclical predictability of the equity premium. The results do not support the hypothesis that regime-dependent relations between the predictors and the risk premium are a main driver of countercyclical predictability. Finally, when implemented as a trading strategy, the improvement of predictability caused by regime-dependent and time-varying volatility models do not translate into additional economic value.

2 Literature Review

This section clarifies the results and methods used by the literature in more detail. Furthermore, it elaborates on the countercyclical equity premium predictability phenomenon, and how earlier literature incorporates these aspects.

2.1 Equity Risk Premium Predictors

A comprehensive research paper in the area of equity risk premium forecasting is that of Welch and Goyal (2007). In this field of wide-ranging results, they reexamine some of the most prominent candidate variables for equity premium forecasting. They find that all the evaluated variables yield poor forecasting performance, often due to unstable and spurious relationships. They do leave open the possibility that the inclusion of more variables or non-linear models could still be promising ventures of future research.

Indeed, it is what much subsequent research reports. A prominent example is Bai (2010).

Inspired by the methods used by Stock and Watson (2002), she employs a set of over 120 macroeconomic predictors to forecast the equity premium. To solve the dimensionality problem, she uses PCA to summarize the information into factors, and selection techniques to select the factors most effective in forecasting. This so-called dynamic factor model (DFM) approach improves the accuracy of the forecasts. Using the same DFM method and variables, but with pre-selection techniques, Çakmaklı and van Dijk (2016) find similar results. Neely et al. (2014) also adopt the DFM approach and achieve improved forecasting accuracy, using a range of technical indicators as predictors. The same result is found by Baker and Wurgler (2007), who construct an investor sentiment index by applying PCA to a set of sentiment proxies.

It is not clear whether PCA is actually the best dimension reduction technique to use when optimizing forecasting performance. From an econometric point of view, PCA converts the information of the variables into a set of linearly uncorrelated factors. The method subsequently orders the factors according to their eigenvalue size. As a consequence, it sorts the factors according to the amount of variability within the original set of variables that they explain. It does not necessarily mean that the information contained by the factors is also most effective at forecasting a variable not included in the original data set, as shown by Bai and Ng (2008). This is a problem that Kelly and Pruitt (2013) tackle by using the partial least squares method to construct the factors. As opposed to PCA, PLS constructs factors by maximizing the covariance with the forecast target. Huang et al. (2015) apply the method to the sentiment index study of Baker and Wurgler (2007), and report improved forecasting results. This makes economic sense, as PCA is liable to include noise from the full set of predictors that does not have predictive power. In contrast, PLS attempts to maximize the information that is relevant for forecasting, making the factors a closer proxy to the underlying drivers of equity returns. Hence, this thesis uses the PLS method for the construction of factors across different categories.

2.2 Countercyclical Equity Premium Predictability

A common pattern in the literature of forecasting the equity premium (dating back to early findings of Fama and French (1989)) is that most studies report a higher degree of relative predictability during recessions compared to during expansions. This result is present regardless of the predictor set, although different predictors do capture different stages of the recession well. Bai (2010) highlights two explanations. The first is that predictability is most pronounced during economic regime switches, leading to uncertain, unsettled investors, and hence an opportunity for a well-specified model to capture economically significant returns. This is also the explanation put forward by Pesaran and Timmermann (1995). A similar explanation is that investors have a difficult time identifying turning points of business cycles, leading to well-specified models being able to outperform them. A more straightforward explanation is put forward by Henkel et al. (2011). During times of worsening economic conditions, investors demand a higher risk premium, and hence volatility of both the predictors and returns increase as well. Thus, the adjustments to the equity premium per unit of change in its predictors are larger and more pronounced during times of bad economic conditions.

2.2.1 Regime-Dependency

One way to overcome the problem of countercyclical predictability is to directly assume regime-dependency in the forecasting model. Sander (2018) shows several ways to allow for different parameter coefficients during recessions and expansions. In one approach, he uses a simple recession signal based on external data and uses different estimates in the case of a recession or an expansion. In another approach known as Markov switching, he allows the data to infer unobserved states. MS allows the model parameters to differ between these different states. Although Sander (2018) finds stronger results for the former approach, this thesis sticks to the latter.[3]

The Markov switching model is more common in the literature, and implemented by Henkel et al. (2011), Guidolin and Timmermann (2007) and Guidolin and Hyde (2012). All three report improved forecasting performance using the MS specification compared to constant parameter specifications. The approach of Henkel et al. (2011) is closest to this thesis. They assume two states, use several predictors and implement a Bayesian estimation approach. They specifically find improved forecasting performance during expansions, which is where constant parameter models tend to struggle. A major point where this thesis differs from Henkel et al. (2011) is that they use iterated forecasting in a VAR setting, whereas this thesis uses direct forecasting in a univariate setting. Refer to section 3 for the motivation. Another difference is their use of a much more limited set of individual predictive variables

[3] The reason is that Sander (2018) only evaluates one step ahead forecasts. It bypasses the problem that the recession signal approach does not include a method of forecasting a recession, which the MS model does have.

2.2.2 Time-Varying Volatility

Motivated by the arguments of Henkel et al. (2011) on the volatile nature of both returns and predictors during times of recession, one can also attribute the countercyclical pattern directly to the volatility process of returns. If the volatility process is indeed the main source of the countercyclical pattern, a model accounting for a time-varying volatility process should yield gains in forecasting performance over both a constant-parameter model and a regime-dependent model. Earlier literature in the field of equity premium forecasting incorporates this by including measures of volatility as predictors into a VAR framework (Campbell, Giglio, Polk, & Turley, 2018). However, this thesis takes an approach more similar to Geweke and Amisano (2010) and Nakajima (2012), who study the effect of different specifications of the shocks on equity returns. Both find stronger performance in forecasting the predictive density by stochastic volatility models over models such as GARCH. The points of departure from these papers are that this thesis uses data of monthly frequency as opposed to daily, and implements predictors into the forecast equation. As Cenesizoglu and Timmermann (2012) argue that economic state variables and the equity return distribution are related, the combination of predictors and time-varying volatility used by this thesis could yield additional benefits

The combination of predictive models and stochastic volatility is common in the macroeconomic literature but remains a relatively underexplored area in the field of forecasting equity returns. In the macroeconomic literature, examples of SV implementations are Clark and Ravazzolo (2015) and Pettenuzzo and Timmermann (2017). Similar to the predictive framework of this thesis, they predict macroeconomic variables of monthly frequency using a set of current predictors. After incorporating stochastic volatility into the model, they report improved forecasting accuracy for both point forecasts and density forecasts. Pettenuzzo et al. (2014) show that the SV model is also appropriate for a predictive equity premium model with monthly frequency. Although their findings are relatively unrelated to this thesis, their model functions as a starting point for the specification shown in section 3.2.

3 Econometric Methods and Techniques

The predictive regression in equation 1 forms the basis of the other, more complex, forecasting models. It takes the following form:

Constant parameter linear prediction model (Linear model):

$$r_{t+h}^{(h)} = \alpha^{(h)} + \boldsymbol{\beta}'^{(h)}\boldsymbol{x}_t + \sigma^{(h)} z_{t+h}^{(h)}, \quad z_t^{(h)} \sim \mathcal{N}(0,1), \tag{1}$$

where $r_{t+h}^{(h)}$ is the h-step ahead forecast of excess equity returns, $\alpha^{(h)}$ is the intercept parameter, $\boldsymbol{\beta}^{(h)}$ is a $(k \times 1)$ vector of parameters, $\boldsymbol{x}_t$ is a $(k \times 1)$ vector of predictors, $\sigma^{(h)}$ is the scale parameter of the error process, and $z_{t+h}^{(h)}$ the error of the standard normal distribution.

Note that the superscripts (h) reflect that the parameter estimates are dependent on the forecast horizon. This direct forecasting method means that each excess return forecast is dependent on only the known values of its predictors. The consequence is that each forecast horizon uses different parameter estimates. An alternative approach is iterated forecasting. In that case, the excess returns depend on the $(t-1)$ lags of its predictors, which requires an $(h-1)$-step ahead forecast of each predictor for each forecast of excess returns. The method requires a vector time series model containing all predictors and the dependent variable. Rolling the model forward generates multistep ahead forecasts. The latter approach has some advantages. For instance, by using a vector autoregressive model (VAR), the model takes the covariation of the error processes into account. Furthermore, any h-step ahead forecast of the vector of variables depends on an $(h-1)$ forecast or observation of the vector. Making use of more recent information in this way is likely a realistic assumption and a closer approximation to the correctly specified model. However, the most significant drawback is the large number of parameters that such a VAR system requires. This could erode forecasting performance as the number of parameters increases, and also adds substantial computation time.

According to Stock and Watson (2002), direct forecasting is an alternative approach that reduces the number of parameters in an effective manner.[4] They recognize that a multistep forecast is linear in its predictors. Hence, an h-step ahead projection is appropriate to construct the forecasts directly. As a consequence, forecasts of the predictors are not needed, and hence a VAR system is not necessary. This reduces the number of parameters to a significant degree. Marcellino et al. (2006) tackle the trade-off between iterated and direct forecasting more specifically. They note that iterated forecasting produces more efficient parameters than the direct approach if the model is correctly specified. However, it also suffers from more bias when it is not the case. They point out that the theoretical literature tends to see direct forecasting as a method that produces more robust results (Ing, 2003; Bhansali, 1999). Bhansali (1999) also makes the specific recommendation

[4] They use direct forecasting in the context of macroeconomic forecasting.

that direct forecasting is recommended in the case when the time series process is particularly complex and the amount of true parameters is very large. Similar to Stock and Watson (2002), this thesis uses a relatively large set of predictors and also uses more complex model specifications to produce forecasts. Hence, the number of parameters and the robustness of the results are the most important elements driving the choice between direct and iterated forecasting. This thesis opts for direct forecasting as it provides clear benefits in these areas.

3.1 Constructing Predictors

The first key question is how to choose the content of the $\boldsymbol{x}_t$ vector of predictors. As explained in section 1 and 2, multiple categories of predictors are available, each containing a vast amount of variables. Including all variables in the predictive equation increases estimation uncertainty which adversely affects the forecasting performance. The solution is to summarize the information of each category into smaller sets of factors.[5] Hence, this thesis makes assumptions on the structure of the available data and extracts factors based on these assumptions. The interpretation that follows is a more general case of the interpretation given by Huang et al. (2015) that allows for multiple categories and drivers.

For the constant, linear case, the assumption is that the h-step ahead expected equity premium is explained by the following underlying relation:

$$\begin{aligned} r_{t+h} &= E_t[r_{t+h}] + \epsilon_{t+h} \\ &= \alpha_u^{(h)} + \sum_{i=1}^{K_A} \beta_{i,t}^{(A)} f_{i,t}^{(A)} + \sum_{i=1}^{K_B} \beta_{i,t}^{(B)} f_{i,t}^{(B)} + \sum_{i=1}^{K_C} \beta_{i,t}^{(C)} f_{i,t}^{(C)} + \epsilon_{t+h}, \end{aligned} \tag{2}$$

where $\alpha_u^{(h)}$ is the long-term unconditional expectation, $f_{i,t}^{(c)}$ denotes the ith factor of the indicated category c at time t, $\beta_{i,t}^{(C)}$ is the corresponding regressor and K_c denotes the amount of factors of category c. Note again that as the forecasting procedure optimizes for different forecast horizons, it is the case for the factors as well. The illustration assumes three categories. All factors are unobserved and need to be inferred from a set of proxies. Let $\boldsymbol{p}_t^{(c)} = (\tilde{p}_{1,t}^{(c)}, ..., \tilde{p}_{N_c,t}^{(c)})'$ represent a vector of observed proxies for an arbitrary category at time t and assume that each individual proxy $\tilde{p}_{i,t}^{(c)}$ follows a factor structure in the following fashion:

[5] In line with the methods used by Stock and Watson (2002), Çakmaklı and van Dijk (2016), Bai (2010), Baker and Wurgler (2007), and others.

$$\tilde{p}_{i,t}^{(c)} = \delta_{i,0}^{(c)} + \sum_{j=1}^{K_c} \delta_{i,j}^{(c)} f_{j,t}^{(c)} + \delta_{i,w}^{(c)} w_t^{(c)} + e_{i,t}^{(c)}, \quad i = 1, ..., N_c, \tag{3}$$

where each δ denotes a parameter of the regression equation, $f_{j,t}^{(c)}$ denotes an unobserved factor that drives the equity premium, $w_t^{(c)}$ denotes the error component common to all proxies of the given category that is irrelevant towards forecasting returns, and $e_{i,t}^{(c)}$ denotes the idiosyncratic error related to the particular proxy variable. The goal is to infer all factors $f_j^{(c)}$ that drive the equity return from the set of available proxies. As the common PCA method is unable to distinguish between between the factors $f_{j,t}^{(c)}$ and the common error component $w_t^{(c)}$, the PLS method explained by Kelly and Pruitt (2013) is necessary.

Generally, PLS infers the factors $f_{j,t}^{(c)}$ from the cross-section of proxies based on their covariance with the h-step ahead equity returns. As the existence of multiple drivers is possibly appropriate for certain categories, the general case of PLS estimation is necessary, as opposed to the 1-factor case of Kelly and Pruitt (2013). In matrix notation, the proxies can be denoted as follows:

$$\boldsymbol{P}^{(c)} = \boldsymbol{F}^{(c)} \boldsymbol{Q}'^{(c)} + \boldsymbol{U}^{(c)}, \tag{4}$$

where $\boldsymbol{P}^{(c)}$ is the $T \times N_c$ matrix of proxies for category c, $\boldsymbol{F}^{(c)}$ is the $T \times K_c$ matrix of factors for category c, $\boldsymbol{Q}^{(c)}$ is a $N_c \times K_c$ orthogonal matrix, and $\boldsymbol{U}^{(c)}$ is an error matrix. The factor matrix $\boldsymbol{F}^{(c)}$ is constructed by maximizing the covariance between $\boldsymbol{F}^{(c)}$ and the h-step ahead equity return vector $\boldsymbol{r}^{(h)}$. The most common method for implementation is the nonlinear iterative partial least squares (NIPALS) algorithm.[6]

3.2 Advanced Models

To incorporate the phenomena of regime-dependence and time-varying volatility, this thesis expands the linear model of equation 1 in several ways.

3.2.1 Markov Switching Models

A Markov switching as stated here is able to take regime-dependency into account. The first version reads according to the following specification:
where the subscript $\zeta_{t+h}^{(h)}$ indicates that the parameter is dependent on the state $\zeta_{t+h}^{(h)}$ at time $t+h$,

[6] Refer to Ng (2013) for more detailed information on the method.

Basic Markov switching model (Basic MS model):

$$
\begin{aligned}
r_{t+h}^{(h)} &= \alpha_{\zeta_{t+h}^{(h)}}^{(h)} + \boldsymbol{\beta}'^{(h)}\boldsymbol{x}_t + \sigma_{\zeta_{t+h}^{(h)}}^{(h)} z_{t+h}^{(h)}, \qquad z_t^{(h)} \sim \mathcal{N}(0,1) \\
p_{ij}^{(h)} &= \Pr[\zeta_t^{(h)} = i | \zeta_{t-1}^{(h)} = j] \qquad \zeta_t^{(h)} \in \{1,..,M\},
\end{aligned}
\tag{5}
$$

and $p_{ij}^{(h)}$ denotes the transition probability from state j to state i. The amount of possible states M is restricted to 2, under the assumption that the states will roughly correspond to periods of high and low volatility, which would capture the countercyclical nature of the estimates. However, the basic MS model of equation 5 does not yet allow for regime-dependency in the predictor parameters. It only allows for regime-dependency in the intercept and volatility. Hence, it is an intermediate step towards the more general model, defined as follows:

Full Markov switching model (Full MS model):

$$
\begin{aligned}
r_{t+h}^{(h)} &= \alpha_{\zeta_{t+h}^{(h)}}^{(h)} + \boldsymbol{\beta}'^{(h)}_{\zeta_{t+h}^{(h)}}\boldsymbol{x}_t + \sigma_{\zeta_{t+h}^{(h)}}^{(h)} z_{t+h}^{(h)}, \qquad z_t^{(h)} \sim \mathcal{N}(0,1) \\
p_{ij}^{(h)} &= \Pr[\zeta_t^{(h)} = i | \zeta_{t-1}^{(h)} = j] \qquad \zeta_t^{(h)} \in \{1,..,M\}
\end{aligned}
\tag{6}
$$

A common tool for estimating Markov switching models is a Bayesian Markov Chain Monte Carlo (MCMC) algorithm. The Bayesian approach can effectively estimate the latent states and regression coefficients simultaneously and thus allow for incorporating the estimation uncertainty. As an alternative method like the Hamilton filter always conditions on the parameter estimates when estimating the latent states, it cannot incorporate uncertainty in the same way.

3.2.2 Time-Varying Volatility Models

The regime-dependent models already allow for the presence of time-varying volatility. However, as the volatility process of the error term is dependent on the binary state, it only allows for two different magnitudes of volatility. This section considers more general specifications of the volatility process, allowing for the magnitude of volatility to take different continuous values over time.

First, consider the common GARCH(1,1) model. It is one of the most well-known and most used tools of estimating time-varying volatility. As Nakajima (2012) shows that more advanced specifications of GARCH at best yield only small benefits over the basic GARCH(1,1) model in the area of equity return forecasting, this thesis sticks to the basic specification. It is defined as follows:

where ψ_t is the information set up to and including time t, ${\sigma_t^{(h)}}^2$ is the conditional variance at time

GARCH(1,1) model (GARCH model):

$$
\begin{aligned}
r_{t+h}^{(h)} &= \alpha^{(h)} + \boldsymbol{\beta}'^{(h)}\boldsymbol{x}_t + \epsilon_{t+h}^{(h)}, \qquad & \epsilon_t^{(h)}|\psi_{t-1} \sim \mathcal{N}(0, \sigma_{t+h}^{(h)^2}) \\
\sigma_{t+h}^{(h)^2} &= a_0^{(h)} + a_1^{(h)}\epsilon_{t+h-1}^{(h)^2} + b_1^{(h)}\sigma_{t+h-1}^{(h)^2}, &
\end{aligned} \tag{7}
$$

t, $a_0^{(h)} > 0$, $a_1^{(h)} \geq 0$ and $b_1^{(h)}$ are the parameters of the volatility process, with $a_1^{(h)} + b_1^{(h)} < 1$. Note that the future volatility is not forecast directly, but in an iterated fashion. This is possible as the volatility forecast is not dependent on the predictors, and hence does not require a multivariate system for its forecasts. Another motivation for the use of the iterated approach for forecasting volatility is the fact that it yields superior forecasting results (Ghysels, Valkanov, & Serrano, 2009).

