Risk Aversion and the Forecasting Performance of Implied Expected Returns in Bull and Bear Markets

Author:
Sietse Romijn - 400950

Master thesis in Econometrics and Management Science - Quantatitive Finance
Erasmus School of Economics, Erasmus University Rotterdam

Supervisor: Karolina Scholtus
Second Assessor: Dick van Dijk

Abstract

In order to obtain more accurate return forecasts, this paper proposes several methods to improve the almost always positive implied return forecasts. Firstly, we estimate a constant and time-varying volatility risk premium and link these to the risk aversion, which is an input for implied return forecasts. Secondly, because implied return forecasts are almost always positive, forecasting accuracy is low during bear markets. To remedy this problem, we implement a dynamic forecast combination approach based on the output of Markov switching models, or on the level of volatility. We find that a constant volatility risk premium only marginally affects forecasting performance. On the other hand, a time-varying volatility risk premium reduces the root-mean-squared forecast error (RMSFE) from 53.66‰ to 53.08‰. The forecast combination approach based on a Markov switching model leads to further considerable improvement and achieves the RMSFE of 51.21‰. This improvement mainly comes from more accurate forecasts in bear markets where the RMSFE decreases from 72.07‰ to 62.30‰.

Keywords: implied return, volatility risk premium, risk aversion, forecast combination

Contents

1 Introduction

Ever since the introduction of modern portfolio theory by Markowitz [1952], investors and academics have been concerned with finding portfolios that lay on the efficient frontier. These mean-variance efficient portfolios are also relevant when reverse engineering expected returns, as in Sharpe [1974]. Reverse engineering of expected returns means that instead of computing optimal weights from expected returns and a covariance estimate, we extract implied expected returns from the covariance and a portfolio that is assumed to be mean-variance efficient. In an investable universe, with known expected return vector μ, covariance matrix Σ and a mean-variance efficient portfolio w, there exists a linear relation between the asset's expected return μ and their covariance with a mean-variance efficient portfolio, Σw. This relation is used to compute the implied expected returns, which are given by: $\mu = \ell\iota + \gamma\Sigma w$, where ℓ is the risk-free rate and γ is the risk aversion parameter. A complication in determining the implied expected returns is that the composition of the mean-variance efficient portfolio is unobservable in practice. We therefore use various portfolios that proxy a mean-variance efficient portfolio. In this paper we aim to improve the implied expected return forecasts, by linking the risk aversion parameter to the volatility risk premium. Moreover, we apply a dynamic forecast combination approach with weights dependent on the state of the equity market.

Under the capital asset pricing model (CAPM) formulated by Sharpe [1964], the market capitalization weighted portfolio is mean-variance efficient. Hence it has been the standard choice as mean-variance efficient portfolio proxy in the literature. Ardia and Boudt [2015] investigate this choice in the context of implied expected returns. They examine the forecasting accuracy of implied expected returns that are based on other mean-variance efficient portfolio proxies. They conclude that risk-based portfolio proxies perform marginally better than the market capitalization weighted portfolio, when computing implied expected returns. These risk-based portfolios include the inverse volatility weighted-, the equal risk contribution- and maximum diversification portfolio. Furthermore, they conclude that implied expected return forecasts are more accurate and more stable than time-series based forecast models. The latter includes models such as the constant mean and the Fama and French [1993] three factor model.

A downside of basic implied expected return forecasts is that they are almost always positive. Between 1970 and 2016, one out of 3240 forecasts is negative[1]. Therefore, forecasting accuracy during bear markets is likely low, because bear markets are characterized as periods with negative or low returns. Moreover, bear markets usually have higher volatility, which leads to even higher implied return forecasts. Although implied return forecasts are almost always positive, we can obtain negative return forecasts by making a forecast combination of them. This holds, because forecast combination weights do not have to sum to one, as recommended by Granger and Ramanathan [1984]. A forecast combination of only implied expected returns can therefore solve the problem of almost positive implied expected returns in bear markets. On the other hand, almost always positive return forecasts are not a problem in bull markets. Applying the same forecast combination weights in bull and bear markets, thus likely leads to a drop in forecasting performance in bull markets. We therefore implement a forecast combination approach of implied expected return forecasts with dynamic weights, dependent on the regime of the equity market, similar to Deutsch et al. [1994]. Because bull and bear markets have different forecast combination weights, the forecasting performance in bull markets is not necessarily negatively impacted. Since implied return forecasts are likely biased in bear markets, gains from implementing this forecast combination approach can come from reducing forecast bias in bear markets and from diversification.

[1]This one negative forecasts is in the high-tech industry at the height of the 2008 financial crisis, for the maximum diversification portfolio proxy.