Next, consider the following volatility specification:

Stationary stochastic volatility model (Stationary SV model):

$$
\begin{aligned}
r_{t+h}^{(h)} &= \alpha^{(h)} + \boldsymbol{\beta}'^{(h)}\mathbf{x}_t + s_{t+h}^{(h)} z_{t+h}^{(h)}, \qquad & z_t^{(h)} \sim \mathcal{N}(0,1) \\
h_{t+h}^{(h)} &= \log(s_{t+h}^{(h)^2}) & \\
h_{t+h}^{(h)} &= \gamma_0^{(h)} + \gamma_1^{(h)}(h_{t+h-1}^{(h)} - \gamma_0^{(h)}) + \sigma_\eta^{(h)}\eta_{t+h}^{(h)}, & \eta_t^{(h)} \sim \mathcal{N}(0,1),
\end{aligned} \tag{8}
$$

where $s_{t+h}^{(h)} z_t^{(h)}$ is the conditional volatility at time t, $\gamma_0^{(h)}$, $|\gamma_1^{(h)}| < 1$ and $\sigma_\eta^{(h)}$ are the parameters of the volatility process, and $\eta_t^{(h)}$ is the shock of the volatility process at time t. Stationarity of the volatility process is suitable when volatility is not expected to potentially blow up over the forecast horizon (see Clark and Ravazzolo (2015)). However, some research shows that assuming a random walk (RW) specification of the volatility process can improve forecast results (i.e. Nakajima (2012)). By imposing the restrictions $\gamma_0^{(h)} = 0$ and $\gamma_1^{(h)} = 1$ to equation 8, the stationary SV model turns into a random walk SV model. As Jacquier et al. (1994) show, satisfactory estimation of the stochastic volatility model requires the use of a Bayesian MCMC algorithm. Hence, in order to be consistent, the GARCH model is also estimated via MCMC.

This thesis opts for the use of a Hamiltonian Monte Carlo (HMC) algorithm called the No-U-Turn Sampler (NUTS) as the MCMC sampler of choice. The motivation is that NUTS leads to fast convergence of the sampler and produces a smaller degree of autocorrelated draws compared to other MCMC samplers (Hoffman & Gelman, 2014). As stochastic volatility models especially tend to suffer from a large degree of autocorrelation in its draws, the choice is particularly appropriate for the purposes of this research. See appendix A for more information on the MCMC algorithm.

3.3 Priors

Following the practice of Pettenuzzo et al. (2014) and Pettenuzzo and Timmermann (2017), this thesis calibrates several of the prior's hyperparameters using a small training sample of size $T_t = 60$ (5 years) at the start of the full data set. It ensures the sampler uses economically realistic distributions as priors, which improves the stability of the MCMC algorithm.

As is standard practice, the models assume that the parameters α and $\boldsymbol{\beta}$ follow a normal distribution independent from the error process of the returns.

$$\begin{bmatrix} \alpha \\ \boldsymbol{\beta} \end{bmatrix} \sim \mathcal{N}(\underline{\boldsymbol{b}}, \underline{\boldsymbol{V}}), \text{ where}$$
$$\underline{\boldsymbol{b}} = \begin{bmatrix} \bar{r}_{T_t} \\ 0 \end{bmatrix}, \underline{\boldsymbol{V}} = \begin{bmatrix} \underline{\psi}^2 s^2_{r,T_t} & 0 \\ 0 & \underline{\psi} s^2_{r,T_t} \boldsymbol{S}^{-1}_{x,T_t} \end{bmatrix}. \tag{9}$$

In this case $\bar{r}_{T_t}$ denotes the sample average of the equity premium across the training sample, s^2_{r,T_t} denotes the sample variance, and $\boldsymbol{S}_{x,T_t}$ denotes the sample variance matrix of the matrix of predictors, with zeroes on the off-diagonals. $\underline{\psi}$ controls the tightness of the prior with $\underline{\psi} \to \infty$ corresponding to a diffuse prior. Following Pettenuzzo et al. (2014) the model sets $\underline{\psi} = 2.5$. Note that these hyperparameters are different for each forecast horizon h as the predictor set differs. Note also that the priors apply to the linear model of equation 1, but also for any variant of α and $\boldsymbol{\beta}$ in any of the more advanced models.

Next, the error precision of the innovation process of returns σ^{-2} follows the following gamma distribution:[7]

$$\sigma^{-2} \sim \text{Gam}(\underline{v}_0 \, T_t \,/\, 2,\ s^{-2}_{r,T_t} \,/\, (2 \, \underline{v}_0 \, T_t)), \tag{10}$$

where $\underline{v}_0$ determines the degree of informativeness of the prior with $\underline{v}_0 \to 0$ describing a diffuse prior. Following Pettenuzzo et al. (2014), $\underline{v}_0$ is set to 0.1.

The Markov switching models of equation 5 and 6 require additional priors on the transition probabilities p_{ij}. As the amount of states is restricted to 2, imposing a univariate beta-distributed prior on the transition probabilities p_{11} and p_{22} is sufficient. Adapted from Pettenuzzo and Tim-

[7]For clarity, this thesis uses the gamma distribution corresponding to the probability density function $f(x;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$, as opposed to the one used by Pettenuzzo et al. (2014).

mermann (2017), they are as follows:

$$p_{ii} \sim \text{Beta}(\text{e}^2,\ \text{e}^{1/2}),\ \text{for } i \in \{1,2\}. \tag{11}$$

The GARCH model in equation 7 requires additional prior specifications for parameters a_0, a_1 and b_1. As proposed by Clark and Ravazzolo (2015) they are set as uninformative uniform distributions. The hyperparameters correspond to the appropriate boundaries imposed by the GARCH model, as follows:

$$\begin{aligned} a_0 &\sim U(0, 10^4), \\ a_1 &\sim U(0,1), \\ b_1 &\sim U(0, 1-a_1). \end{aligned} \tag{12}$$

The models using a stochastic volatility process (equation 8) require priors on the initial volatility h_1, and the parameters γ_0, γ_1, and σ_η. Again, the priors follow those of either Pettenuzzo et al. (2014) or Pettenuzzo and Timmermann (2017) for the random walk specification or the stationary specification, respectively:

$$\begin{aligned} h_1 &\sim \mathcal{N}(\ln(s_{r,T_t}), \underline{k}_h), \\ \begin{bmatrix} \gamma_0 \\ \gamma_1 \end{bmatrix} &\sim \mathcal{N}\left(\begin{bmatrix} \underline{m}_{\gamma_0} \\ \underline{m}_{\gamma_1} \end{bmatrix}, \begin{bmatrix} \underline{V}_{\gamma_0} & 0 \\ 0 & \underline{V}_{\gamma_1} \end{bmatrix} \right), \\ \sigma_\eta^{-2} &\sim \text{Gam}(1\ /\ 2,\ 1\ /\ (2\ \underline{k}_\eta)), \end{aligned} \tag{13}$$

with $\underline{k}_h = 10$, $\underline{m}_{\gamma_0} = 0$, $\underline{m}_{\gamma_1} = 0.9$, $\underline{V}_{\gamma_0} = 0.25$, $\underline{V}_{\gamma_1} = 10^{-4}$ and $\underline{k}_\eta = 0.01$.

3.4 Forecasting Study

Bayesian analysis makes use of Bayes' rule in order to produce sequences of draws from the joint posterior density distribution. The posterior density is proportional to the product of the prior density and the likelihood function, as follows:[8]

$$p(\boldsymbol{\theta}_m \mid \boldsymbol{D}_{m,\tau}, m) \propto p(\boldsymbol{\theta}_m \mid m)\ p(\boldsymbol{D}_{m,\tau} \mid \boldsymbol{\theta}_m, m), \tag{14}$$

where $\boldsymbol{\theta}_m$ denotes the vector of both latent states and parameters associated with model m, and $\boldsymbol{D}_{m,\tau}$ denotes the matrix with observations of the equity premium and its predictors included in

[8]This section is largely inspired by the explanation of Geweke and Amisano (2010).

the respective sample window ending at time τ. For details on the prior distributions and the parameters, refer to section 3.3. The MCMC sampler expresses the joint posterior density by simulating a sequence of draws of $\boldsymbol{\theta}_m^{(s)}$, with $s = 1...S$. Inference can be made on the sequence of draws.

In the forecasting study, the value of interest is not necessarily the density of the joint posterior density, but rather the predictive density of the equity risk premium $r_{\tau+h}$. A general representation reads as follows:

$$p(r_{\tau+h} \mid \boldsymbol{D}_{m,\tau}, m) = \int_{\boldsymbol{\theta}_m} p(r_{\tau+h} \mid \boldsymbol{D}_{m,\tau}, \boldsymbol{\theta}_m, m)\, p(\boldsymbol{\theta}_m \mid \boldsymbol{D}_{m,\tau})\, d\boldsymbol{\theta}_m. \tag{15}$$

Inference on the predictive density can be made after again simulating a sequence of draws $r_{\tau+h}^{(s)}$. The sampler achieves this by simulating a single value of $r_{\tau+h}^{(s)}$ from each density represented by $p(r_{\tau+h} \mid \boldsymbol{D}_{m,\tau}, \boldsymbol{\theta}_m^{(s)}, m)$. The vector $\boldsymbol{\theta}_m^{(s)}$ is a single draw produced by sampling from the joint posterior density. This thesis applies several evaluation criteria to make inference on the sequences of draws from the predictive density. These criteria assess the accuracy of the forecasts and thus the performance of the different forecasting models.

In order to measure the accuracy of the different forecasting models in a systematized way, this thesis performs a pseudo-out-of-sample forecasting study. It means that in each iteration, the study estimates the model over a sample called the expanding window. The initial expanding window lasts from time period 1 to W. After the estimation, the sampler generates forecasts for the forecast horizon (h) in question, where $h \in \{1, 3, 6, 12\}$, meaning for time $W + h$. Then, the study expands by 1 period. The window is now 1 to $W + 1$ and the sampler generates forecasts for the period $W + h + 1$. The study repeats the process until no more observations are available.

The initial window size W is fixed at 15 years, corresponding to 180 monthly observations.[9] Moreover, the full procedure holds out 5 years of observations, corresponding to 60 monthly observations. This period functions as a training sample that calibrates the prior parameters of the Bayesian estimation procedure. The sample is not used within the expanding window procedure itself, which means it takes place before $t = 1$.

The evaluation criteria of the forecasting performance of the different models (m) are the following, starting with the out-of-sample (OOS) R^2 measure of Campbell and Thompson (2007):

[9] In accordance with the sample size used by Henkel et al. (2011).

$$R^2_{\text{OOS,m}} = 1 - \frac{\sum_{\tau=1}^{T}(r_\tau - \hat{r}_{m,\tau})^2}{\sum_{\tau=1}^{T}(r_\tau - \hat{r}_{\text{B},\tau})^2}, \tag{16}$$

where r_τ is the observed value of the equity risk premium at time τ and $\hat{r}_{m,\tau}$ is the posterior mean of the predictive density of model m. The denominator includes the posterior mean of the predictive density of benchmark model B. The measure assesses the accuracy of the posterior mean of the predictive density across all points in time for which forecasts are made. As can be inferred from the equation, a positive R^2 value indicates that a model outperforms the benchmark. As the value approaches 1, the model approaches perfect forecasts. A negative value shows that the model underperforms the benchmark, with lower values showing unfavourable performance compared to the benchmark.

To get a sense of how the forecast accuracy develops over time, the cumulative sum of squared prediction error difference (CSSED) proposed by Welch and Goyal (2007) is appropriate.

$$\text{CSSED}_{m,t} = \sum_{\tau=1}^{t} \left((r_\tau - \hat{r}_{\text{B},\tau})^2 - (r_\tau - \hat{r}_{m,\tau})^2\right). \tag{17}$$

This measure yields a value for every point in time for which forecasts are made. When shown on graph, a rising value indicates that the relative forecasting performance of the model m is increasing during the indicated period, and vice versa.

Moreover, the log predictive score (LS) is the common method to assess the accuracy of the posterior predictive density of the forecasts. Geweke and Amisano (2010) motivate the use of the measure. This thesis uses the quadratic approximation of Adolfson et al. (2007):

$$\begin{aligned} \text{LS}_{m,\tau}(r_\tau) &= -0.5 \ \left(\log(2\pi) + \log|\hat{V}_{m,\tau}| + (r_\tau - \hat{r}_{m,\tau})^2 \ \hat{V}_{m,\tau}^{-1}\right) \\ \text{LSD}_m &= \sum_{\tau=1}^{T}(\text{LS}_{m,\tau} - \text{LS}_{\text{B},\tau}), \end{aligned} \tag{18}$$

where $\hat{V}_{m,\tau}$ denotes the sample variance of the posterior forecast density at time τ for model m. The log-score differential (LSD) measure provides a value indicating the aggregate relative performance of model m compared to the benchmark model in the area of density forecasts.

Finally, another suitable measure to assess the accuracy of the forecast density makes use of the predictive likelihood. Geweke and Amisano (2010) provide a suitable method to simulate the predictive likelihood of each forecast, and compare the number to that of a benchmark, yielding a log Bayes factor (log BF). The predictive likelihood of a model m is denoted as follows:

$$PL_m(\tau) = p(r_\tau \mid \boldsymbol{x}_{\tau-h}, m) = \int_{\boldsymbol{\theta}_m} p(r_\tau \mid \boldsymbol{x}_{\tau-h}, m, \boldsymbol{\theta}_m)\, p(\boldsymbol{\theta}_m \mid \boldsymbol{x}_{\tau-h}, m)\, d\boldsymbol{\theta}_m,$$
$$\text{approximated by: } S^{-1}\sum_{s=1}^{S} p(r_\tau \mid \boldsymbol{x}_{\tau-h}, m, \boldsymbol{\theta}^{(s)}_{m,\tau-h}), \tag{19}$$

where S denotes the amount of draws in an iteration of the sampler, and $\boldsymbol{\theta}_m$ denotes the set of parameters estimated by the model. The approximation makes use of the posterior simulator of $\boldsymbol{\theta}^{(s)}_{m,\tau}$ produced by the MCMC algorithm. As Geweke and Amisano (2010) note, summing over the logarithmic values of the predictive likelihoods for all τ yields the logarithmic value of the full marginal likelihood (meaning over all the sample windows). The value can then be used to compare two competing models by means of the log Bayes factor, decomposed as follows:

$$\log\, BF_m = \log\left[\frac{p(\boldsymbol{r}_T \mid m)}{p(\boldsymbol{r}_T \mid B)}\right] = \sum_{\tau=1}^{T} \log\left[\frac{PL_m(\tau)}{PL_B(\tau)}\right], \tag{20}$$

where $PL_m(\tau)/PL_B(\tau)$ is the predictive Bayes factor in favor of model m over benchmark model B for observation τ. The predictive Bayes factors are approximated by using equation 19.

The log Bayes factor is also appropriate for measuring the evolution of the relative predictive performance over time. The only necessary change to equation 20 is the need to sum up until time t instead of time T, as follows:

$$\log\, BF_{m,t} = \log\left[\frac{p(\boldsymbol{r}_t \mid m)}{p(\boldsymbol{r}_t \mid B)}\right] = \sum_{\tau=1}^{t} \log\left[\frac{PL_m(\tau)}{PL_B(\tau)}\right], \tag{21}$$

Kass and Raftery (1995) propose rules of thumb for interpreting the evidence generated by the log Bayes factors. The evidence refers to the hypothesis that the model has more success at predicting the actual data than the benchmark. When the value is negative, one can also assess the opposite hypothesis by taking the absolute value of the log Bayes factor and using these same rules. They are as follows:

$$\begin{aligned}
0 \leq \log\, BF_{m,t} < \ 1&: \text{ Not worth more than a bare mention}\\
1 \leq \log\, BF_{m,t} < \ 3&: \text{ Positive evidence}\\
3 \leq \log\, BF_{m,t} < \ 5&: \text{ Strong evidence}\\
5 \leq \log\, BF_{m,t} \qquad &: \text{ Very strong evidence}
\end{aligned} \tag{22}$$

4 Data

This thesis uses data which correspond to four different categories. All data series consist of monthly observations. The full sample consists of 492 monthly observations, from January 1975 to December 2014. The first category is that of the equity risk premium itself. Its definition is the monthly market return minus the monthly risk-free return. For the market return, this research uses the value-weighted return of the S&P 500, including dividends.[10] The risk-free rate is defined as the 1-month T-bill rate.[11]

The next category consists of macroeconomic variables. This thesis deviates slightly from Çakmaklı and van Dijk (2016) by collapsing their definitions of financial variables and macroeconomic variables into a single macroeconomic category. The reason is that the definitions of several financial variables are very close to those of similar variables included in the set of macroeconomic variables. This leaves a small number of financial variables, namely the variables of the dividend yield, price-earnings ratio, the default spread and the additional cyclically-adjusted price-earnings ratio. The full sample analysis in appendix B points out that these variables are not clearly complementary to the other macroeconomic variables. Hence, they belong in the same category. The remainder of the full list of macroeconomic variables follows that of Çakmaklı and van Dijk (2016) as close as possible. Due to changes to the data sets since the publication of their paper, this thesis makes some adjustments and omissions. It leaves a total of 104 macroeconomic variables, for which appendix C provides a full list.

The category of sentiment variables follows the research of Baker and Wurgler (2007). The variables are five series of proxies of investor sentiment. These are the value-weighted dividend premium, the first-day returns on IPOs, IPO volume, the closed-end fund discount, and the equity share in new issues.[12] See Baker and Wurgler (2007) for more details. A sixth variable, turnover on the New York stock exchange, has been dropped since the original publication of the paper. The authors cite the explosion in high-frequency trading and the migration to different trading venues as reasons to believe that the definition of the variable has changed and is no longer a good proxy for investor sentiment. The value-weighted dividend premium has a total of 16 missing observations, which are filled using linear interpolation.

The construction of the technical variables follows Neely et al. (2014). They amount to 14

[10] The series is available on the CRSP database and reported in percentage points.
[11] The series is available on the website of Kenneth French.
[12] The data are available on the website of Jeffrey Wurgler.

indicators, based on three popular trading strategies. The strategies are momentum, moving-average (MA), and moving-average of volume (MA^{OBV}). The indicators always take on the value 1 or 0.[13]

Neely et al. (2014) construct the momentum indicator as follows:

$$S_{i,t}^{\text{MOM}(\tilde{m})} = \begin{cases} 1, & \text{if } P_t \geq P_{t-\tilde{m}} \\ 0, & \text{if } P_t < P_{t-\tilde{m}}, \end{cases} \tag{23}$$

where P_t is the index price at time t. The indicator signals positive momentum if the current price is greater or equal than the price $\tilde{m}$ periods ago. Momentum variables are produced for $\tilde{m} = \{9, 12\}$.

The MA strategy looks as follows:

$$\begin{aligned} S_{i,t}^{\text{MA}(\tilde{s},\tilde{l})} &= \begin{cases} 1, & \text{if } \text{MA}_{\tilde{s},t} \geq \text{MA}_{\tilde{l},t} \\ 0, & \text{if } \text{MA}_{\tilde{s},t} < \text{MA}_{\tilde{l},t} \end{cases}. \\ \text{MA}_{j,t} &= (1/j) \sum_{i=0}^{j-1} P_{t-i} \text{ for } j = \tilde{s}, \tilde{l}. \end{aligned} \tag{24}$$

The strategy provides a positive trading signal if the moving average with a short look-back period is greater or equal than the moving average with a long look-back period. This thesis uses the look-back periods of $\tilde{s} = \{1, 2, 3\}$ and $\tilde{l} = \{9, 12\}$, producing 6 MA variables.

Finally, the moving-average of volume signals are calculated as follows:

$$\begin{aligned} S_{i,t}^{\text{VOL}(\tilde{s},\tilde{l})} &= \begin{cases} 1, & \text{if } \text{MA}_{\tilde{s},t}^{\text{OBV}} \geq \text{MA}_{\tilde{l},t}^{\text{OBV}} \\ 0, & \text{if } \text{MA}_{\tilde{s},t}^{\text{OBV}} < \text{MA}_{\tilde{l},t}^{\text{OBV}} \end{cases}. \\ \text{MA}_{j,t}^{\text{OBV}} &= (1/j) \sum_{i=0}^{j-1} \text{OBV}_{t-i} \text{ for } j = \tilde{s}, \tilde{l}. \\ \text{OBV}_t &= \sum_{k=1}^{t} \text{VOL}_k D_k, \end{aligned} \tag{25}$$

where OBV denotes the on-balance volume. This is a measure that increases at time t by the amount of traded volume at time t (VOL_t) if the price at period t has increased, and decreases by the amount if the price has decreased. The increase or decrease in price is indicated by D_t, which

[13] The data used to construct these variables are the price of the S&P 500 composite index, available on the CRSP database, and its traded volume, available on Yahoo Finance.

takes on the value 1 at periods of a price increase, and -1 at periods of a price decrease. Hence, the signal describes a moving-average strategy that is positive when the short moving-average measure of OBV is greater or equal than the long measure, and vice versa. For more details, see Neely et al. (2014). Appendix B presents arguments for separating the technical and sentiment variables into two different categories.

5 Results

This section presents the main results of the research. As the thesis consists of two main research questions, it produces two sets of results. The first addresses the question what set of predictors is most effective at forecasting the equity risk premium. In other words: it assesses whether adding any category of predictors to the forecast equation increases the predictive accuracy. To measure this, this research compares the forecast accuracy of linear models of multiple possible combinations of the three predictor categories: macroeconomic variables, technical indicators and investor sentiment.[14] The benchmark for the evaluation criteria is an uninformative historical average (HA) model. The benchmark uses the average excess return over the sample window as its forecast for all forecast horizons. Section 5.1 presents the results.

The second set of results assesses the question whether forecast performance improves by implementing models that take regime-dependency or time-varying volatility into account. The results include the performance of the five forecasting models of equation 5 to 8. Important information for the performance of these models in relation to countercyclical predictability is how they perform specifically during recessions and expansion. Evaluation criteria that only sum over forecasts made during either recessions or expansions account for this.[15] Section 5.2 presents the results.

5.1 Results Linear Prediction Models

Table 1 shows the forecast performance for several different variations of the linear prediction model when evaluating the accuracy of the point forecasts. The first observation based on the results is the general poor performance of the macroeconomic factors. Regardless of the amount of factors used, models using only macroeconomic factors underperform the historical average benchmark. In fact, adding additional factors to the predictive equation tends to decrease the accuracy further.

[14] Appendix B presents a preliminary analysis that analyzes the content of the created factors, and argues for the separation of the categories into the three categories chosen here.

[15] The source of the recession indicator is the National Bureau of Economic Research (NBER).

Table 1: Out-of-sample R^2 values per linear prediction model

1995:12-2014:12	R^2 | Overall				R^2 | Expansion				R^2 | Recession			
Forecast horizon	1	3	6	12	1	3	6	12	1	3	6	12
Macroeconomic												
1F	-0.092	-0.069	-0.098	-0.048	-0.081	-0.072	-0.081	-0.073	-0.123	-0.059	-0.142	0.018
2F	-0.072	-0.067	-0.100	-0.092	-0.080	-0.124	-0.086	-0.124	-0.050	0.084	-0.135	-0.006
3F	-0.150	-0.096	-0.257	-0.170	-0.149	-0.162	-0.197	-0.236	-0.153	0.079	-0.416	0.004
Technical												
1F	0.114	-0.005	-0.018	-0.039	0.116	-0.007	-0.041	-0.055	0.109	0.001	0.046	0.003
2F	0.279	0.004	-0.074	-0.050	0.293	0.005	-0.094	-0.070	0.239	0.002	-0.021	0.005
3F	0.295	-0.008	-0.071	-0.089	0.316	-0.014	-0.091	-0.106	0.238	0.008	-0.017	-0.046
Sentiment												
1F	-0.034	-0.024	-0.008	0.045	-0.032	-0.010	0.001	0.042	-0.037	-0.060	-0.033	0.051
2F	-0.045	-0.042	0.000	0.055	-0.049	-0.031	0.000	0.044	-0.032	-0.070	-0.002	0.085
3F	-0.043	-0.047	-0.013	0.062	-0.046	-0.037	-0.015	0.051	-0.035	-0.075	-0.010	0.090
Combined												
1F	0.053	-0.075	-0.084	-0.062	0.066	-0.095	-0.087	-0.103	0.019	-0.022	-0.076	0.045
2F	0.075	-0.080	-0.074	-0.060	0.096	-0.115	-0.113	-0.086	0.018	0.014	0.029	0.008
3F	-0.019	-0.128	-0.223	-0.153	0.028	-0.169	-0.182	-0.213	-0.141	-0.019	-0.333	0.004
4F	0.041	-0.153	-0.206	-0.155	0.082	-0.225	-0.176	-0.223	-0.069	0.038	-0.288	0.024
Macro+Tech+Sent												
1F+1F+1F	0.062	-0.084	-0.099	-0.036	0.072	-0.073	-0.105	-0.068	0.034	-0.111	-0.085	0.047
2F+1F+1F	0.048	-0.089	-0.111	-0.061	0.054	-0.130	-0.108	-0.094	0.034	0.021	-0.117	0.026
1F+2F+1F	0.243	-0.075	-0.132	-0.027	0.265	-0.067	-0.131	-0.060	0.183	-0.095	-0.136	0.062
Tech+Sent												
1F+1F	0.106	-0.022	-0.018	0.005	0.098	-0.011	-0.033	-0.011	0.125	-0.050	0.023	0.045
2F+1F	0.271	-0.011	-0.076	-0.008	0.286	-0.002	-0.087	-0.026	0.231	-0.033	-0.047	0.041

Notes: The column on the left represent the predictors. *#F* indicates the number of factors the PLS procedure generates of the respective category. *Combined* means that PLS generates factors from the combined set of all predictive variables. *(Macro+)Tech+Sent* means that a certain number of factors are generated from each of the two or three categories separately. The accuracy is measured relative to the historical average benchmark model, over the out-of-sample period of 1995:12 to 2014:12. The *overall* column represents the entire period, whereas the *expansion* and *recession* columns represent the sub-samples of economic expansion or recession, respectively, as measured by NBER. See equation 16 in section 3.4 for more information on the evaluation criterion.

Both of these findings are not entirely in line with the findings of Çakmaklı and van Dijk (2016) and Bai (2010), who find positive performance of macroeconomic factor models compared to an uninformative benchmark. Although their use of a different evaluation sample compared to this thesis could play a role in this, it is likely that macroeconomic models require some pre-selection or post-selection techniques of the variables[16] to improve the accuracy. The results do not support the hypothesis that PLS is a good replacement for these techniques in the case of macroeconomic variables. However, there is evidence of the existence of the countercyclical pattern of relative predictability for the macroeconomic models. The R^2 values are generally higher in times of recession than during times of expansion.

Factor models generated from technical indicators yield the strongest forecasting performance of all categories. At the 1-month forecast horizon, technical factor models consistently outperform

[16] Used by Bai (2010) and Çakmaklı and van Dijk (2016) respectively.

Table 2: LSD values per linear prediction model

1995:12-2014:12	LSD | Overall				LSD | Expansion				LSD | Recession			
Forecast horizon	1	3	6	12	1	3	6	12	1	3	6	12
Macroeconomic												
1F	-11.13	-8.68	-11.54	-5.25	-6.55	-5.98	-5.91	-4.81	-4.58	-2.70	-5.63	-0.44
2F	-8.45	-8.57	-12.92	-13.07	-5.49	-9.91	-6.66	-9.50	-2.96	1.34	-6.26	-3.57
3F	-19.62	-12.97	-33.53	-23.22	-12.62	-13.60	-17.20	-19.38	-7.01	0.63	-16.33	-3.84
Technical												
1F	14.14	-0.12	-1.30	-4.57	11.96	-0.85	-3.32	-4.52	2.18	0.74	2.02	-0.05
2F	37.07	0.80	-7.21	-5.44	35.64	0.36	-6.56	-5.10	1.43	0.44	-0.65	-0.35
3F	39.74	-0.59	-7.60	-10.12	39.01	-0.66	-6.83	-8.46	0.73	0.07	-0.77	-1.67
Sentiment												
1F	-3.43	-2.27	-0.72	5.77	-2.76	-0.92	0.34	4.91	-0.67	-1.35	-1.07	0.87
2F	-4.87	-5.02	0.92	6.61	-3.87	-3.11	0.91	4.79	-1.00	-1.92	0.02	1.83
3F	-5.17	-4.59	-0.28	8.26	-3.99	-2.88	-0.72	5.98	-1.18	-1.71	0.44	2.28
Combined												
1F	6.46	-8.78	-9.68	-7.72	8.15	-7.82	-6.28	-7.34	-1.69	-0.96	-3.40	-0.39
2F	8.26	-10.25	-10.41	-9.21	11.78	-9.09	-9.51	-6.04	-3.52	-1.16	-0.90	-3.18
3F	-9.70	-18.02	-31.18	-23.24	4.51	-14.75	-16.54	-18.86	-14.22	-3.27	-14.65	-4.38
4F	-7.60	-22.89	-31.08	-25.65	8.52	-21.03	-16.91	-20.94	-16.12	-1.86	-14.16	-4.71
Macro+Tech+Sent												
1F+1F+1F	8.65	-9.32	-11.12	-3.45	9.32	-5.39	-7.17	-2.83	-0.66	-3.93	-3.96	-0.61
2F+1F+1F	5.70	-10.96	-14.04	-6.82	7.36	-10.41	-8.76	-4.39	-1.66	-0.55	-5.27	-2.43
1F+2F+1F	32.00	-8.54	-14.70	-2.14	33.04	-4.58	-9.44	-2.42	-1.04	-3.96	-5.26	0.29
Tech+Sent												
1F+1F	13.25	-2.32	-1.05	1.03	11.09	-0.96	-2.11	0.47	2.16	-1.36	1.06	0.56
2F+1F	36.57	-1.07	-7.91	0.19	35.22	-0.07	-6.43	-0.73	1.35	-1.00	-1.48	0.93

Notes: The column on the left represent the predictors. *#F* indicates the number of factors the PLS procedure generates of the respective category. *Combined* means that PLS generates factors from the combined set of all predictive variables. *(Macro+)Tech+Sent* means that a certain number of factors are generated from each of the two or three categories separately. The accuracy is measured relative to the historical average benchmark model, over the out-of-sample period of 1995:12 to 2014:12. The *overall* column represents the entire period, whereas the *expansion* and *recession* columns represent the sub-samples of economic expansion or recession, respectively, as measured by NBER. See equation 18 in section 3.4 for more information on the evaluation criterion.

the uninformative benchmark. The strong performance is present during both recessions and expansions. At longer forecast horizons, the technical models do not perform as well. These results are in line with the findings of Neely et al. (2014). In fact, the results show that a 2-factor model can improve the performance of the 1-factor model considered by Neely et al. (2014) even further. Evidence of the existence of a countercyclical pattern is mixed. Although the results show stronger performance of the technical factor models during recessions when considering longer time horizons (6 and 12 months), the pattern is not present when considering shorter time horizons (1 and 3 months).

Models generated from sentiment variables show the opposite pattern of the technical models. The models slightly underperform the benchmark at the 1 and 3-month forecast horizon, but consistently outperform it on the 12-month forecast horizon. It is in line with the analysis in

Table 3: Log BF values per linear prediction model

1995:12-2014:12	log BF | Overall				log BF | Expansion				log BF | Recession			
Forecast horizon	1	3	6	12	1	3	6	12	1	3	6	12
Macroeconomic												
1F	-10.12	-8.62	-12.29	-5.07	-5.59	-5.37	-6.21	-4.24	-4.53	-3.25	-6.07	-0.83
2F	-8.26	-9.13	-13.55	-12.76	-5.43	-10.31	-7.23	-9.58	-2.83	1.18	-6.31	-3.18
3F	-18.52	-13.75	-34.31	-23.61	-11.91	-13.84	-17.90	-19.61	-6.61	0.09	-16.42	-4.01
Technical												
1F	14.99	-0.56	-1.55	-4.07	12.89	-0.66	-3.21	-4.20	2.10	0.10	1.66	0.13
2F	37.97	0.52	-7.75	-5.45	36.52	0.69	-7.26	-5.52	1.45	-0.17	-0.49	0.07
3F	40.23	-1.23	-7.45	-9.72	39.16	-1.13	-6.89	-8.24	1.07	-0.10	-0.56	-1.48
Sentiment												
1F	-2.92	-2.50	-0.38	5.96	-1.97	-0.59	0.56	4.89	-0.95	-1.92	-0.94	1.07
2F	-4.53	-5.11	0.35	7.71	-3.75	-2.94	0.23	5.33	-0.77	-2.17	0.12	2.38
3F	-4.20	-5.22	-0.62	7.97	-3.48	-3.08	-0.56	5.30	-0.72	-2.14	-0.06	2.67
Combined												
1F	6.99	-9.36	-10.34	-7.76	8.57	-7.55	-6.65	-7.28	-1.58	-1.81	-3.68	-0.48
2F	8.40	-11.38	-11.15	-9.09	11.47	-9.63	-9.94	-6.45	-3.06	-1.76	-1.21	-2.64
3F	-9.67	-18.36	-31.61	-22.88	4.46	-14.45	-17.00	-18.62	-14.13	-3.92	-14.61	-4.26
4F	-6.65	-23.70	-31.16	-24.74	8.57	-21.15	-16.79	-20.49	-15.22	-2.55	-14.37	-4.25
Macro+Tech+Sent												
1F+1F+1F	8.57	-9.71	-11.47	-2.81	9.32	-5.08	-7.52	-2.55	-0.75	-4.63	-3.95	-0.25
2F+1F+1F	6.45	-11.26	-14.24	-7.22	8.07	-10.29	-8.92	-5.01	-1.62	-0.97	-5.32	-2.20
1F+2F+1F	32.74	-8.41	-15.42	-1.92	33.48	-4.16	-9.90	-2.31	-0.74	-4.25	-5.52	0.39
Tech+Sent												
1F+1F	14.02	-2.32	-1.22	1.36	11.50	-1.02	-2.11	0.30	2.53	-1.30	0.89	1.06
2F+1F	37.21	-0.99	-7.86	-0.19	35.62	-0.06	-6.48	-0.90	1.58	-0.93	-1.38	0.71

Notes: The column on the left represent the predictors. *#F* indicates the number of factors the PLS procedure generates of the respective category. *Combined* means that PLS generates factors from the combined set of all predictive variables. *(Macro+)Tech+Sent* means that a certain number of factors are generated from each of the two or three categories separately. The accuracy is measured relative to the historical average benchmark model, over the out-of-sample period of 1995:12 to 2014:12. The *overall* column represents the entire period, whereas the *expansion* and *recession* columns represent the sub-samples of economic expansion or recession, respectively, as measured by NBER. See equation 20 in section 3.4 for more information on the evaluation criterion.

appendix B that shows how the technical and sentiment variables are complementary. However, note that the performance at the 1-month horizon is slightly weaker than the performance reported by Huang et al. (2015). A difference of the evaluated sample and the removal of one of the sentiment variables since the publication of their paper can account for this difference. At the 1 and 12-month horizon, the countercyclical predictability pattern is also present.

The results report two approaches of putting the different categories together into a single model. The first is to combine all variables into a single data set from which the PLS procedure generates factors. Table 1 denotes it as the *combined* prediction model. The other approach is to generate factors from each category of variables (macroeconomic, technical and sentiment) separately. The tables denotes it as the *macro+tech+sent* case or *tech+sent*, where macroeconomic variables are excluded. When comparing models of the *combined* case to the *macro+tech+sent* case with equal amounts of predictive variables, the results show that the *macro+tech+sent* case yields stronger results across the board. It suggests that the strategy of separating different classes of variables before generating factors is a fruitful strategy to use. The weakness of the macroeconomic factors drive much of the difference. The inclusion of macroeconomic factors reduces the accuracy of the models in most cases. This is in line with the findings of Huang et al. (2015), who show that a factor model generated from technical indicators performs much better than a factor model generated from both technical and macroeconomic variables. As a consequence, the *tech+sent* models are some of the strongest models considered in the set of linear prediction models. As they utilize the strengths of both the technical and sentiment variables, they perform relatively well at both short and long forecast horizons.

Tables 2 and 3 show the forecast accuracy of the linear models when evaluating the forecast densities. In general, the results confirm the findings based on table 1. When considering the forecast densities, the existence of the countercyclical pattern of relative predictability is clearer than in the case of point forecasts. In the latter case, the countercyclical pattern was often limited to the 6 and 12-month forecast horizon. For the predictive densities, the pattern is also present for most 1 and 3-month forecast horizons. The only case where the models clearly perform better during expansions than in recessions is for models including technical factors at the 1-month forecast horizon. It shows that there is still room for gains to be made in the forecast accuracy when fully taking the countercyclical pattern into account.

To sum up, the results suggest that among the categories of equity premium predictors found in the literature, factors generated from technical indicators provide by far the most value. Combining factors of technical indicators with factors of investor sentiment provides some additional value, making the predictability more robust at longer forecast horizons. However, the results also suggest that the predictive value of macroeconomic factors is more limited. When using PLS, the results do not yield evidence of positive overall predictive performance of macroeconomic factors. It is both the case when macroeconomic factors are the only predictive variables, and also when combined with technical and sentiment factors. It suggests that pre or post-selection techniques introduced by Çakmaklı and van Dijk (2016) and Bai (2010) respectively are necessary to derive value from macroeconomic factors. Also note that the results do not find confirmation of the existence of countercyclical predictability across all settings at the evaluated sample period. In fact, the performance of the technical factors at the 1-month forecast horizon shows evidence of cyclicality rather than countercyclicality.