To determine whether forecast combinations of only implied expected returns are enough to improve the forecasting performance in bear markets, we examine the impact of including time-series based forecasts in the forecast combination.

Another purpose of this thesis is to further improve the forecasting accuracy of implied expected return forecasts, by more accurately capturing the (time-varying) risk aversion parameter. Ardia and Boudt [2015] and Romijn [2016] choose a constant risk aversion parameter $\gamma = 2.4$ and also present results for $\gamma = 1$ and $\gamma = 5$ in their sensitivity studies. However, these values are chosen quite arbitrarily. Hence we incorporate the approach of Bollerslev et al. [2011] in which they estimate the risk aversion parameter from implied and realized volatility. Bollerslev et al. [2011] use GMM estimation of the cross-conditional moments between risk-neutral and objective expectations of realized volatility to identify the stochastic volatility risk premium. Subsequently, they switch the sign of the volatility risk premium to obtain investor risk aversion. Our approach is different as we choose a constant, to multiply the volatility risk premium with, that minimizes the forecast error of the corresponding implied expected returns over past observations. We deviate from the approach of Bollerslev et al. [2011], because we are concerned with the forecasting performance of the implied expected returns that result from the risk aversion estimate, while they are not.

Guiso et al. [2013] document an increase in investor risk aversion from before the financial crisis of 2008 to after the crisis. This supports the idea that investor risk aversion is time-varying and higher after economic downturn. We therefore also estimate a time-varying volatility risk premium, as in Bollerslev et al. [2011]. Their approach introduces time-variation in the volatility risk premium by linking it to macro-economic state variables. It is possible in this approach for risk aversion to be negative. This implies that the average investor is risk seeking, which means that the utility function of the average investor is not globally concave. Hartley and Farrell [2002] find evidence of risk seeking behavior and local convexity of the utility function. We investigate whether the time-varying risk aversion estimate better captures investor risk aversion than the constant alternative. More importantly, we examine whether it improves implied return forecasts. A consequence of a negative risk aversion parameter is that implied return forecasts can be negative. We analyze if this remedies the problem of positive implied return forecasts.

Several alternative methods for estimating a time-varying risk aversion parameter have been implemented in the literature. One of these approaches simply sets the risk aversion parameter to that month's difference between the implied and realized volatility. The downside of this approach is that it attributes every change in the data to a change in risk aversion. This leads to a very volatile time-series of risk aversion. Conversely, approaches from the consumption-based asset pricing models (Campbell and Cochrane [1999]) lead to remarkably smooth graphs of risk aversion and show too little variation. The approach of Bollerslev et al. [2011] leads to a series of risk aversion that avoids excessive short-term random variation, but is still linked to economic circumstances, therefore allowing for easier economic interpretation.

We apply both extensions to implied expected returns and evaluate the out-of-sample performance of the resulting forecasts of ten US industry portfolios, over the period June 1998 to December 2016. We compare these forecasts to implied expected returns with a risk aversion parameter of $\gamma = 2.4$ and no forecast combinations. We analyze forecasting performance of implied expected returns conditional and unconditional on the state of the equity market. Showing both results is important because a big improvement in forecasting accuracy in bear markets does not always lead to an improvement over the whole sample. Our main focus is on the difference in forecast performance between bull and bear markets, since our extensions aim to solve the problem of poor forecasting performance in bear markets.

We find that a constant volatility risk premium only marginally affects the forecasting accuracy compared to the benchmark. On the other hand, a time-varying volatility risk premium significantly(10%) improves the forecasting accuracy for four out of the five mean-variance efficient portfolio proxies. It reduces the RMSFE from an average of 53.66‰ in the benchmark to 53.08‰. This is a consequence of the improvement of the RMSFE in bear markets from an average of 72.71‰ to 70.99‰. A time-varying volatility risk premium therefore improves the forecasting accuracy of implied expected returns in bear markets. However, the difference between the forecasting performance in bull and bear markets is still substantial.

The biggest gain in forecasting accuracy is made by introducing forecast combinations with Markov switching models to implied return forecasts. A forecast combination of only implied return forecasts significantly(5%) reduces the RMSFE from 53.66‰ to 51.21‰. We conclude that for such forecasts combinations, the weights sum to less than unity in bear markets while summing to values higher than unity in bull markets. This results in implied return forecast combinations being able to forecast negative returns. Consequently, they improve the hit ratio from 59.42% to 63.54%. Furthermore, these forecasts significantly(1%) reduce the RMSFE in bear markets from 72.07‰ to 62.30‰. This alleviates the problem of substantial differences in RMSFE between bull and bear markets for implied return forecasts and shows that a forecast combination approach based on the state of the equity market is a suitable solution for the poor performance of implied expected returns in bear markets. Including time-series based forecasts slightly increases the RMSFE to 51.23‰ over the whole sample. However, it lowers the RMSFE in bear markets from 62.30‰ to 60.10‰. This is due to the Fama and French [1993] three factor model, which reports the lowest RMSFE in bear markets, in the benchmark case.

2 Methodology

2.1 Implied Expected Return

We now introduce the implied expected return methodology and the relevant proxies that are considered for mean-variance efficient portfolios, similar to Ardia and Boudt [2015] and Romijn [2016]. We consider a market with N risky securities and a risk-free asset, we denote a portfolio of risky securities by the (N x 1) vector w. The expected return is denoted by the (N x 1) vector μ and the covariance matrix is denoted as $\Sigma(N \text{ x } N)$. For simplicity we drop the time dimensions in this section. But we consider the risk aversion parameter, the risk-free rate, the conditional moments of the return distribution and the mean-variance efficient portfolio proxy to be time varying.

For some value of the risk aversion parameter $\gamma, \gamma < \infty$, a vector of weights w^* (N x 1) is mean-variance efficient if it maximizes the mean-variance utility function:

$$w^* \equiv \underset{w}{\operatorname{argmax}} \left\{ \mu^{'} w - \frac{\gamma}{2} \mathrm{w}^{'} \Sigma \mathrm{w} \right\}, \tag{1}$$

subject to $w^{'} \iota = 1$. The motivation for reverse engineering implied expected returns from this optimization problem, is that the determination of the optimal weights is very sensitive to changes in expected returns. Moreover, the optimized weights frequently exhibit extreme positions, which may be impossible for investors to implement. On the other hand, a good proxy for a mean-variance efficient portfolio is often available and the covariance matrix can generally be more accurately estimated than the expected return.

The solution to (1), under the budget constraint ($w^{'}\iota = 1$), equals:

$$w = \frac{1}{\gamma}\Sigma^{-1}(\mu - \ell\iota), \tag{2}$$

and from (2) we obtain the expression for the implied expected returns:

$$\mu = \ell\iota + \gamma\Sigma w, \tag{3}$$

where the intercept parameter ℓ is a consequence of imposing the budget restriction ($w^{'}\iota = 1$). Herold [2005] argues that it is necessary to impose this restriction in order to obtain meaningful implied expected returns. We set ℓ to the risk-free rate, as in Black and Litterman [1991]. Consequently, the implied expected return of a risk-less asset is equal to the risk-free rate. The risk aversion parameter γ is discussed in Section 2.5. Implied expected returns based on (3) with a risk aversion parameter of $\gamma = 2.4$, as in Ardia and Boudt [2015], will henceforward be referred to as the benchmark.

From (3) it becomes clear that implied expected returns in industry i can be negative when the portfolio weight w_i is negative. However, for four out of the five portfolio proxies considered in this paper, a negative weight is impossible by construction. This holds for the market capitalization weighted-, the equally weighted-, inverse volatility weighted- and equal risk contribution- portfolio. On the other hand, while the maximum diversification portfolio can have a negative weight in an industry, this only leads to one negative forecast out of 3240 forecasts. Ardia and Boudt [2015] consider three more portfolio proxies, but these proxies do not lead to negative forecasts as well.

2.2 Proxies for Mean-Variance Efficient Portfolios

We now introduce and discuss the different mean-variance efficient portfolio proxies used in this paper. We also present various assumptions under which these portfolio proxies are mean-variance efficient. It is not necessary that the assumptions hold for implied expected return forecasts to be more accurate than other return forecasts. The forecast error that follows from the portfolio proxy not being truly mean-variance efficient can be smaller than the forecast error that follows from other return forecasting models. We have discussed the market capitalization weighted portfolio before and we denote it henceforward as w_{mkt}.

2.2.1 Equally Weighted Portfolio

DeMiguel et al. [2009] showed that equally weighing each asset in the investable universe outperforms most portfolio optimization techniques in terms of Sharpe ratio and turnover for various popular datasets. As will be discussed in Section 3, our data consists of several industry portfolios. Because the number of firms differs across industries, the weight given to industry portfolio i equals the number of firms in industry portfolio i divided by the total number of firms in all ten industry portfolios. The equally weighted portfolio is denoted as w_{ew}. It is mean-variance efficient when the expected excess returns ($\mu - \ell\iota$) are proportional to the sum of their covariances ($\Sigma\iota$).