5.2 Results Advanced Prediction Models

Table 4: Out-of-sample R^2 values per advanced prediction model

1995:12-2014:12	R^2 | Overall				R^2 | Expansion				R^2 | Recession			
Forecast horizon	1	3	6	12	1	3	6	12	1	3	6	12
Linear	0.243	-0.075	-0.132	-0.027	0.265	-0.067	-0.131	-0.060	0.183	-0.095	-0.136	0.062
GARCH	0.252	-0.072	-0.137	-0.019	0.268	-0.053	-0.127	-0.050	0.208	-0.121	-0.164	0.063
SV (Random Walk)	0.238	-0.057	-0.125	-0.002	0.261	-0.036	-0.118	-0.026	0.178	-0.114	-0.141	0.062
SV (Stationary)	0.240	-0.055	-0.117	0.007	0.262	-0.032	-0.112	-0.012	0.183	-0.116	-0.132	0.056
MS (Basic)	0.242	-0.068	-0.114	0.008	0.267	-0.051	-0.110	-0.017	0.173	-0.115	-0.126	0.073
MS (Full)	0.238	-0.067	-0.129	-0.032	0.264	-0.052	-0.117	-0.071	0.171	-0.107	-0.159	0.072

Notes: The column on the left represent the predictors and model specification (see section 3.2 for more details on the models). Each model uses the *Macro+Tech+Sent 1F+2F+1F* specification of the previous section, meaning that for the categories of macroeconomic, technical, and sentiment variables the PLS procedure generates one, two, and one factor, respectively. The accuracy is measured relative to the historical average benchmark model, over the out-of-sample period of 1995:12 to 2014:12. The *overall* column represents the entire period, whereas the *expansion* and *recession* columns represent the sub-samples of economic expansion or recession, respectively, as measured by NBER. See equation 16 in section 3.4 for more information on the evaluation criterion.

Table 5: LSD values per advanced prediction model

1995:12-2014:12	LSD | Overall				LSD | Expansion				LSD | Recession			
Forecast horizon	1	3	6	12	1	3	6	12	1	3	6	12
Linear	32.00	-8.54	-14.70	-2.14	33.04	-4.58	-9.44	-2.42	-1.04	-3.96	-5.26	0.29
GARCH	42.63	-3.00	-12.99	-5.38	36.66	1.44	-7.00	-0.98	5.97	-4.44	-5.98	-4.40
SV (Random Walk)	40.68	0.90	-10.91	-24.67	35.05	5.73	-3.07	-3.41	5.63	-4.84	-7.84	-21.27
SV (Stationary)	43.62	0.86	-15.08	-7.03	38.77	5.45	-5.69	2.89	4.85	-4.59	-9.39	-9.92
MS (Basic)	23.95	-28.53	-44.50	-31.79	25.97	-11.28	-17.68	-5.88	-2.02	-17.25	-26.82	-25.90
MS (Full)	26.68	-21.00	-30.64	-10.08	31.73	-7.84	-10.87	-3.58	-5.05	-13.16	-19.77	-6.50

Notes: The column on the left represent the predictors and model specification (see section 3.2 for more details on the models). Each model uses the *Macro+Tech+Sent 1F+2F+1F* specification of the previous section, meaning that for the categories of macroeconomic, technical, and sentiment variables the PLS procedure generates one, two, and one factor, respectively. The accuracy is measured relative to the historical average benchmark model, over the out-of-sample period of 1995:12 to 2014:12. The *overall* column represents the entire period, whereas the *expansion* and *recession* columns represent the sub-samples of economic expansion or recession, respectively, as measured by NBER. See equation 18 in section 3.4 for more information on the evaluation criterion.

Table 4 presents the results that measure the forecast accuracy of the advanced econometric models explained in section 3.2. The predictors of each model consist of a single macroeconomic factor, two technical factors, and a single sentiment factor. The reason behind testing this specific predictor set is two-fold. First of all, tables 1 to 3 show that models consisting of two technical factors and one sentiment factor show consistently good results across multiple forecast horizons. In an attempt to attain high predictive performance, these predictors show to be the most suitable. Secondly, the macroeconomic factors showcase the highest degree of countercyclical predictability. As the advanced models attempt to account for the countercyclical pattern, it is expected that the

Table 6: Log BF values per advanced prediction model

1995:12-2014:12	log BF | Overall				log BF | Expansion				log BF | Recession			
Forecast horizon	1	3	6	12	1	3	6	12	1	3	6	12
Linear	32.74	-8.41	-15.42	-1.92	33.48	-4.16	-9.90	-2.31	-0.74	-4.25	-5.52	0.39
GARCH	43.34	-0.66	-11.60	-0.82	37.32	2.66	-6.80	0.07	6.02	-3.32	-4.80	-0.90
SV (Random Walk)	43.34	3.57	-8.38	-8.28	37.26	5.77	-3.54	-1.36	6.07	-2.20	-4.84	-6.92
SV (Stationary)	44.74	1.59	-15.36	-6.95	39.45	4.39	-9.27	-3.54	5.30	-2.81	-6.09	-3.41
MS (Basic)	42.80	3.89	-5.60	2.98	38.71	5.03	-2.87	2.49	4.09	-1.14	-2.73	0.49
MS (Full)	40.63	3.41	-11.11	2.58	37.58	2.48	-6.82	0.79	3.05	0.93	-4.29	1.79

Notes: The column on the left represent the predictors and model specification (see section 3.2 for more details on the models). Each model uses the *Macro+Tech+Sent 1F+2F+1F* specification of the previous section, meaning that for the categories of macroeconomic, technical, and sentiment variables the PLS procedure generates one, two, and one factor, respectively. The accuracy is measured relative to the historical average benchmark model, over the out-of-sample period of 1995:12 to 2014:12. The *overall* column represents the entire period, whereas the *expansion* and *recession* columns represent the sub-samples of economic expansion or recession, respectively, as measured by NBER. See equation 20 in section 3.4 for more information on the evaluation criterion.

predictive value of macroeconomic predictors is most inclined to increase.

When comparing the R^2 values of the advanced models to the linear model, the differences are quite small across the board. There is also no clear winner among any of the models in terms of accuracy. At the 1-month forecast horizon, the GARCH model attains small improvements over the linear model. The improvement is larger during times of recession. At the 12-month forecast horizon, the stationary stochastic volatility model performs best among the time-varying volatility models. The results show that it is able to outperform both the linear models and the historical average benchmark at this time horizon. Among the Markov switching models, there is no large difference between the basic MS model and the full MS model. In fact, at the 12-month forecast horizon, the basic MS model is performing slightly better and is able to beat both the linear model and historical average benchmark. This contradicts the hypothesis that the countercyclical pattern of equity premium predictability is driven by a regime-dependent nature of the relation between the predictors and the equity premium. Assuming regime-dependency of the predictors' parameters is associated with little to no improvements of the accuracy of point forecasts. The fact that the MS models do not produce much superior out-of-sample predictions compared to the linear model contradicts the findings of Henkel et al. (2011). The major difference between the full MS model of this thesis and that of Henkel et al. (2011) is their use of only four individual predictors. In their paper Çakmaklı and van Dijk (2016) report that the use of dimension reduction techniques yields much more consistent predictive performance than the use of individual predictors. Hence, one could infer that the PLS method already accounts for much of the regime-dependent nature of the individual macroeconomic predictors. It makes the use of explicit regime-dependent parameters superfluous, which is what the results confirm.

Tables 5 and 6 show the forecast accuracy of the advanced models when evaluating the predictive densities. Differences between the linear models and the advanced models are more obvious in this set of results compared to those of table 4. The accuracy of the predictive density is especially relevant for models assuming time-varying volatility, as the error process is specified more explicitly compared to linear models.

At the shorter forecast horizons, all time-varying volatility models show improvements over the linear model when focusing on the accuracy of the predictive densities. Only at the 12-month forecast horizon does the linear model remain superior over most time-varying volatility models. The improvements occur during both times of expansion and recession. The best-performing time-varying volatility model is the stationary SV model, which outperforms the historical average benchmark at both the 1 and 3-month forecast horizon. The random walk SV model yields similar results, but struggles at the 12-month time horizon. A likely explanation is that the random walk SV is a more restricted version of the stationary SV model. If the restriction is unjustified, the specification error accumulates when rolled further into the future. In contrast, both the GARCH and stationary SV model are more general and are able to maintain decent performance on longer time horizons. In fact, the Bayes factor of the GARCH model is higher at the 12-month forecast horizon than that of the linear model.

With regards to the Markov switching models, the evidence is mixed between the log score differentials and the Bayes factors. As the log score specifically penalizes the variance of the posterior distributions, it clearly hurts the performance of the basic MS model. Appendix D provides an indication of why this might be the case. The basic MS model tends to assign relatively indecisive probabilities to the unobserved states. It means that the predictive distributions is usually an almost equal combination between two very dissimilar distributions, which increases the variance of the predictive distributions. The Bayes factors might provide a clearer picture on the accuracy of the predictive densities. Indeed, the results are more in line with those of the R^2 measure. Both Markov switching models provide clear benefits over the linear model. This is the case at each forecast horizon. The improvements occur during both expansions and recessions. However, when comparing the two Markov switching models, the Bayes factors of the basic MS model are slightly superior to those of the full MS model. It again contradicts the hypothesis that regime-dependent relations between predictors and the equity premium drive the phenomenon of countercyclical predictability.

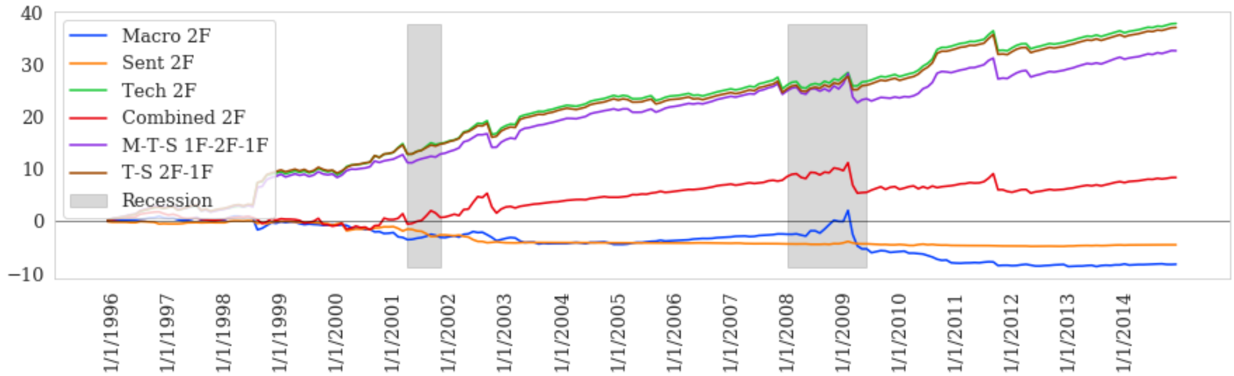
To sum up, alternative specifications of the equity premium's volatility process can lead to improved forecasting performance. However, the evidence on which specification is superior is mixed. At short forecasting horizons, GARCH and stochastic volatility models yield the strongest results. At larger forecast horizons, however, the Markov switching model that specifies both the intercept and the volatility parameter as regime-dependent provides stronger results. As countercyclical predictability is not present at all forecast horizons for the evaluated predictor set, it is unclear whether the improvement is attained by successfully accounting for this phenomenon. However, at the forecast horizons for which the phenomenon is present (6 and 12 months), expansionary periods attain the largest improvements of predictive accuracy. At the same time, at the horizon for which there is clear cyclical predictability, the biggest gain in improvement occurs during recessions. Hence, the results suggest that alternative specifications of the volatility process do lead to more consistent results across the business cycle as well as higher accuracy overall. The results do not support the hypothesis that regime-dependent relations between predictors and the equity premium are the main driver of (counter)cyclical predictability.

5.3 Predictive Performance Over Time

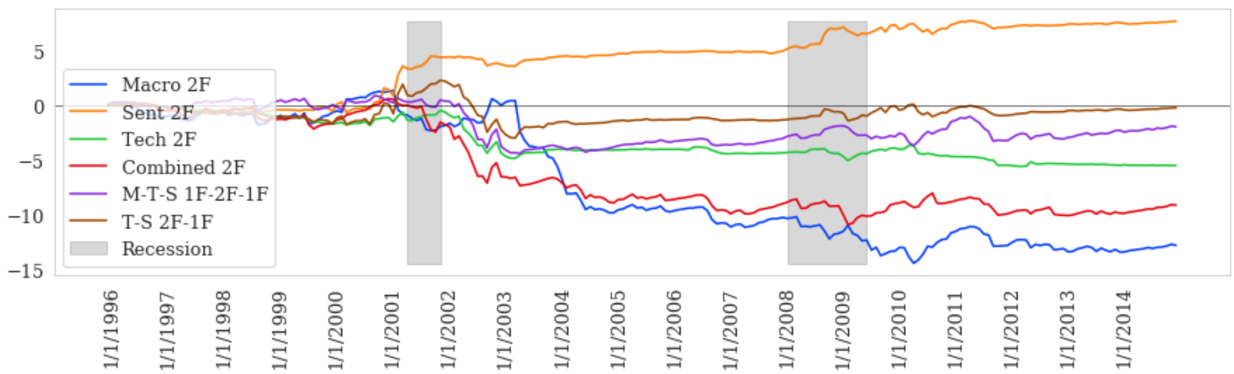
In order to dig deeper into the predictive performance of both the linear and advanced models, this section presents the evolution of the predictive performance over time. The results restrict themselves to the log Bayes factors over time for a selection of the models presented in the previous paragraphs. Appendix E presents additional results.

Figure 1 displays the log Bayes factor over time for a selection of the linear predictive models. It shows 2-factor models generated from the macroeconomic, sentiment, and technical category respectively, and also three models using multiple categories. Starting with the forecast horizon of 1 month, the predictive performance shows to be relatively stable. The models that end up as the best-performing models (such as the technical factor model, the *macro+tech+sent* model, and the *tech+sent* model) start their strong performance early in the evaluated sample period, and continue to increase further in time. A slight exception is the macroeconomic model, which shows to have a few periods in which it is outperforming the benchmark. It is the case during, for instance, the recession in the early 2000's, and the run-up to and first half of the 2008 recession. However, a significant decrease in predictability towards the end of the 2008 recession and the period afterwards reduces the performance of the model to a large degree.

Figure 1: Log Bayes factors over time for linear prediction models



(a) Forecast horizon = 1 month



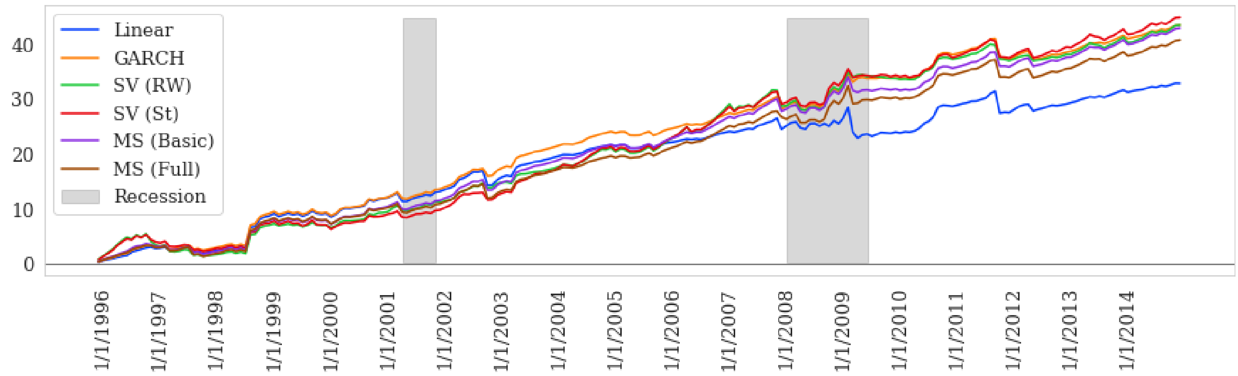
(b) Forecast horizon = 12 months

Note: The figures represent the evolution of the log Bayes factors of a selection of the linear prediction models over time. The naming convention is similar to tables 1 to 6. However, *Macro*, *Tech*, and *Sent* are abbreviated as *M*, *T*, and *S*, respectively, in some of the legend entries. Gray areas represent time periods denoted as recessions by NBER. The horizontal line at 0 represents the performance of the historical average benchmark model. Hence, upward sloping lines represent models that are outperforming the benchmark at the respective period, and vice versa. A value above zero means that a respective model has outperformed the benchmark up to the respective time period, and vice versa. See equation 21 for details on the evaluation criterion.

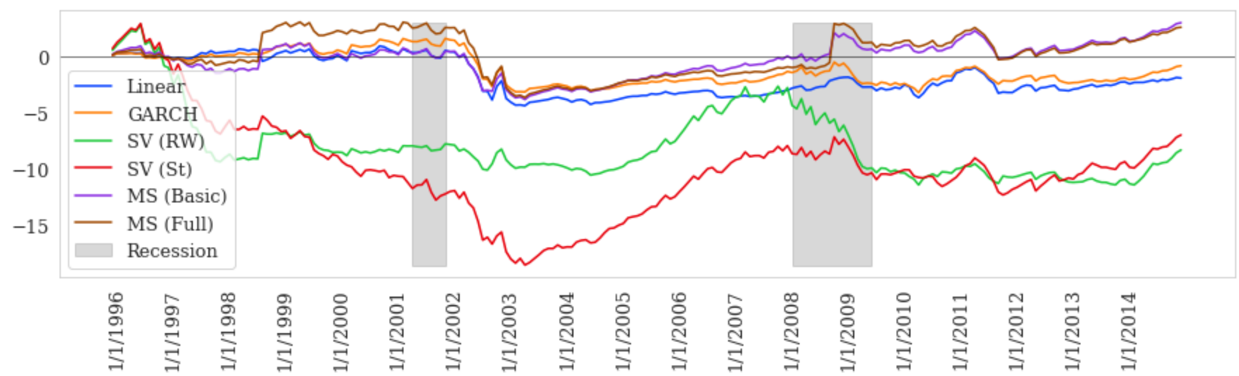
At the 12-month forecast horizon, the predictability of the different models also show some interesting patterns. The predictability of the models seems relatively in line with the historical average benchmark during the more quiet economic periods. However, in and around the two recessionary periods, the models start to diverge. A clear winner at this forecast horizon is the sentiment factor model. However, it clearly derives most of its strong performance in the run-up to the dot-com crash and the two recessionary periods, as the performance is relatively in line with the benchmark during other periods.

These results show that the distinction between the performance during periods of expansion and recession made earlier does not tell the full story of the countercyclical predictability phenomenon.

Figure 2: Log Bayes factors over time for advanced prediction models



(a) Forecast horizon = 1 month



(b) Forecast horizon = 12 months

Note: The figures represent the evolution of the log Bayes factors of a selection of the advanced prediction models over time. The factors used as predictors consist of one macroeconomic factor, two technical factors, and one sentiment factor (hence, the Macro+Tech+Sent 1F-2F-1F model referred to earlier). The naming convention is the same as maintained in tables 1 to 6. *RW* refers to the random walk variant of SV models, whilst *St* refers to the stationary variant. Gray areas represent time periods denoted as recessions by NBER. The horizontal line at 0 represents the performance of the historical average benchmark model. Hence, upward sloping lines represent models that are outperforming the benchmark at the respective period, and vice versa. A value above zero means that a respective model has outperformed the benchmark up to the respective time period, and vice versa. See equation 21 for details on the evaluation criterion.

Periods just before and after the identified recessions also have a pattern of predictability distinct from more stable periods.

Figure 2 shows the log Bayes factors over time for the advanced predictive models. The predictor set used by each model is the one using a macroeconomic factor, two technical factors, and a sentiment factor. These are the same models as those in table 4 to 6. At the 1-month forecast horizon, the models almost all follow a very similar trajectory. Up until the 2008 recession, there is also not much difference between the advanced models and the linear model. Only from 2008 onward, at which a structural break seems to occur, do the performances of the linear model versus the advanced models diverge in favour of the advanced models. The advanced models all follow a similar trajectory throughout the entire evaluated sample, with the stochastic volatility models ending on top.

At the 12-month forecast horizon, larger differences exist. Especially the two stochastic volatility models have prolonged periods of very weak performance, as well as long periods of extremely strong performance. There seems to be a structural break around 2003, when the weak performance of these models turns into very strong performance, lasting until the 2008 financial crisis. The GARCH model is the only time-varying volatility model characterized by a different pattern. Its performance diverges very little with the linear model, suggesting that the model might reduce to something close to a constant-volatility model. The accuracy of the two MS models again diverges very little across the predicted period, suggesting that the full MS model reduces towards the basic MS model. The performance of the MS models again follow the pattern of the linear model closely, but slightly favouring the MS models, across the predicted period.

6 Market Timing Analysis

As an additional robustness check, it is important to analyze whether the forecast accuracy of the predictive models actually translates into economic value for investors. Hence, this section measures the investment performance an arbitrary investor can attain when using the 1-step ahead forecasts as part of a monthly market timing strategy.

6.1 Methodology

The setup for the market timing study is similar to those of Çakmaklı and van Dijk (2016) and Neely et al. (2014). A mean-variance investor chooses to invest a portion of the investor's wealth in the stock market (the S&P 500 index in this case) and another portion in a risk-free product (the 1-month T-bill). The investment period lasts from December 1995 to December 2014. The investor adjusts the proportions invested each month, based on the new information generated by the respective forecast model. The market timing study considers a utility-theory framework, which assumes that the investor is risk-averse and attempts to maximize utility by investing based on the risk premium forecasts. The maximization problem reads as follows:

$$\max_{w_{t+1}} \mathrm{U}(r_{p,t+1}) = \max_{w_{t+1}} \mathrm{E}_t \ (r_{p,t+1}) - \frac{1}{2}\gamma \ \mathrm{Var}_t \ (r_{p,t+1}), \tag{26}$$

where w_t is the proportion allocated to stocks at time t, $r_{p,t}$ is the return of the allocated portfolio at time t, $\text{U}(r_{p,t})$ is the utility of the portfolio, $\text{E}_t(r_{p,t+1})$ is the conditional expected value of the portfolio, γ is the investor's coefficient of relative risk aversion, and $\text{Var}_t(r_{p,t+1})$ is the conditional variance of the portfolio return. When using two products, a market portfolio and a risk-free product, the portfolio return is given by:

$$r_{p,t+1} = r_{f,t+1} + w_{t+1} r_{e,t+1}, \tag{27}$$

where $r_{f,t}$ is the risk-free rate at time t, and $r_{e,t}$ is the market's risk premium at time t. Assuming that the next month's risk-free rate is fixed at time t, the optimal weight for stocks, following Çakmaklı and van Dijk (2016), is as follows:

$$w^*_{t+1} = \frac{\text{E}_t(r_{e,t+1})}{\gamma \text{Var}_t(r_{e,t+1})} \tag{28}$$

The conditional expectation of the risk premium $\text{E}_t(r_{e,t+1})$ is calculated by taking the mean of the sample of the posterior predictive distribution, generated by the Bayesian MCMC sampler: $\text{E}_t(r_{e,t+1}) = \text{E}_t(\{r^{(1)}_{e,t+1}, ..., r^{(S)}_{e,t+1}\})$. As for the conditional variance $\text{Var}_t(r_{e,t+1})$, this thesis considers two approaches. The first is the more common approach of using posterior mean of the scale parameter of the volatility process at the forecast period σ_{t+1}: $\text{Var}_t(r_{e,t+1}) = \text{E}_t(\{\sigma^{2,(1)}_{t+1}, ..., \sigma^{2,(S)}_{t+1}\})$. This represents the expected volatility of next month's returns. An additional step of the MCMC sampler can generate a sample of draws of this parameter. Taking the mean of this sample results in the expected variance of next month's equity premium. The scale parameter of the volatility process is either constant in the case of constant-volatility models, or a forecast resulting from the estimated volatility process, such as Markov switching, stochastic volatility, or GARCH. The second approach is similar in philosophy to papers like Barberis (2000) and Diris et al. (2014). They propose a fully Bayesian process in which the MCMC sampler produces both draws of parameters and the weights that determine the optimal portfolio allocation. Such a method ensures the optimal weights take the uncertainty of the forecast estimates into account. However, this thesis stays within the context of the proposed out-of-sample study by instead incorporating the parameter uncertainty separately into equation 28. Hence, the second method plugs in the variance of the sample of the posterior predictive distribution generated by the Bayesian MCMC sampler, meaning that $\text{Var}_t(r_{e,t+1}) = \text{Var}_t(\{r^{(1)}_{e,t+1}, ..., r^{(S)}_{e,t+1}\})$. Instead of next month's volatility, it represents the full uncertainty around next month's returns. Hence, when the model signals that returns are more uncertain, the investor's allocation into stocks will be closer to zero.

The market timing study also takes into account transaction costs, following the method introduced by DeMiguel et al. (2007). The process calculates the turnover each month, which is the sum of the absolute changes of weight invested in both products. It represents the percentage of wealth traded in the respective month. Over the turnover, 0.1% transaction costs are calculated, which are subtracted from the monthly portfolio return. Çakmaklı and van Dijk (2016) report this number as *medium transaction costs*.[17] The reported results also follow Çakmaklı and van Dijk (2016) in the fact that the proportion invested in the market is constrained at -1 and 2, allowing for limited short-selling and leverage.[18]

6.2 Results

The study uses two evaluation criteria to assess the risk-adjusted profitability of the investment strategies: the Sharpe ratio and the certainty equivalent return (CER). They read as follows:

$$\begin{aligned} \text{Sharpe Ratio: } &= \frac{\hat{\mu}_p}{\hat{\sigma}_p} \\ \text{CER: } &= \hat{\mu}_p - \frac{1}{2}\gamma\hat{\sigma}_p^2, \end{aligned} \tag{29}$$

with $\hat{\mu}_p$ representing the mean of the respective portfolio's monthly returns and $\hat{\sigma}_p$ denoting the standard deviation of the portfolio's returns. The results report the annualized versions of both measures. Where relevant, the discussion of the results includes outcomes of the robust Sharpe ratio difference test of Ledoit and Wolf (2008). The test allows the researcher to evaluate the hypothesis of whether the investment performance of one model is significantly better than another.

Table 7 presents the results of the market strategies that make use of all different instances of the models introduced in earlier sections. The first thing to note is that the difference between the two methods of optimizing the weights is minimal. Using the return's volatility forecast leads to very similar weights as using the estimation uncertainty of the return forecasts.[19] This is in line with the results of Barberis (2000), who shows that for certain time periods, at short investment horizons, taking parameter uncertainty into account makes little difference to the resulting weights.

Among the different sets of predictors, the results are similar to those of section 5.1. Any model involving technical factors produces the best forecasts and also the best investment performance

[17] Unreported results show that the results are similar when higher or lower transaction costs are used.
[18] Appendix E.2 reports results where the constraint is 0 to 1, not allowing for short-selling and leverage.
[19] Appendix F provides an example for the weights and portfolio returns of one of the models across the entire evaluated time period.

Table 7: Results market timing study

$\gamma = 6$	Weights Determined Using Volatility Forecast						Weights Determined Using Estimation Uncertainty					
1995:12-2014:12	Sharpe Ratio			CER			Sharpe Ratio			CER		
	Full	Exp	Rec	Full	Exp	Rec	Full	Exp	Rec	Full	Exp	Rec
Constant Weights												
100%	0.465	0.754	-0.811	4.10%	7.31%	-17.54%						
50%	0.464	0.753	-0.812	4.36%	6.54%	-10.97%						
0%	-	-	-	2.50%	2.59%	1.81%						
Benchmark												
Historical Average	0.459	0.679	-0.916	4.26%	6.36%	-10.48%	0.459	0.679	-0.910	4.27%	6.36%	-10.40%
Linear Models												
Macroeconomic												
1F	0.054	0.099	-0.153	-4.77%	-2.83%	-19.23%	0.045	0.090	-0.162	-4.79%	-2.86%	-19.20%
2F	0.157	0.119	0.394	-3.09%	-3.25%	-2.24%	0.160	0.122	0.398	-2.99%	-3.15%	-2.14%
3F	-0.051	-0.146	0.501	-6.48%	-7.15%	-1.26%	-0.052	-0.146	0.493	-6.43%	-7.09%	-1.34%
Technical												
1F	1.427	1.533	0.883	19.59%	21.28%	6.60%	1.428	1.536	0.881	19.50%	21.18%	6.57%
2F	2.134	2.147	2.045	39.12%	38.95%	39.67%	2.134	2.148	2.046	39.11%	38.95%	39.64%
3F	2.154	2.173	2.032	39.71%	39.67%	39.27%	2.155	2.174	2.033	39.69%	39.65%	39.23%
Sentiment												
1F	0.209	0.407	-0.736	-0.02%	3.14%	-21.40%	0.208	0.408	-0.749	0.14%	3.22%	-20.76%
2F	0.101	0.266	-0.677	-1.12%	1.60%	-19.98%	0.097	0.266	-0.696	-1.03%	1.69%	-19.83%
3F	0.122	0.294	-0.686	-0.95%	1.87%	-20.40%	0.123	0.295	-0.677	-0.78%	1.98%	-19.92%
Combined												
1F	1.213	1.269	0.871	15.55%	16.63%	6.92%	1.207	1.263	0.870	15.41%	16.47%	6.94%
2F	1.290	1.377	0.825	16.87%	18.33%	5.59%	1.294	1.379	0.830	16.91%	18.34%	5.83%
3F	1.040	1.052	0.975	12.08%	12.41%	8.91%	1.040	1.052	0.973	12.08%	12.42%	8.87%
4F	1.267	1.234	1.486	16.78%	16.11%	21.70%	1.269	1.237	1.485	16.81%	16.16%	21.67%
Tech+Sent												
1F+1F	1.383	1.482	0.836	19.00%	20.77%	5.46%	1.386	1.486	0.836	19.02%	20.80%	5.44%
2F+1F	2.158	2.175	2.047	39.66%	39.56%	39.64%	2.159	2.174	2.051	39.58%	39.49%	39.47%
Macro+Tech+Sent												
1F+1F+1F	1.228	1.287	0.878	15.83%	16.96%	6.81%	1.226	1.287	0.872	15.78%	16.92%	6.67%
2F+1F+1F	1.122	1.164	0.886	13.74%	14.59%	6.69%	1.120	1.160	0.888	13.68%	14.51%	6.75%
1F+2F+1F	2.048	2.091	1.739	35.62%	36.39%	29.05%	2.045	2.088	1.741	35.50%	36.25%	29.09%
Advanced Models (Macro+Tech+Sent 1F+2F+1F)												
GARCH	2.024	2.055	1.780	34.71%	35.28%	29.65%	2.025	2.058	1.772	34.74%	35.35%	29.43%
SV (Random Walk)	1.960	2.016	1.568	32.55%	33.61%	23.93%	1.957	2.010	1.577	32.41%	33.43%	24.13%
SV (Stationary)	2.021	2.071	1.657	34.67%	35.67%	26.47%	2.019	2.069	1.659	34.63%	35.61%	26.52%
MS (Basic)	2.034	2.082	1.699	35.15%	36.00%	27.95%	2.073	2.122	1.717	36.80%	37.81%	28.51%
MS (Full)	2.095	2.120	1.896	37.08%	37.48%	33.28%	2.110	2.134	1.905	37.86%	38.33%	33.54%

Notes: This table shows the results of a hypothetical investment strategy where an investor allocates a proportion of wealth in the market and another in a risk-free product. As a comparison, the table also shows the results of investing using constant weights, where the investor rebalances each month in order to invest the respective proportion in the market and the remaining wealth in the risk-free product. The evaluated investment period is 1995:12 to 2014:12, where the investor reallocates the investments monthly by optimizing the weights according to equation 28 with $\gamma = 6$. The weight invested in the market is constrained between -1 and 2, allowing for a limited degree of short-selling and leverage. 0.1% transaction costs over the respective turnover is subtracted from each portfolio's monthly returns. The column on the left represent the predictors and model specification (see section 3.2 for more details on the models). Each advanced model uses the *Macro+Tech+Sent 1F+2F+1F* specification of the previous section, meaning that for the categories of macroeconomic, technical, and sentiment variables the PLS procedure generates one, two, and one factor, respectively. The *overall* column represents the entire period, whereas the *expansion* and *recession* columns represent the sub-samples of economic expansion or recession, respectively, as measured by NBER. See equation 29 for the evaluation criteria.

when using these forecasts for market timing. All models involving at least one technical factor outperform both the historical average model and the constant weights strategies. Using the Ledoit and Wolf (2008) Sharpe ratio difference test, the results show that two technical factors significantly outperform the use of a single technical factor at the 5% significance level. However, adding a third factor does not significantly improve the investment performance further. The test can also assess the difference of the Sharpe ratio between the predictive models and the historical average model. It shows that all models using technical factors significantly outperform the benchmark at the 5% level. The same goes for the *combined* models, with the exception of the version that uses three factors. The conclusions hold for *volatility forecast* case and the *estimation uncertainty* case.

Note that the market timing study only takes into account the 1-month forecast horizon. Hence, the sentiment and macroeconomic factors do not produce results that are able to outperform either the benchmark model or the constant weights strategies. When assessing the difference between the Sharpe ratio of the *tech+sent 2F+1F* and the 2-factor technical model, one can see whether a sentiment factor provides any additional economic value over using only two technical factors. At the 5% significance level, there is no evidence that this is the case.

Whereas section 5.2 shows that the advanced models provide added value over the linear model, it does not seem to translate into additional economic value. The stochastic volatility, GARCH and Markov switching models all yield very similar Sharpe ratios and CERs to their linear counterpart (the *macro+tech+sent 1F+2F+1F* model). The best-performing advanced model in this case is the full Markov switching model, which does yield slightly higher performance measures than its linear counterpart. However, when testing the difference of the full MS model with its linear counterpart using the Ledoit and Wolf (2008) test, there is no evidence that it is significant at the 5% level. The test yields the same result when comparing any of the other advanced models with the linear model. An indication for why this is the case can be found in section 5.3. It shows that the strong performance of the advanced models occurs not across the entire sample, but rather during more limited time periods at the end of the sample. This small time period does not seem long and fruitful enough for the investment strategy to reap the benefits of additional predictive power in terms of returns. Hence, a longer time period of strong performance by the advanced models may be needed for the investor to reap the full benefits of them.

In conclusion, the market timing study confirms most of the results of the forecast study. At the 1-month forecast horizon, the technical indicators yield the most economic value, especially when multiple factors are used. On the other hand, the macroeconomic and sentiment indicators do not provide economic value over the benchmark model or basic constant weight strategies. Whilst the forecast study finds that the advanced models improve the forecast accuracy of the predictive models, the market timing study shows that this improvement cannot be converted into additional economic value. This may be attributed to the short time period at which the advanced models outperform their linear counterpart. Finally, the results show that there is little difference between optimizing weights using volatility forecasts or using estimation uncertainty, as both methods lead to similar weights and almost identical investment performance.

7 Conclusion

This thesis uses various categories of PLS factors to forecast the equity risk premium. The results show that factors generated from technical indicators yield the highest predictive power at short forecast horizons. Factors generated from sentiment factors have value at longer forecast horizons. Combining the two categories yields robust forecast performance across multiple forecast horizons. The results do not provide evidence that macroeconomic factors generate superior forecast performance over an uninformative model. An analysis that translates the 1-month ahead forecasts into a monthly updated trading strategy confirms that any positive forecast performance also translates into additional economic value.

Furthermore, the results show that alternative specifications of the equity return's volatility process leads to increased predictability. At short forecast horizons, stochastic volatility models produces the strongest results. However, at longer forecast horizons, a regime-dependent volatility process implemented via a Markov switching model proves superior over both linear and stochastic volatility models. Both methods lead to more consistent results across the business cycle. This suggests that, to a certain degree, they effectively account for (counter)cyclical relative predictability of equity returns. However, the additional predictability of regime-dependent and time-varying volatility models do not translate into additional economic value over linear models. Moreover, the results do not provide evidence that regime-dependent relations between predictors and the equity premium are a main driver of the countercyclical pattern.

The results point to some promising avenues of future research. The different methods of incorporating time-varying volatility have varying degrees of success across different forecast horizons and time periods. This suggests that finding improved specifications of the volatility process that are more robust across multiple settings could increase the predictability further. Possible candidates could be the inclusion of jumps, a leverage effect, and heavy-tailed errors, as proposed by Nakajima (2012). Additionally, the results show the value of separating predictive variables into complementary and mutually exclusive categories. Generating factors from separate categories leads to superior predictive accuracy compared to generating factors from a single large category. Hence, separating the macroeconomic categories into multiple complementary categories could be a direction of improvement of predictability.

References

Adolfson, M., Lindé, J., & Villani, M. (2007). Forecasting performance of an open economy DSGE model. *Econometric Reviews*, *26*(2-4), 289–328.

Ang, A., & Bekaert, G. (2006). Stock return predictability: Is it there? *The Review of Financial Studies*, *20*(3), 651–707.

Bai, J. (2010). Equity premium predictions with adaptive macro indexes. *FRB of New York Staff Report*(475).

Bai, J., & Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, *146*(2), 304–317.

Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*, *21*(2), 129–152.

Barberis, N. (2000). Investing for the long run when returns are predictable. *The Journal of Finance*, *55*(1), 225–264.

Bhansali, R. (1999). Parameter estimation and model selection for multistep prediction of a time series: A review. *Asymptotics, Nonparametrics, and Time Series*, 201.

Brown, G. W., & Cliff, M. T. (2004). Investor sentiment and the near-term stock market. *Journal of Empirical Finance*, *11*(1), 1–27.

Çakmaklı, C., & van Dijk, D. (2016). Getting the most out of macroeconomic information for predicting excess stock returns. *International Journal of Forecasting*, *32*(3), 650–668.

Campbell, J. Y., Giglio, S., Polk, C., & Turley, R. (2018). An intertemporal CAPM with stochastic volatility. *Journal of Financial Economics*, *128*(2), 207–233.

Campbell, J. Y., & Thompson, S. B. (2007). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, *21*(4), 1509–1531.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of statistical software*, *76*(1).

Cenesizoglu, T., & Timmermann, A. (2012). Do return prediction models add economic value? *Journal of Banking & Finance*, *36*(11), 2974–2987.

Clark, T. E., & Ravazzolo, F. (2015). Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics*, *30*(4), 551–575.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, *1*(2), 223-236.

DeMiguel, V., Garlappi, L., & Uppal, R. (2007). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *The Review of Financial Studies*, *22*(5), 1915–1953.