2.2.2 Inverse Volatility Weighted Portfolio

The inverse volatility weighted portfolio, as proposed by Leote De Carvalho et al. [2012], is mean-variance efficient under the condition that all Sharpe ratios are the same and all pairwise correlations are equal. The inverse volatility weighted portfolio is defined as:

$$w_{iv} = \left(\frac{\frac{1}{\sigma_1}}{\sum_{j=1}^{N} \frac{1}{\sigma_j}}, ..., \frac{\frac{1}{\sigma_N}}{\sum_{j=1}^{N} \frac{1}{\sigma_j}} \right)^{'} . \tag{4}$$

2.2.3 Equal Risk Contribution Portfolio

All assets in the equal risk contribution portfolio contribute equally to the volatility of the whole portfolio. The contribution of an asset i to the portfolio volatility is measured by the percentage volatility contribution:

$$\%RC_i = \frac{w_i[\Sigma w]_i}{w'\Sigma}, \tag{5}$$

where $[\Sigma w]_i$ is the volatility of asset i with portfolio w. The percentage volatility contribution of all N assets in the equal risk contribution portfolio should equal $1/N$. Therefore, portfolio weights satisfy:

$$w_{erc} = \underset{w}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N} (\%RC_i - \frac{1}{N})^2 \right\}, \tag{6}$$

under the constraint that $w^{'}\iota = 1$. The equal risk contribution portfolio is mean-variance efficient under the condition that all assets within the portfolio contribute equally to the portfolio excess return.

2.2.4 Maximum Diversification Portfolio

Let $\boldsymbol{\sigma} = \sqrt{\text{diag}(\Sigma)}$ be the vector of standard deviations of the assets. Because of diversification effects and sub-additivity, the portfolio's standard deviation is always less than or equal to the weighted average volatility:

$$\sqrt{w'\Sigma w} \leq w^{'}\boldsymbol{\sigma}. \tag{7}$$

Choueifaty and Coignard [2008] give the diversification ratio of a portfolio w by:

$$DR(w) \equiv \frac{w^{'}\boldsymbol{\sigma}}{\sqrt{w'\Sigma w}} \geq 1. \tag{8}$$

A portfolio w with a low portfolio standard deviation, $\sqrt{w'\Sigma w}$, relative to the weighted average volatility, $w^{'}\boldsymbol{\sigma}$, therefore obtains a high diversification ratio. Choueifaty and Coignard [2008] define the maximum diversification portfolio, w_{md}, as the portfolio that has the highest diversification ratio. It is obtained numerically, by maximizing the diversification ratio in (8). w_{md} is mean-variance efficient under the condition that the expected return of all assets in the portfolio are proportional to their volatility.

2.3 Identification and Forecasting of Bull and Bear Regimes

Because the problem of almost always positive implied return forecasts is specific to bear markets, we introduce parametric Markov regime-switching models and semi-parametric rule-based models, so we can identify and forecast bull and bear states in the equity market. The latter are characterized as periods with negative or lower returns, higher volatility and increased correlation (Campbell et al. [2002]). Identification of the state of the equity market is necessary for evaluating the forecasting performance of implied expected returns in a specific regime. This provides insight into whether implied expected returns actually perform worse in bear markets. Moreover, in our later application of forecast combinations, the weights differ across regimes. Identification of the regime is therefore crucial for determining forecast combination weights.

These methods for classifying the state of the equity market have been previously implemented and evaluated in the literature, Kole and van Dijk [2017] make a mean-variance based comparison of rule-based methods and Markov switching models. They conclude that Markov switching models perform worse when identifying the state of the equity market compared to rule-based algorithms, such as the algorithm of Lunde and Timmermann [2004]. However, Markov switching models as in Hamilton [1990] are better at forecasting the state of the equity market, because rule-based methods only look at the changes in the mean return, while Markov switching models incorporate volatility. Therefore rule-based methods are slower in picking up changes from one regime to another. Furthermore, in the algorithm of Lunde and Timmermann [2004] the price level has to drop 15% before a bear market is declared. If this drop does not occur in one time period, the algorithm is slower than Markov switching models by construction. Hence, Markov switching models are preferred when forecasting the regimes.

The state of the equity market at time t is denoted as S_t, where $S_t = 1$ and $S_t = 0$ correspond to a bull and bear market, respectively. For rule-based identification we implement the algorithm of Lunde and Timmermann [2004], which selects peaks and troughs of a stock market price index. In order for a new peak(trough) to be identified, it has to be a fraction $\lambda_1(\lambda_2)$ higher(lower) than the last peak(trough). A detailed description of the procedure is available in Section A.1. In order to forecast the state of the equity market in the next period we use a dynamic logit model, as in Kole and van Dijk [2017]. In this model, the effect of a predictor variable depends on the current state of the equity market.

Conditional on the current state q and predictor variables z_t, the probability of state s occurring at time $t+1$ is

$$\pi_{qs,t+1} = \Pr[S_{t+1} = s | S_t = q, z_t] = \Lambda(\beta_q^{'} z_t), \quad s, q \in \{0, 1\} \tag{9}$$

where $\Lambda(x)$ denotes the logistic function, β_q is the coefficient vector and z_t includes a constant. To make a forecast we require the current state at time t which may be unknown with the approach of Lunde and Timmermann [2004]. In this case we compute the forecast recursively. Assume that the last known extreme is identified at $t^\star < t$, the state of the equity market at t+1, S_{t+1} is constructed as in Kole and van Dijk [2017]:

$$\Pr[S_{\tau+1} | z_\tau] = \sum_{q \in \{0,1\}} \Pr[S_{\tau+1} = s | S_\tau = q, z_\tau] \Pr[S_\tau = q | z_{\tau-1}] \quad t^\star < \tau \leq t+1, . \tag{10}$$

When applying Markov switching models we assume that the state of the market follows a first-order Markov chain with transition probabilities $\pi_{qst} \equiv \Pr\left[S_t = s | S_{t-1} = q, z_{t-1}\right], s, q \in \{0, 1\}$, as in Kole and van Dijk [2017]. We consider a model with time-varying and constant transition probabilities[2]. When the transition probabilities are time-varying they are linked to predictor variables z_{t-1} by a logit specification, as in (9). A detailed description of the predictor variables used is available in Section 3. Although the Markov switching models characterize a distribution of returns, it is only used for forecasting the state of the equity market and should be seen as separate from the implied expected return forecasts. We consider them as separate, because obtaining regime-specific implied expected return estimates from the covariance in the Markov switching model likely reduces the forecast accuracy in bear markets. This holds, because the variance in bear markets is generally higher, which leads to higher implied return forecasts in bear markets.

[2]Models are estimated in MATLAB by implementing Perlin [2015] and Ding [2012].

2.4 Dynamic Forecast Combinations

To remedy the poor performance of implied expected returns in bear markets, we introduce a dynamic forecast combination approach, as in Deutsch et al. [1994].The weights applied to each forecast can be made conditional on the forecast of the state of the equity market in the next period. A formula, similar to Deutsch et al. [1994], for the forecast combination f_{it}^c at time t in industry i is then given by:

$$f_{it}^c = \Pr[S_t = 1|\mathcal{I}_{t-1}] \times \alpha_1^{'} f_{it} + \Pr[S_t = 0|\mathcal{I}_{t-1}] \times \alpha_2^{'} f_{it} \tag{11}$$

where $\mathcal{I}_{t-1}$ denotes the information available at time t - 1 and α_1 and α_2 are vector of weights assigned to the vector of forecasts f_{it}, at time t in industry i. In this paper, we make forecast combinations of only implied return forecasts and of implied returns and time-series based forecasts. In the former case, the resulting forecast can still be seen as a implied return forecast. We present an argument for this in Section A.3 in the Appendix. Further, $\Pr[S_{t+1} = 1|\mathcal{I}_t]$ is obtained from the Markov switching models or the algorithm of Lunde and Timmermann [2004] described in Section 2.3. The vector of weights α_1 and α_2 are obtained from a switching linear regression model:

$$\begin{cases} y_{it} = \alpha_1^{'} f_{it} + \epsilon_{it} & \text{if } S_t = 1 \\ y_{it} = \alpha_2^{'} f_{it} + \epsilon_{it} & \text{if } S_t = 0 \end{cases} \tag{12}$$

where y_{it} and f_{it} are the return and a vector of return forecasts, at time t in industry i, respectively. Despite having observations for ten different industries, we do not estimate industry specific parameters as that will increase the parameters in α_1 and α_2 tenfold. We choose to pool industry data and estimate the model using pooled OLS[3]. The latent state of the equity market S_t can be obtained either trough Markov switching models or the algorithm of Lunde and Timmermann [2004]. The former identify a bull(bear) market when the smoothed probability of a bull(bear) market exceeds a threshold. Because we apply these probabilities to the estimation of forecast combination weights we want to be relatively certain about the state of the equity market. We therefore identify a period as a bull market when the smoothed probability exceeds 70%. On the other hand, bear markets are identified when their probability exceeds 60%. Periods with probabilities in between these thresholds are discarded while estimating the weights. The threshold for bear markets is lower than for a bull market, because a higher threshold reduces the sample size of bear market observations too much. The lowest number of forecasts we estimate the weights over is 160.