Diris, B., Palm, F., & Schotman, P. (2014). Long-term strategic asset allocation: an out-of-sample evaluation. *Management Science*, *61*(9), 2185–2202.

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

Geweke, J., & Amisano, G. (2010). Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting*, *26*(2), 216–230.

Ghysels, E., Valkanov, R. I., & Serrano, A. R. (2009). Multi-period forecasts of volatility: Direct, iterated, and mixed-data approaches. In *Efa 2009 bergen meetings paper.*

Guidolin, M., & Hyde, S. (2012). Can VAR models capture regime shifts in asset returns? a long-horizon strategic asset allocation perspective. *Journal of Banking & Finance*, *36*(3), 695–716.

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

Hamilton, J. D., & Lin, G. (1996). Stock market volatility and the business cycle. *Journal of Applied Econometrics*, *11*(5), 573–593.

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

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, *15*(1), 1593–1623.

Huang, D., Jiang, F., Tu, J., & Zhou, G. (2015). Investor sentiment aligned: A powerful predictor of stock returns. *The Review of Financial Studies*, *28*(3), 791–837.

Ing, C.-K. (2003). Multistep prediction in autoregressive processes. *Econometric Theory*, *19*(2), 254–279.

Jacquier, E., Polson, N. G., & Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business & Economic Statistics*, *12*(4), 69–87.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795.

Kelly, B., & Pruitt, S. (2013). Market expectations in the cross-section of present values. *The Journal of Finance*, *68*(5), 1721–1756.

Ledoit, O., & Wolf, M. (2008). Robust performance hypothesis testing with the Sharpe ratio.

Journal of Empirical Finance, *15*(5), 850–859.

Marcellino, M., Stock, J. H., & Watson, M. W. (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics*, *135*(1-2), 499–526.

Nakajima, J. (2012). Bayesian analysis of generalized autoregressive conditional heteroskedasticity and stochastic volatility: Modeling leverage, jumps and heavy-tails for financial time series. *The Japanese Economic Review*, *63*(1), 81–103.

Neely, C. J., Rapach, D. E., Tu, J., & Zhou, G. (2014). Forecasting the equity risk premium: the role of technical indicators. *Management Science*, *60*(7), 1772–1791.

Ng, K. S. (2013). A simple explanation of partial least squares. *The Australian National University, Canberra*.

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. (2017). Forecasting macroeconomic variables under model instability. *Journal of Business & Economic Statistics*, *35*(2), 183–201.

Pettenuzzo, D., Timmermann, A., & Valkanov, R. (2014). Forecasting stock returns under economic constraints. *Journal of Financial Economics*, *114*(3), 517–553.

Sander, M. (2018). Market timing over the business cycle. *Journal of Empirical Finance*, *46*, 130–145.

Shiller, R. J. (2015). *Irrational exuberance: Revised and expanded third edition.* Princeton University Press.

Stan Development Team. (2018). Stan modeling language users guide and reference manual, version 2.18.0. [Computer software manual]. Retrieved from http://mc-stan.org

Stock, J. H., & Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, *20*(2), 147–162.

Welch, I., & Goyal, A. (2007). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, *21*(4), 1455–1508.

Appendix A MCMC Algorithm Details

A.1 Hamiltonian Monte Carlo: No-U-Turn Sampler

This thesis makes use of a Hamiltonian Monte Carlo algorithm to generate draws directly from the joint posterior density distribution. HMC is a variant of an MCMC algorithm that generates proposals based on Hamiltonian dynamics. The use of HMC has been relatively rare in the Bayesian literature up until recently, as the performance is very dependent on the researcher's choice of several tuning variables. With the introduction of the No-U-Turn Sampler (Hoffman & Gelman, 2014), HMC algorithms can now be implemented without the need to set the tuning variables manually, as this is done automatically by the NUTS algorithm. Hoffman and Gelman (2014) show that the implementation of their algorithm can lead to faster convergence and a smaller degree of autocorrelated draws compared to other MCMC samplers such as Gibbs and random-walk Metropolis.

The introduction of the Stan programming language further facilitates the implementation of performing Bayesian analysis via the NUTS algorithm. The platform has interfaces available for many popular programming languages, such as Matlab, Python and R. This thesis makes use the Stan language via Python at each estimation step of the forecasting procedure. See the overview paper (Carpenter et al., 2017) and the user manual (Stan Development Team, 2018) for more information.[20]

A.2 MCMC Sampler Settings

Table 8 shows the appropriate settings used by the MCMC sampler for the different models. The settings follow the recommendations of Clark and Ravazzolo (2015). Burn-in samples are draws at the start of a set of draws that are thrown away, to ensure that only converged samples are used to perform analysis on. Larger thinning intervals are appropriate for models such as the stochastic volatility specifications, in order to reduce the inevitable autocorrelation of the draws. Four sampler chains are run in parallel in order to make optimal use of computation power. It also provides the possibility of an extra check to ensure that the posteriors converge to the same distribution, regardless of different initial values or sampling paths. The total amount of draws leftover at the end of the procedure is the sample size S on which posterior analysis can be performed.

[20] The Stan user manual also functions as a starting point for the programming of several models used in this paper, such as the linear models and SV models.

Table 8: MCMC sampler settings per model

Model	Burn-in	Thin	Chains	Total leftover draws
Linear	2500	1	4	5000
Basic MS	2500	2	4	5000
Full MS	2500	2	4	5000
GARCH	2500	1	4	5000
SV (Stationary)	2500	8	4	5000
SV (RW)	2500	8	4	5000

Notes: For each model explained in section 3.2, this table lists the appropriate amount of burn-in samples, thinning intervals, number of parallel chains and total amount of draws leftover to use in the posterior analysis.

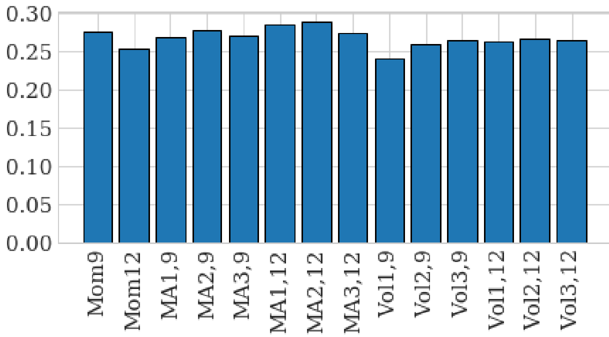
Appendix B Full Sample PLS Analysis

This section illustrates the contents of the factors generated by the PLS procedure. For illustration purposes, the results are derived by applying the PLS procedure to the full sample, applied to the 1-step ahead case. In the forecasting study, however, the procedure is repeated for each sample window and differs for each forecast horizon.

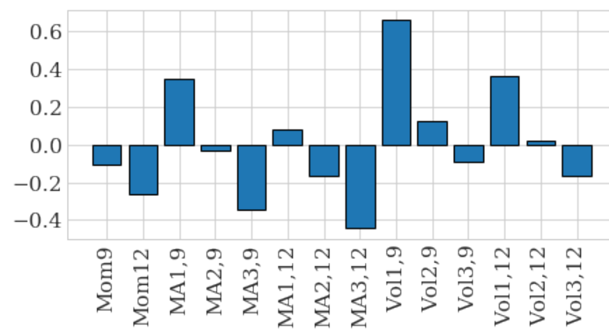
Figure 3 shows the loadings of the first two factors of the technical and sentiment categories, and the first two factors of the category combining the two. An important observation with regards to the combined category is that the first factor loads primarily on the technical indicators, whereas the second factor primarily loads on the sentiment indicators. This is important as Neely et al. (2014) note that there is reason to believe that technical indicators and sentiment indicators capture a similar underlying driver of the equity premium. As the PLS procedure largely separates the two categories into different factors, it actually indicates that the relation is not as apparent as expected. Furthermore, it provides evidence that the two categories are complementary, making it appropriate to treat them as separate categories in any further analysis shown in this thesis.

Figure 4 shows the loadings of the first three macroeconomic factors produced by the PLS procedure. There is not a very clear pattern regarding the interpretation of the factors, as they load on a wide-ranging set of variables. A similar observation can be made with regards to the content of the factors in figure 5. This figure shows the factor loadings when performing PLS on all available predictor variables. Although the different categories tend to stick load similarly on different factors, each factors still loads on a wide-ranging set of variables.

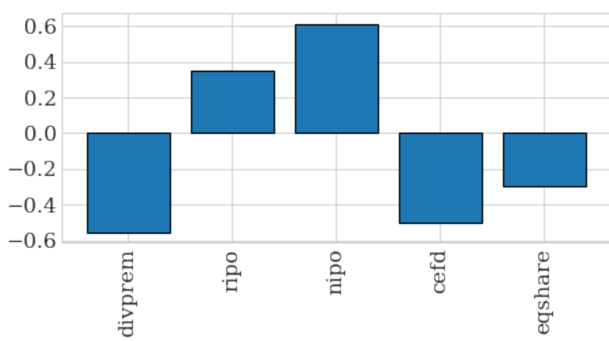
Figure 3: Technical and sentiment factor loadings



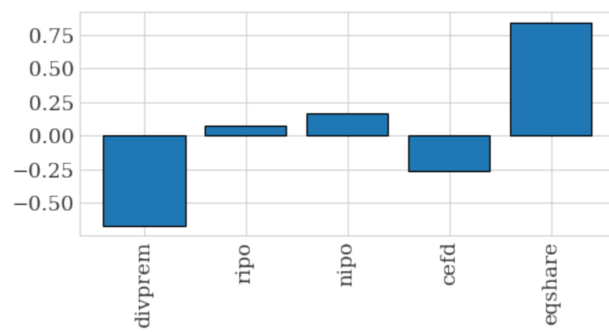
(a) Technical factor 1 loadings



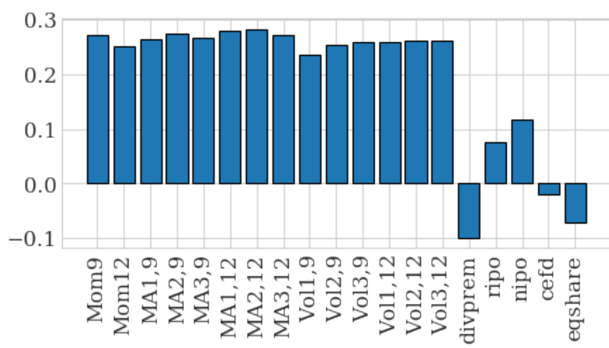
(b) Technical factor 2 loadings



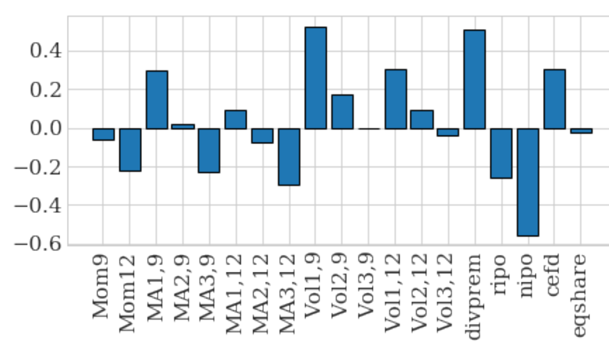
(c) Sentiment factor 1 loadings



(d) Sentiment factor 2 loadings



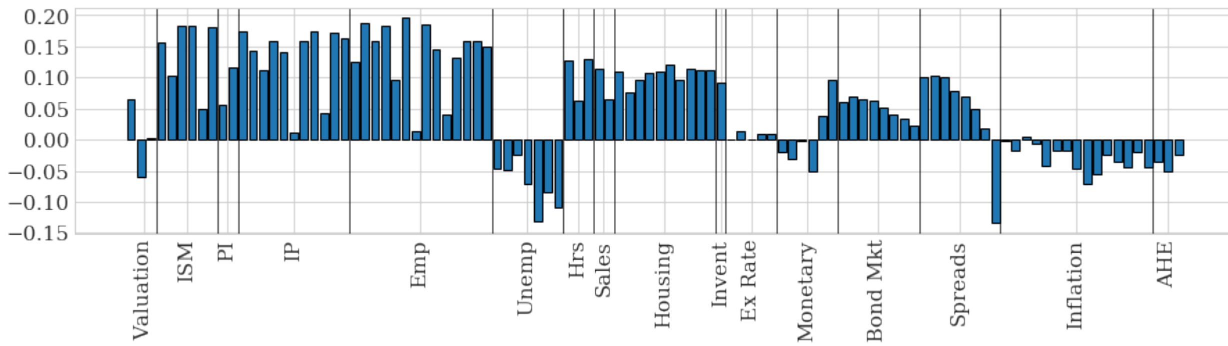
(e) Technical + sentiment factor 1 loadings



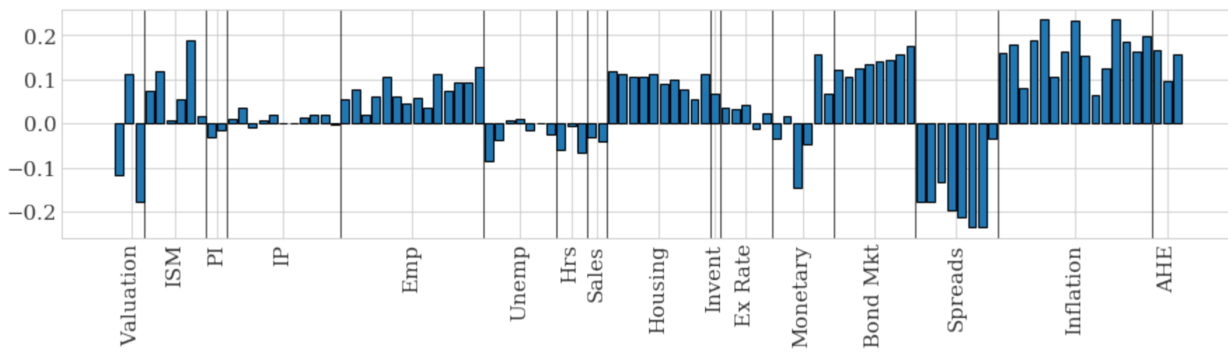
(f) Technical + sentiment factor 2 loadings

Note: The figures show loadings of the first factors of the technical and sentiment categories, and the first two factors of the combined category of technical and sentiment. The variables are explained in section 4. The variables of the sentiment variables represent value-weighted dividend premium, first-day returns on IPOs, IPO volume, closed-end fund discount, and the equity share in new issues.

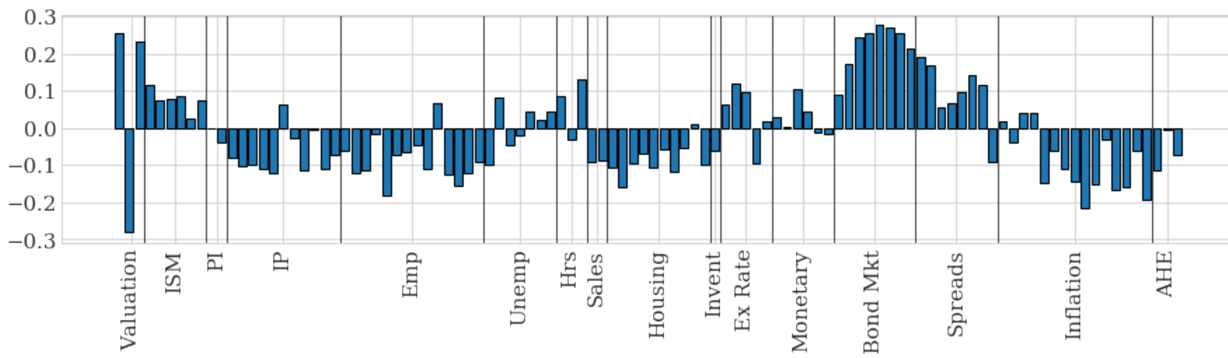
Figure 4: Macroeconomic factor loadings



(a) Factor 1



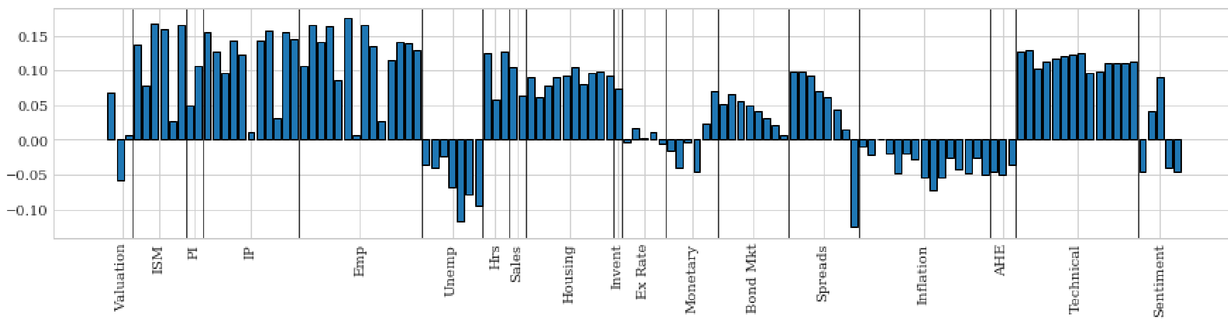
(b) Factor 2



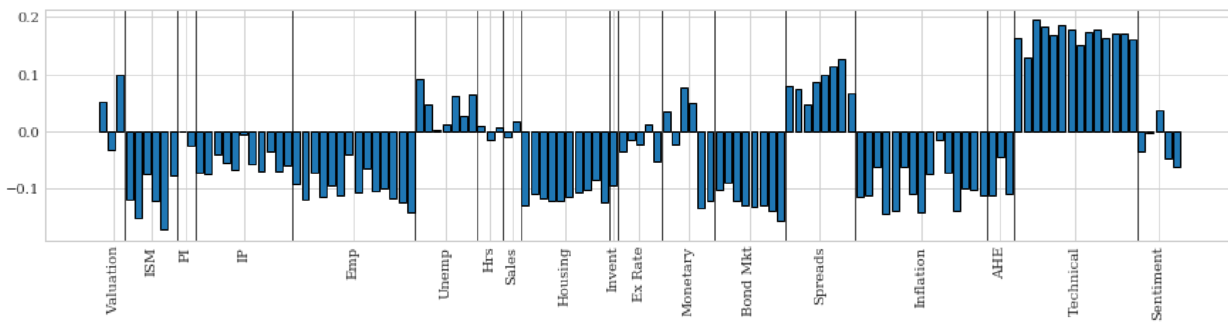
(c) Factor 3

Note: The figures show loadings of the first three macroeconomic factors, grouped by category. The categories are respectively: valuation ratios, ISM indices, personal income, industrial production, employment, unemployment, hours worked, sales and consumption, housing, inventory, exchange rates, monetary indicators, bond market yields, bond market spreads, inflation, and average hourly earnings. The order of the variables is equal to the order in table 9 of appendix. C.

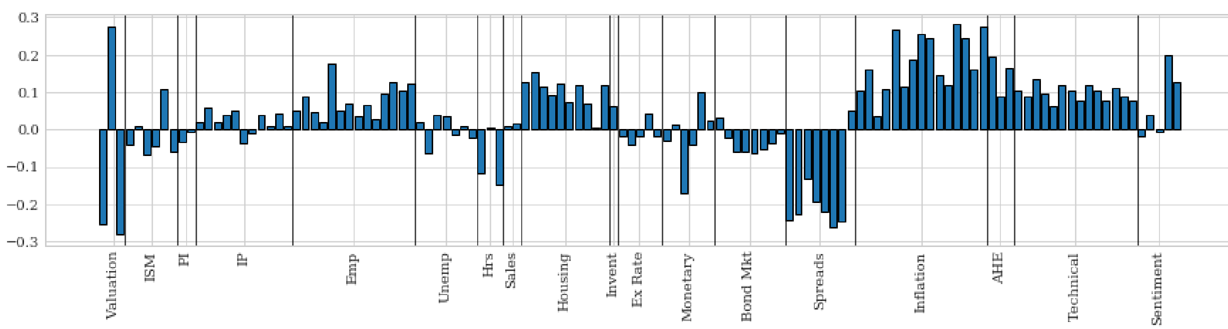
Figure 5: All categories factor loadings



(a) Factor 1



(b) Factor 2



(c) Factor 3

Note: The figures show loadings of the first three factors when performing PLS on all available predictor variables. The categories are respectively: valuation ratios, ISM indices, personal income, industrial production, employment, unemployment, hours worked, sales and consumption, housing, inventory, exchange rates, monetary indicators, bond market yields, bond market spreads, inflation, average hourly earnings, technical indicators, and sentiment indicators. The order of the variables is equal to the order in table 9 of appendix.

Appendix C Macroeconomic Data Set

This section provides a detailed overview of the large collection of macroeconomic variables. The source of all data is the FRED, unless indicated otherwise in the description (i.e. Shiller or Quandl).[21] To ensure stationarity, the series are transformed, according to the description in the online appendix of Çakmaklı and van Dijk (2016). Also in accordance with the process of Çakmaklı and van Dijk (2016), outliers of these transformed series are removed. The definition of an outlier are observations that deviate from the median value of the sample by more than six interquartile ranges. These outliers are replaced by the median value of the previous five observations. Wherever variables used by Çakmaklı and van Dijk (2016) were unavailable at publically available sources, either suitable replacements were selected, or the variable was left out. As some series were discontinued, they have been left out as well.

Table 9: Macroeconomic variables used to extract macroeconomic factors

Short Name	**Transf.**	**Description**
Div Yield	ln	Dividend Yield (Shiller)
PE	ln	PE Ratio (Shiller)
CAPE	ln	Cyclically-adjusted PE Ratio (Shiller)
ISM emp	lv	Manufacturing Employment Index (percent) (Quandl)
ISM inv	lv	Manufacturing Inventories Index (percent) (Quandl)
ISM new ordrs	lv	Manufacturing New Order Index (percent) (Quandl)
ISM pmi	lv	PMI Composite Index (percent) (Quandl)
ISM price	lv	Manufacturing Prices Index (percent) (Quandl)
ISM prodn	lv	Manufacturing Production Index (percent) (Quandl)
PI	Δln	Real Disposable Personal Income, Billions of Chained 2012 \$, SAAR
PI less transfers	Δln	Real Pers. Income excl. current transfers, Bill. Chained 2012 \$, SAAR
IP Total	Δln	Industrial Production Index, Index 2012=100, SA
IP bus eqpt	Δln	Industrial Production: Business Equipment, Index 2012=100, SA
IP cons gds	Δln	Industrial Production: Consumer Goods, Index 2012=100, SA
IP constr sup	Δln	Industrial Production: Construction supplies, Index 2012=100, SA
IP final prod	Δln	Industrial Production: Final Products (Market Group), Index 2012=100, SA
IP elct gas	Δln	Industrial Production: Electric and Gas Utilities, Index 2012=100, SA
IP matls	Δln	Industrial Production: Materials, Index 2012=100, SA
IP mfg	Δln	Industrial Production: Manufacturing (SIC), Index 2012=100, SA
IP min	Δln	Industrial Production: Mining, Index 2012=100, Monthly, SA
IP nind sup	Δln	Industrial Production: Nonindustrial supplies, Index 2012=100, SA
Cap util	Δlv	Capacity Utilization: Total Industry, Percent of Capacity, SA

continued on next page

[21] The Shiller dataset refers to the set on the website of Robert Shiller. Refer to (Shiller, 2015) for details.

continued from previous page

Short Name	Transf.	Description
Emp CPS total	Δln	Civilian Employment Level, Thousands of Persons, SA
Emp total	Δln	All Employees: Total Nonfarm Payrolls, Thousands of Persons, SA
Emp const	Δln	All Employees: Construction, Thousands of Persons, SA
Emp dble gds	Δln	All Employees: Durable Goods, Thousands of Persons, SA
Emp fire	Δln	All Employees: Financial Activities, Thousands of Persons, SA
Emp gds prod	Δln	All Employees: Goods-Producing Industries, Thousands of Persons, SA
Emp govt	Δln	All Employees: Government, Thousands of Persons, SA
Emp mfg	Δln	All Employees: Manufacturing, Thousands of Persons, SA
Emp ndble gds	Δln	All Employees: Nondurable goods, Thousands of Persons, SA
Emp mining log	Δln	All Employees: Mining and logging, Thousands of Persons, SA
Emp retail	Δln	All Employees: Retail Trade, Thousands of Persons, SA
Emp service	Δln	All Employees: Private Service-Providing, Thousands of Persons, SA
Emp TTU	Δln	All Employees: Trade, Transportation and Utilities, Thous. of Persons, SA
Emp Wholesale	Δln	All Employees: Wholesale Trade, Thous. of Persons, SA
U all	lv	Civilian Unemployment Rate, Percent, SA
U mean duration	Δlv	Average (Mean) Duration of Unemployment, Weeks, SA
U u5 weeks	Δln	Civilians Unemployed for Less Than 5 Weeks, Thousands of Persons, SA
U 5-14 weeks	Δln	Civilians Unemployed for 5 to 14 Weeks, Thousands of Persons, SA
U 15+ weeks	Δln	Civilians Unemployed for 15 Weeks and Over, Thousands of Persons, SA
U 15-26 weeks	Δln	Civilians Unemployed for 15 to 26 Weeks, Thousands of Persons, SA
U 27+ weeks	Δln	Civilians Unemployed for 27 Weeks and Over, Thousands of Persons, SA
Avg hrs	lv	Avg Weekly Hrs of Employees: Goods-Producing, Hours, SA
Overtime mfg	Δlv	Avg Weekly Overtime Hrs of Employees: Manufacturing, Hours, SA
Avg hrs mfg	lv	Avg Weekly Hrs of Employees: Manufacturing, Hours, SA
M&T sales	Δln	Real Manuf. and Trade Industries Sales, Mill. of Chained 2012 $, SA
Consumption	Δln	Real pers. consumption exp. (Chained index), Index 2012=100, SA
Hstarts total	ln	Housing Starts: Total: New Houses Started, Thous., SAAR
Hstarts NE	ln	Housing Starts in Northeast Census Region, Thous., SAAR
Hstarts MW	ln	Housing Starts in Midwest Census Region, Thous., SAAR
Hstarts South	ln	Housing Starts in South Census Region, Thous., SAAR
Hstarts West	ln	Housing Starts in West Census Region, Thous., SAAR
BP total	ln	Houses Auth. by Building Permits, Thous., SAAR
BP NW	ln	Houses Auth. by Building Permits in the Northeast, Thous., SAAR
BP MW	ln	Houses Auth. by Building Permits in the Midwest, Thous., SAAR
BP South	ln	Houses Auth. by Building Permits in the South, Thous., SAAR
BP West	ln	Houses Auth. by Building Permits in the West, Thous., SAAR
M&T Invent	Δln	Real Manufacturing and Trade Inventories, Chained 2012 Dollars, SA
Ex rate broad	Δln	Trade Weighted $ Index: Broad, Goods, Index Jan 1997=100, Not SA
Ex rate Switz	Δln	Switzerland / U.S. FX Rate, Swiss Francs to One U.S. Dollar, Not SA
Ex rate Japan	Δln	Japan / U.S. FX Rate, Japanese Yen to One U.S. Dollar, Not SA
Ex rate UK	Δln	U.S. / U.K. FX Rate, U.S. Dollars to One British Pound, Not SA
Ex rate Canada	Δln	Canada / U.S. FX Rate, Canadian Dollars to One U.S. Dollar, Not SA

continued on next page

continued from previous page

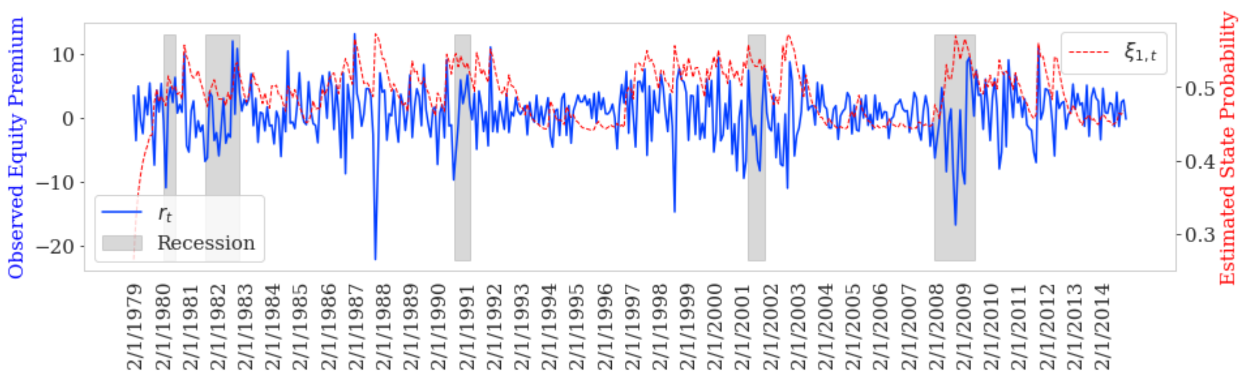
Short Name	**Transf.**	**Description**
M1	Δln	M1 Money Stock, Billions of Dollars, SA
M2	Δln	M2 Money Stock, Billions of Dollars, SA
M2 (real)	Δln	Real M2 Money Stock, Billions of 1982-84 Dollars, SA
MB	Δln	St. Louis Adjusted Monetary Base, Billions of Dollars, SA
C&I loans	Δln	Commercial and Industrial Loans, All Commercial Banks, Bill. of $, SA
Cons Credit	Δln	Total Nonrevolving Credit Owned and Securitized, Outst., Bill. of $, SA
Fedfunds fyff	Δlv	Effective Federal Funds Rate, Percent, Not SA
3 mo T-bill fygm3	Δlv	3-Month Treasury Bill: Secondary Market Rate, Percent, Not SA
6 mo T-bill fygm6	Δlv	6-Month Treasury Bill: Secondary Market Rate, Percent, Not SA
1 yr T-bond fygt1	Δlv	1-Year Treasury Constant Maturity Rate, Percent, Not SA
5 yr T-bond fygt5	Δlv	5-Year Treasury Constant Maturity Rate, Percent, Not SA
10 yr T-bond fygt10	Δlv	10-Year Treasury Constant Maturity Rate, Percent, Not SA
Aaa bond fyaaac	Δlv	Moody's Seasoned Aaa Corporate Bond Yield, Percent, Not SA
Baa bond fybaac	Δlv	Moody's Seasoned Baa Corporate Bond Yield, Percent, Not SA
3 mo-FF spread	lv	fygm3-fyff
6 mo-FF spread	lv	fygm6-fyff
1 yr-FF spread	lv	fygt1-fyff
5 yr-FF spread	lv	fygt5-fyff
10 yr-FF spread	lv	fygt10-fyff
Aaa-FF spread	lv	fyaaac-fyff
Baa-FF spread	lv	fyBaac-fyff
Baa-Aaa spread	lv	fyBaac-fyaaac
PPI comm	Δln	Producer Price Index (PPI) for All Commodities, Index 1982=100, Not SA
PPI fin gds	Δln	PPI: Finished Goods, Index 1982=100, SA
PPI cons gds	Δln	PPI: Finished Consumer Foods, Index 1982=100, SA
PPI int gds	Δln	PPI: Processed Goods for Interm. Demand, Index 1982=100, SA
CPIU all	Δln	Consumer Price Index (CPI): All Items, Index 1982-1984=100, SA
CPIU apparel	Δln	CPI: Apparel, Index 1982-1984=100, SA
CPIU food	Δln	CPI: Food and Beverages, Index 1982-1984=100, SA
CPIU house	Δln	CPI: Housing, Index 1982-1984=100, SA
CPIU med	Δln	CPI: Medical Care, Index 1982-1984=100, SA
CPIU other	Δln	CPI: Other Goods and Services, Index 1982-1984=100, SA
CPIU transp	Δln	CPI: Transportation, Index 1982-1984=100, SA
PCE	Δln	Personal Consumption Exp. (PCE): Chained Index, Index 2012=100, SA
PCE dlbes	Δln	PCE: Durable goods (Chained Index), Index 2012=100, SA
PCE ndble	Δln	PCE: Nondurable goods (Chained Index), Index 2012=100, SA
PCE services	Δln	PCE: Services (Chained Index), Index 2012=100, SA
AHE goods	Δln	Average Hourly Earnings: Goods-Producing, $ per Hour, SA
AHE const	Δln	Average Hourly Earnings: Construction, $ per Hour, SA
AHE mfg	Δln	Average Hourly Earnings: Manufacturing, $ per Hour, SA

Notes: Each row indicates an individual macroeconomic data series. The column 'Short Name' denotes a term for notational convenience, the 'Transf.' column indicates the appropriate transformation, and the 'Description' column lists the description as provided by the data source. As for the transformations: 'lv' indicates no transformation, 'ln' indicates a logarithmic transformation, 'Δlv' indicates taking first differences of the level, and 'Δln' indicates taking first differences of the logarithmic transformation.

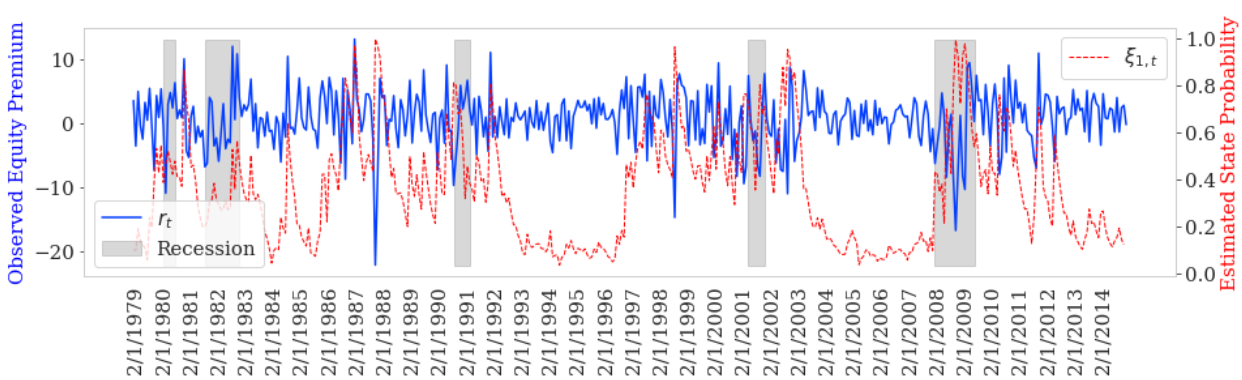
Appendix D Full Sample Analysis of Advanced Models

The sampling process of the models explained in section 3.2 produce in-sample estimates of the underlying state probabilities, or the volatility of the equity premium innovation process. This section provides an illustration of what these estimates look like in practice, in order to provide a better intuition on the way the models function. The estimates used for illustration are applied to the full sample of observations, except for a training sample of 5 years used to calibrate the prior distributions. Each illustration is a 1-step ahead case, using two predictive PLS factors. The PLS procedure derives the factors from the combined data set of macroeconomic, technical and sentiment variables. Note that as the actual forecasting study is applied on expanding windows rather than the full sample, the estimates shown here differ at least slightly from the estimates used in the forecasting procedure.

Figure 6: Volatility estimates of Markov switching models



(a) Basic Markov switching model



(b) Full Markov switching model

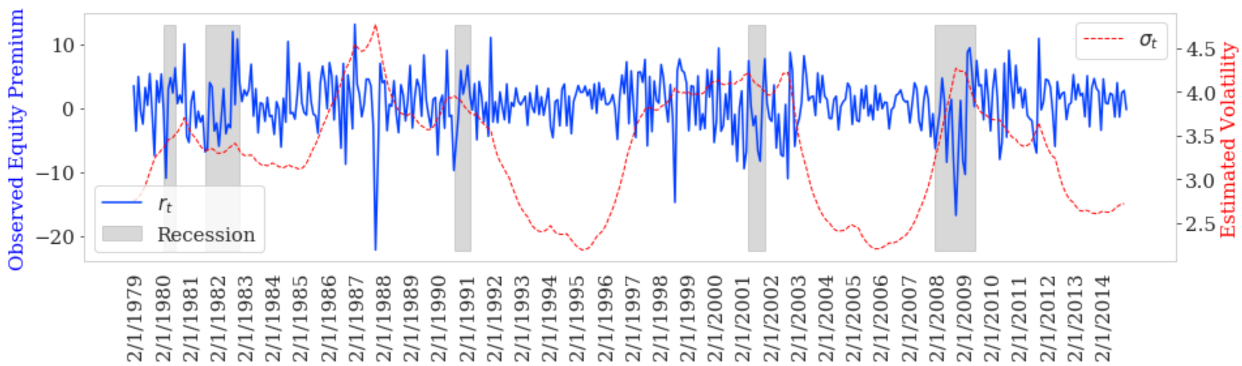
Note: The graphs plot the evolution of the risk premium denoted as r_t and the in-sample estimates of the probability of being in state 1, denoted as $\zeta_{1,t}$ over time. Section 3.2 introduces the different models.

Figure 6 displays both the observed equity premium value and the posterior mean of the probability of being in state 1, over time. The interpretation of state 1 becomes quite clear from observing the graph. The models tend to attribute a high probability to being in state 1 during times of recession, which also corresponds to negative excess returns and higher volatility. However, the estimates cannot be straightforwardly interpreted as a recession indicator, as this would result in many *false positives*, where the model attributes high probabilities to state one when there is in fact no sign of recession. The pattern of the Markov switching volatility model and the full Markov switching model (where the coefficient on the predictors are also time-varying) tends to not differ too much, as seen from the graph.