Because simple forecast combination schemes tend to outperform more complex combination schemes, as noted by Timmermann [2006], we propose a forecast combination method that multiplies the average forecast by a scaling factor, dependent on the regime. With this approach we attempt to fix the bias of implied return forecasts in bear markets. In the case of implied expected returns, not incorporating the scaling factor results in always positive forecasts, which likely does not lead to considerable improvements in bear markets. We give this forecast combination approach by:

$$f_{it}^c = \Pr[S_t = 1|\mathcal{I}_{t-1}] \times \alpha_1 \overline{f_{it}} + \Pr[S_t = 0|\mathcal{I}_{t-1}] \times \alpha_2 \overline{f_{it}}, \tag{13}$$

where $\overline{f_{it}}$ is the average of the vector f_{it} in (11), while α_1 and α_2 are scalars estimated by pooling the industries and different forecast methods together and using a switching linear regression model, similar to (12).

[3]This leads to the forecast combination weights minimizing the mean squared error, because the residual in this model is the forecast error in industry i at time t.

Although Markov switching models incorporate both the mean and the volatility of returns, a more simple approach that only incorporates volatility when making forecast combinations could prove insightful. Therefore, we also employ a smooth transition regression model, as in Lin and Teräsvirta [1994] and Deutsch et al. [1994]. As a measure for volatility we take the CBOE Volatility Index (VIX) which is available from 1990. The VIX reflects the level of implied volatility of liquid options on the S&P500. The smooth transition regression model is then given by:

$$f_{it}^c = (1 + e^{\beta V_{t-1}})^{-1}\alpha_1 f_{it} + (1 - (1 + e^{\beta V_{t-1}})^{-1})\alpha_2 f_{it}, \tag{14}$$

where V_{t-1} is the lagged value of the VIX and β is the slope parameter of the transition function. This parameter is an indicator for the speed of the transition. We perform a grid search for the slope parameter β, choosing the value that minimizes the sum of squared errors over the past 36 months. We can then estimate the forecast combinations weights α_1 and α_2 using pooled OLS. The approach described above allows for a gradual change in the weights of the forecast combination, but perhaps a sudden change is more desirable. Therefore, we also implement a forecast combination approach based on a threshold model, similar to Deutsch et al. [1994]. In this model, the regime is indicated by the level of the VIX. The formula for the forecast combination is then given by:

$$f_{it}^c = \mathbb{1}(V_{t-1} > \theta) \times \alpha_1' f_{it} + \mathbb{1}(V_{t-1} < \theta) \times \alpha_2' f_{it}, \tag{15}$$

where θ is a threshold chosen by the researcher and $\mathbb{1}(\cdot)$ is the indicator function. A good starting point for θ, is the average value of the VIX over the sample. Alternative values will be investigated in our sensitivity study.

Lastly, to discern the gain in forecasting performance that follows from accounting for the state of the equity market. We implement a forecast combination model that does not account for this state. Such a forecast combination is given by:

$$f_{it}^c = \alpha_1' f_{it}, \tag{16}$$

where weights α_1 are estimated by pooled OLS, similar to (12), but without accounting for the state of the equity market. In this paper, the standard approach for determining the forecast combination weights and forecasting the state of the equity market is the Markov switching model with constant transition probabilities. Other approaches are evaluated in our sensitivity study.

2.5 Risk Aversion

Mehra and Prescott [1985] show that the historical equity premium in the US would require an unusually high risk aversion for individual investors. This is called the equity premium puzzle. Because the risk aversion is a main component in the calculation of the implied expected return in (3), it is crucial that it is a good estimate for the true risk aversion of investors. Bollerslev et al. [2011] calculate the volatility risk premium and link it to risk aversion by switching its sign. In the calculation of this volatility risk premium, the monthly realized volatility and implied volatility are needed. For implied volatility we use the CBOE Volatility Index (VIX). Our realized volatilities are based on the summation of daily squared returns of the S&P500 index within a month, this provides a poorer approximation of the true unobserved integrated volatility compared to Bollerslev et al. [2011], who base their realized volatility on five-minute returns. However, their Monte Carlo shows that the finite sample performance of the GMM estimator discussed below, improves with an increased sample size. Because of the increased availability of daily return data, we base our realized volatilities on daily returns. The rest of this section contains a summary of the methodology employed by Bollerslev et al. [2011] supplemented by our own approach to linking the volatility risk premium to the risk aversion parameter.