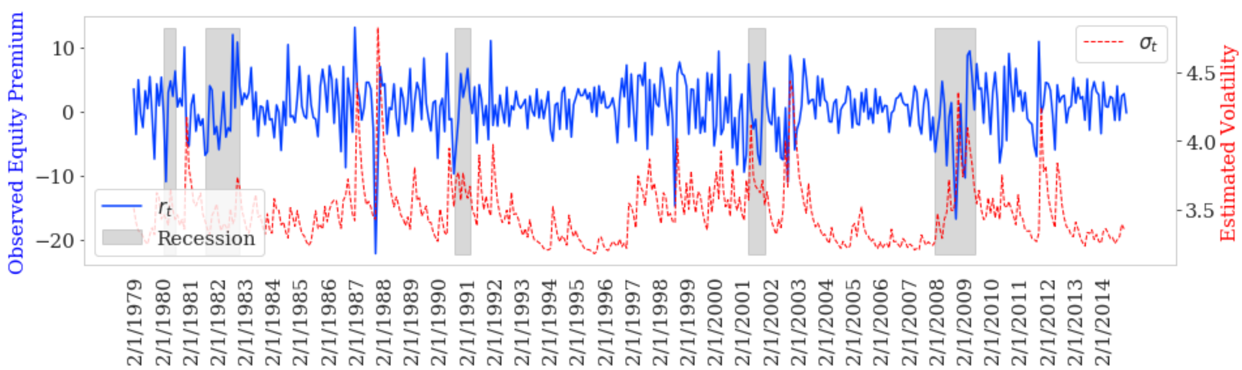
Figure 7 displays both the observed equity premium value and the posterior mean of the volatility estimates produced by the respective models. The results of the volatility estimates look like what would be expected. In prolonged periods where the the equity premium showcases signs of higher volatility, the estimates of the volatility of the innovation process also tend to increase. This tends to occur during recessions, but also during other well-known periods of market volatility such as during the market crash in the late '80s and the period preceding the crash of the dot-com bubble. Between the three different models, the volatility estimates of the random walk SV model looks the most smooth, whereas the GARCH model yields more wide-ranging values.

Figure 7: Volatility estimates of time-varying volatility models

(a) Stationary stochastic volatility model



(b) Random walk stochastic volatility model



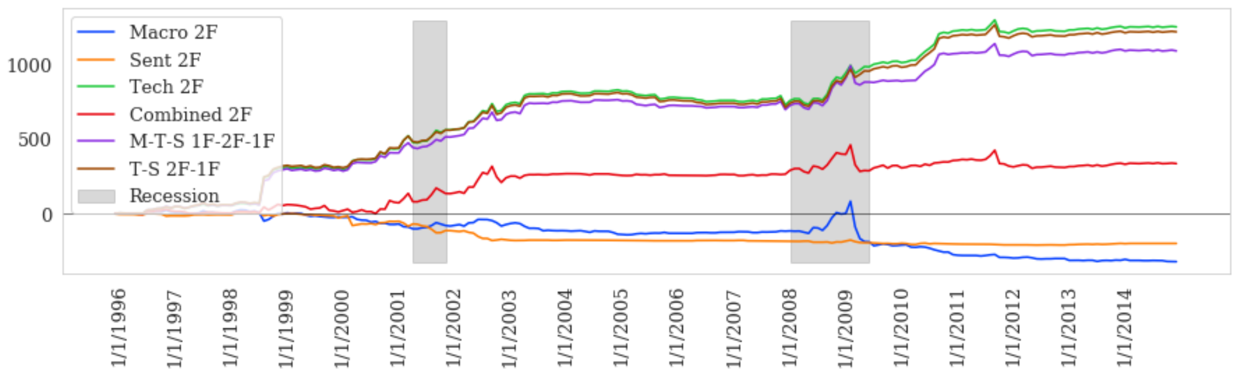
(c) GARCH model

Note: The graphs plot the evolution of the risk premium denoted as r_t and the in-sample estimates of the innovation process σ_t over time. Section 3.2 introduces the different models

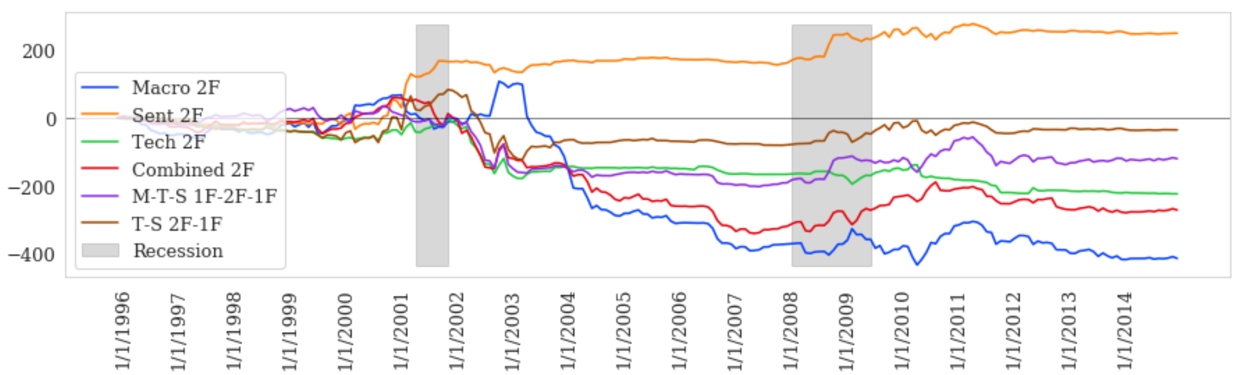
Appendix E Additional Results

E.1 Additional Results Graphs

Figure 8: CSSED over time for linear prediction models



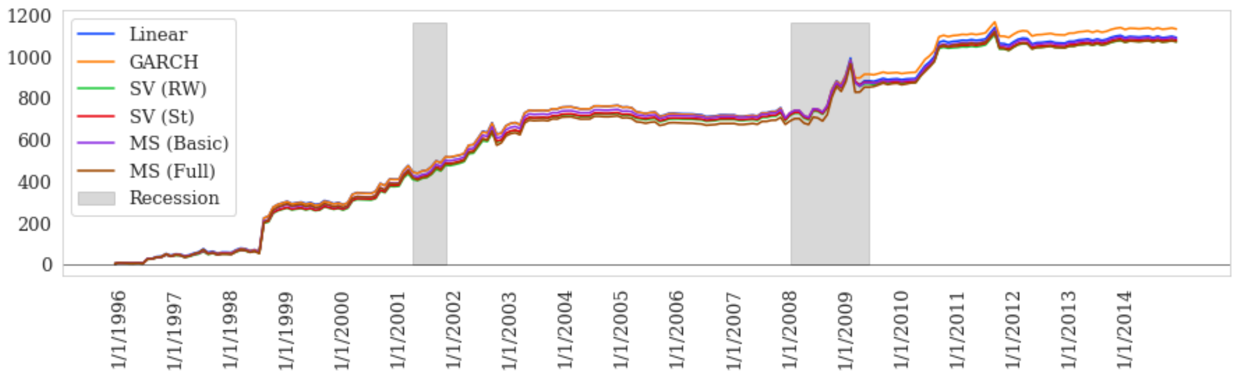
(a) Forecast horizon = 1 month



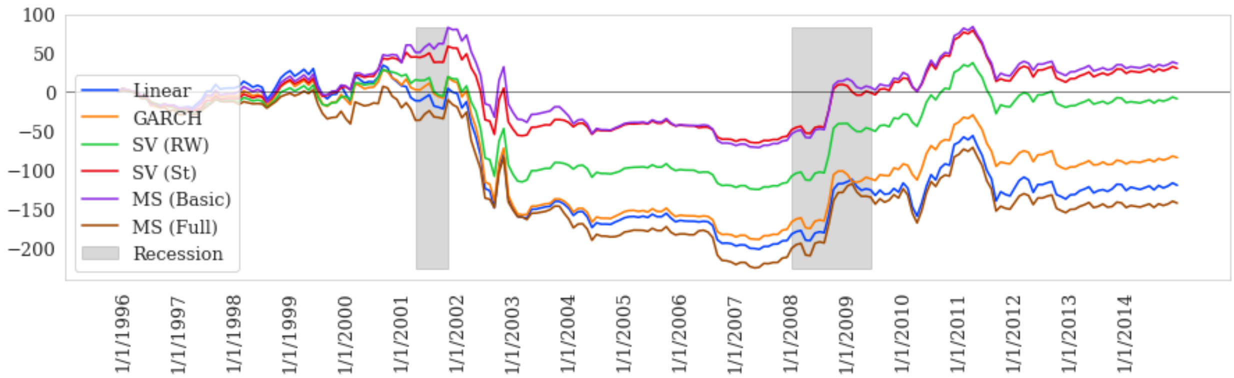
(b) Forecast horizon = 12 months

Note: The figures represent the evolution of the CSSED values of a selection of the linear prediction models over time. The naming convention is similar to tables 1 to 6. However, *Macro*, *Tech*, and *Sent* are abbreviated as *M*, *T*, and *S*, respectively, in some of the legend entries. Gray areas represent time periods denoted as recessions by NBER. The horizontal line at 0 represents the performance of the historical average benchmark model. Hence, upward sloping lines represent models that are outperforming the benchmark at the respective period, and vice versa. A value above zero means that a respective model has outperformed the benchmark up to the respective time period, and vice versa. See equation 17 for details on the evaluation criterion.

Figure 9: CSSED over time for advanced prediction models



(a) Forecast horizon = 1 month



(b) Forecast horizon = 12 months

Note: The figures represent the evolution of the CSSED values of a selection of the advanced prediction models over time. The factors used as predictors consist of one macroeconomic factor, two technical factors, and one sentiment factor (hence, the Macro+Tech+Sent 1F-2F-1F model referred to earlier). The naming convention is the same as maintained in tables 1 to 6. *RW* refers to the random walk variant of SV models, whilst *St* refers to the stationary variant. Gray areas represent time periods denoted as recessions by NBER. The horizontal line at 0 represents the performance of the historical average benchmark model. Hence, upward sloping lines represent models that are outperforming the benchmark at the respective period, and vice versa. A value above zero means that a respective model has outperformed the benchmark up to the respective time period, and vice versa. See equation 17 for details on the evaluation criterion.

E.2 Market Timing Study With No Margin Constraint

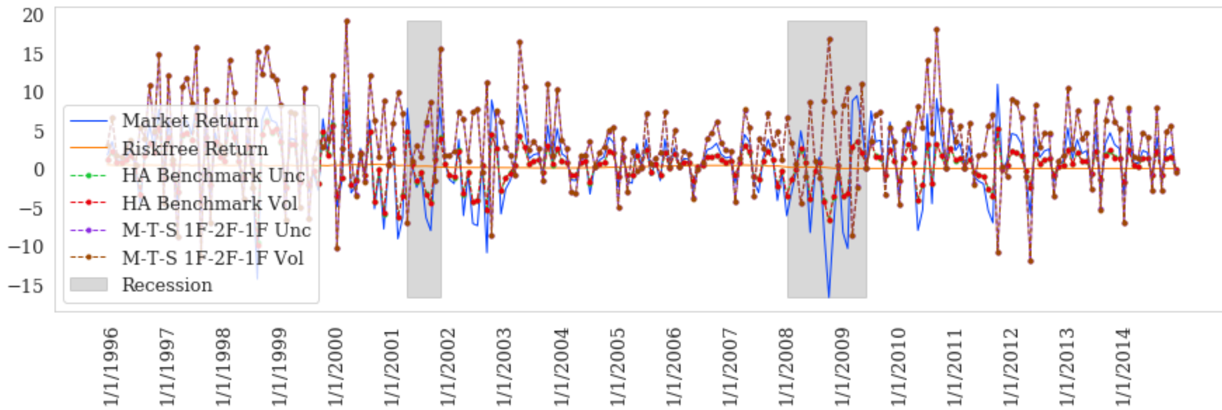
Table 10: Results market timing study with a no margin constraint

$\gamma = 6$	Weights Determined Using Volatility Forecast						Weights Determined Using Estimation Uncertainty					
1995:12-2014:12	Sharpe Ratio			CER			Sharpe Ratio			CER		
	Full	Exp	Rec	Full	Exp	Rec	Full	Exp	Rec	Full	Exp	Rec
Constant Weights												
100%	0.465	0.754	-0.811	4.10%	7.31%	-17.54%						
50%	0.464	0.753	-0.812	4.36%	6.54%	-10.97%						
0%	-	-	-	2.50%	2.59%	1.81%						
Benchmark												
Historical Average	0.459	0.679	-0.916	4.26%	6.36%	-10.48%	0.459	0.679	-0.910	4.27%	6.36%	-10.40%
Linear Models												
Macroeconomic												
1F	0.184	0.292	-0.547	0.95%	2.30%	-9.02%	0.184	0.291	-0.538	0.97%	2.30%	-8.84%
2F	0.253	0.322	-1.254	2.20%	2.82%	-2.41%	0.253	0.321	-1.248	2.21%	2.81%	-2.32%
3F	0.219	0.299	-0.917	2.20%	2.91%	-3.07%	0.217	0.298	-0.914	2.19%	2.90%	-3.10%
Technical												
1F	1.299	1.392	0.679	12.58%	14.09%	1.81%	1.299	1.392	0.679	12.57%	14.09%	1.81%
2F	1.750	1.813	1.210	17.87%	18.86%	10.47%	1.750	1.813	1.210	17.87%	18.85%	10.47%
3F	1.771	1.838	1.204	18.14%	19.18%	10.41%	1.771	1.838	1.203	18.14%	19.18%	10.41%
Sentiment												
1F	0.322	0.549	-0.846	2.42%	5.15%	-16.29%	0.320	0.541	-0.833	2.44%	5.08%	-15.75%
2F	0.228	0.424	-0.780	1.69%	3.96%	-14.20%	0.222	0.416	-0.772	1.65%	3.89%	-14.09%
3F	0.242	0.445	-0.821	1.83%	4.16%	-14.43%	0.238	0.441	-0.813	1.83%	4.13%	-14.27%
Combined												
1F	1.079	1.147	0.664	10.50%	11.65%	2.26%	1.079	1.147	0.665	10.50%	11.64%	2.26%
2F	1.153	1.265	-0.032	10.60%	11.96%	0.75%	1.158	1.265	0.001	10.63%	11.96%	0.99%
3F	1.154	1.216	0.651	10.10%	11.08%	2.96%	1.152	1.215	0.651	10.08%	11.05%	2.93%
4F	1.119	1.175	0.633	10.43%	11.14%	5.07%	1.122	1.177	0.644	10.45%	11.15%	5.15%
Tech+Sent												
1F+1F	1.245	1.333	0.679	11.97%	13.39%	1.81%	1.246	1.334	0.679	11.97%	13.39%	1.81%
2F+1F	1.768	1.834	1.210	18.01%	19.02%	10.47%	1.768	1.834	1.210	18.01%	19.02%	10.47%
Macro+Tech+Sent												
1F+1F+1F	1.045	1.116	-	10.07%	11.22%	1.81%	1.043	1.114	-	10.04%	11.18%	1.81%
2F+1F+1F	1.038	1.109	0.679	10.06%	11.20%	1.81%	1.037	1.108	0.679	10.03%	11.17%	1.81%
1F+2F+1F	1.723	1.808	0.967	17.18%	18.61%	6.81%	1.722	1.807	0.967	17.15%	18.58%	6.81%
Advanced Models (Macro+Tech+Sent 1F+2F+1F)												
GARCH	1.722	1.798	1.061	17.07%	18.40%	7.45%	1.722	1.798	1.057	17.07%	18.40%	7.43%
SV (Random Walk)	1.683	1.770	0.894	16.35%	17.79%	5.96%	1.681	1.768	0.893	16.32%	17.76%	5.95%
SV (Stationary)	1.697	1.781	0.948	16.78%	18.18%	6.69%	1.697	1.781	0.948	16.78%	18.17%	6.69%
MS (Basic)	1.718	1.804	0.960	17.13%	18.56%	6.76%	1.733	1.820	0.962	17.34%	18.81%	6.78%
MS (Full)	1.756	1.821	1.200	17.65%	18.86%	8.82%	1.763	1.830	1.198	17.79%	19.03%	8.76%

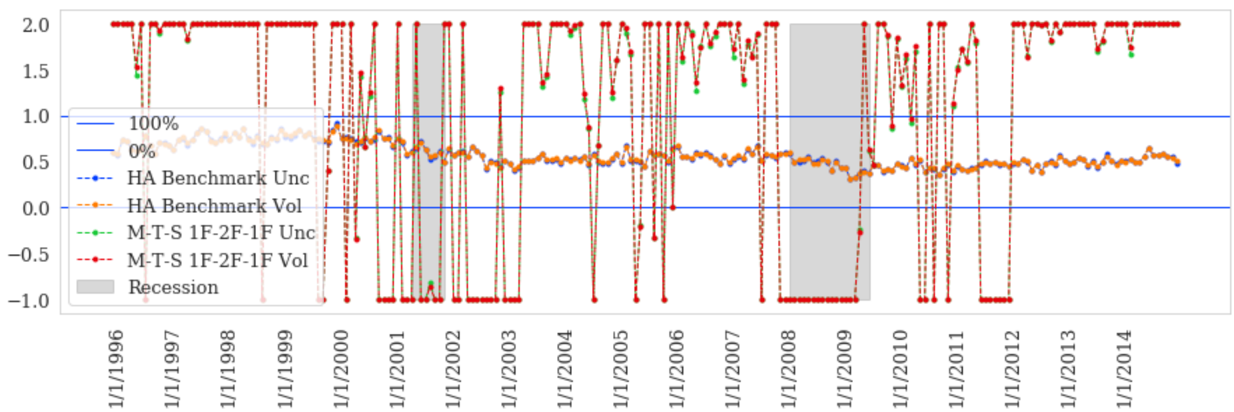
Notes: This table shows the results of a hypothetical investment strategy where an investor allocates a proportion of wealth in the market and another in a risk-free product. As a comparison, the table also shows the results of investing using constant weights, where the investor rebalances each month in order to invest the respective proportion in the market and the remaining wealth in the risk-free product. The evaluated investment period is 1995:12 to 2014:12, where the investor reallocates the investments monthly by optimizing the weights according to equation 28 with $\gamma = 6$. The weight invested in the market is constrained between 0 and 1, not allowing short-selling or leverage. 0.1% transaction costs over the respective turnover is subtracted from each portfolio's monthly returns. The column on the left represent the predictors and model specification (see section 3.2 for more details on the models). Each advanced model uses the *Macro+Tech+Sent 1F+2F+1F* specification of the previous section, meaning that for the categories of macroeconomic, technical, and sentiment variables the PLS procedure generates one, two, and one factor, respectively. The *overall* column represents the entire period, whereas the *expansion* and *recession* columns represent the sub-samples of economic expansion or recession, respectively, as measured by NBER. See equation 29 for the evaluation criteria.

Appendix F Illustration of Market Timing Study

Figure 10: Illustration of market timing study



(a) Portfolio returns over time



(b) Proportion invested in the market over time

Note: This figure shows both the returns of the trading strategy and the proportion of wealth invested in the the market when using two of the respective forecast models to construct a strategy according to the methodology in section 6.1. The *Unc* term denotes that the weights are generated using the uncertainty of the equity premium forecast, whereas *Vol* denotes that that they are generated using the volatility forecasts. *HA* is the historical benchmark model, whereas *M-T-S 1F-2F-1F* is the model that uses a macroeconomic factor, two technical factors, and a sentiment factor as its predictors.