First, a general continuous-time stochastic volatility model for the logarithmic stock price ($p_t = \log S_t$) is assumed:

$$\begin{aligned} dp_t &= \mu_t(\cdot) + \sqrt{V_t} dB_{1t} \\ dV_t &= \kappa(\theta - V_t)dt + \sigma_t(\cdot)dB_{2t} \end{aligned} \tag{17}$$

where V_t is the volatility, and $B_{it}, i = 1, 2$ are Wiener processes with mean zero and variance t. κ is the rate at which the volatility reverts to its long run average θ. Further, $corr(dB_{1t}, dB_{2t}) = \rho$ denotes the leverage effect and the functions $\mu_t(\cdot)$ and $\sigma_t(\cdot)$ should satisfy the usual regularity conditions but can be left unspecified. Assuming no arbitrage, the corresponding risk-neutral distribution, as in Bollerslev et al. [2011], is given as:

$$\begin{aligned} dp_t &= r_t^\star dt + \sqrt{V_t} dB_{1t}^\star \\ dV_t &= \kappa^\star(\theta^\star - V_t)dt + \sigma_t(\cdot)dB_{2t}^\star \end{aligned} \tag{18}$$

where $B_{it}^\star, i = 1, 2$ are Wiener processes with respect to the risk-neutral measure . Further, $corr(dB_{1t}^\star, dB_{2t}^\star) = \rho$ and $r_t^\star$ is the risk-free interest rate. The parameters in (18) are related to the parameters in (18) by, $\kappa^\star = \kappa + \lambda$ and $\theta^\star = \kappa\theta/(\kappa + \lambda)$, where λ is the volatility risk premium. The volatility V_t in (17) and (18) is latent, but the realized volatility provides a simple approach to measure the integrated volatility over a time period. Let $V_{t,t+\Delta}^n$ denote the realized volatility computed by summing the squared returns over the $[t, t + \Delta]$ interval:

$$V_{t,t+\Delta}^n = \sum_{i=1}^{n} \left[p_{t+\frac{i}{n}(\Delta)} - p_{t+\frac{i-1}{n}(\Delta)} \right]^2, \tag{19}$$

from the theory of quadratic variation it then follows,

$$\lim_{n\to\infty} V_{t,t+\Delta}^n \overset{a.s}{\to} V_{t,t+\Delta} = \int_t^{t+\Delta} V_s ds. \tag{20}$$

Hence, for a sufficiently large n, relative to Δ, the realized volatility is a good approximation for the integrated volatility $V_{t,t+\Delta}$. As derived by Bollerslev and Zhou [2002], the conditional expectation of the integrated volatility under the physical measure equals

$$\mathbb{E}(V_{t+\Delta,t+2\Delta}) = \alpha_\Delta \mathbb{E}(V_{t,t+\Delta}|\mathbb{F}_t) + \beta_\Delta, \tag{21}$$

where $\alpha_\Delta = e^{-\kappa\Delta}$ and $\beta_\Delta = \theta(1 - e^{-\kappa\Delta})$ are functions of the parameters κ and θ in (17).

Let $IV_{t,t+\Delta}^\star$ be the implied volatility, computed as a weighted average of a series of Δ-maturity options. As shown by Britten-Jones and Neuberger [2000], this implied volatility is then equal to the risk-neutral expectation of the integrated volatility.

$$IV_{t,t+\Delta}^\star = \mathbb{E}^\star(V_{t,t+\Delta}|\mathbb{F}_t), \tag{22}$$

where $\mathbb{E}^\star(\cdot)$ indicates that the expectation is taken under the risk-neutral measure. Combining the previous results allows us to link the expectation of the integrated volatility under the risk-neutral dynamics in (18) with the objective expectation of the integrated volatility in (17). As shown by Bollerslev and Zhou [2006],

$$\mathbb{E}(V_{t,t+\Delta}|\mathbb{F}_t) = \mathcal{A}_\Delta IV^\star_{t,t+\Delta} + \mathcal{B}_\Delta, \tag{23}$$

where $\mathcal{A}_\Delta = \frac{(1-e^{-\kappa\Delta})/\kappa}{(1-e^{-\kappa^\star\Delta})/\kappa^\star}$ and $\mathcal{B}_\Delta = \theta\left[\Delta - (1-e^{-\kappa\Delta})/\kappa\right] - \mathcal{A}_\Delta\theta^\star\left[\Delta - (1-e^{-\kappa^\star\Delta})/\kappa^\star\right]$ are functions of the parameters κ, θ and λ. The moment restriction in (23), together with the moment restriction in (21), let us identify the constant volatility risk premium λ. We now construct a GMM type estimator, as in Bollerslev et al. [2011]. They add the lagged instrument of realized volatility to the moment conditions in (21) and (23). This means we can test whether the moment conditions are suitable, because the following system of equations is over-identified:

$$f_t(\xi) = \begin{bmatrix} V_{t+\Delta,t+2\Delta} - \alpha_\Delta V_{t,t+\Delta} - \beta_\Delta \\ (V_{t+\Delta,t+2\Delta} - \alpha_\Delta V_{t,t+\Delta} - \beta_\Delta)V_{t-\Delta,t} \\ V_{t,t+\Delta} - \mathcal{A}_\Delta IV^\star_{t,t+\Delta} - \mathcal{B}_\Delta \\ (V_{t,t+\Delta} - \mathcal{A}_\Delta IV^\star_{t,t+\Delta} - \mathcal{B}_\Delta)V_{t-\Delta,t} \end{bmatrix}, \tag{24}$$

where $\xi = (\kappa, \theta, \lambda)'$ and by construction $\mathbb{E}\left[f_t(\xi_0)|\mathbb{F}_t\right] = 0$. The corresponding GMM estimator is given by $\hat{\xi}_T = \text{argmin} \quad g_T(\xi)' W g_T(\xi)$, where $g_T(\xi) \equiv 1/T \sum_{t=2}^{T-2} f_t(\xi)$ is the sample mean of the moment conditions and W denotes the asymptotic covariance matrix of $g_T(\xi_0)$. We use a two-step feasible GMM estimator. Under standard regularity conditions, $\mathbb{J} = min_\xi g_T(\xi)' W g_T(\xi)$ multiplied by the sample size is asymptotically $\chi^2(1)$ distributed. Hence, we can test the over-identifying restrictions by means of an omnibus test. Similar to Bollerslev et al. [2011], we use a heteroskedasticity and autocorrelation consistent covariance estimator, due to the dependence in the moment conditions in (21) and (23).

To make the volatility risk premium time-varying, we assume that it follows an augmented AR(1) process, as in Bollerslev et al. [2011],

$$\lambda_{t+1} = a + b\lambda_t + \sum_{k=1}^{K} c_k \times x_{t,k}, \tag{25}$$

where $x_{t,k}, k = 1,..K$ are macro-economic state variables. A detailed description of these variables can be found in Section 3. We identify the parameters a, b and c_k, by adding additional moment conditions to (24). We add lagged squared realized volatility, lagged implied volatility and every macro-economic state variable, except lagged realized volatility , as instruments for the moment condition in (23). Lagged realized volatility is not added as an instrument, since we already use it as an instrument in (24). This results in the following system of equations:

$$f_t(\psi) = \begin{bmatrix} (24) \\ (V_{t,t+\Delta} - \mathcal{A}_\Delta IV^\star_{t,t+\Delta} - \mathcal{B}_\Delta)V^2_{t-\Delta,t} \\ (V_{t,t+\Delta} - \mathcal{A}_\Delta IV^\star_{t,t+\Delta} - \mathcal{B}_\Delta)IV^\star_{t-\Delta,t} \\ (V_{t,t+\Delta} - \mathcal{A}_\Delta IV^\star_{t,t+\Delta} - \mathcal{B}_\Delta)x_{t,1} \\ \vdots \\ (V_{t,t+\Delta} - \mathcal{A}_\Delta IV^\star_{t,t+\Delta} - \mathcal{B}_\Delta)x_{t,K-1} \end{bmatrix}, \tag{26}$$

where $\psi = (\kappa, \theta, a, b, c_1, \cdots, c_K)$ and $x_{t,k}, k = 1,.., K-1$ are all macro-economic state variables except lagged realized volatility. The top four rows of (26) are equal to the moment conditions in (24). Adding these moment conditions leads to the same asymptotic $\chi^2(1)$ distribution for the GMM omnibus test as in the constant volatility risk premium specification. We can then employ the same GMM estimator, but then over the new moment conditions.

After estimating the volatility risk-premium, either in a time-varying or constant specification, Bollerslev et al. [2011] link it to the risk aversion. They first assume that $\sigma_t(\cdot) = \sigma\sqrt{V_t}$ in (17). It then follows that

$$-\lambda V_t = cov_t(\frac{dm_t}{m_t}, dV_t), \tag{27}$$

where m_t denotes the marginal utility of wealth for the representative investor. Moreover, if we assume that this investor has a power utility function it follows that:

$$cov_t(\frac{dm_t}{m_t}, dV_t) = \gamma\rho\sigma V_t. \tag{28}$$

Combining (27) and (28) shows that the coefficient of constant relative risk aversion $\gamma = \lambda/(\rho\sigma)$. Bollerslev et al. [2011] find that for their sample, $\rho = -0.8$ and $\sigma = 1.2$, such that $\gamma = -\lambda$. Because our sample is different and we are concerned with forecasting performance, we do not estimate ρ and σ. Instead, we choose a value for the multiplication constant $\phi = 1/(\rho\sigma)$ with $\phi < 0$, that minimizes the RMSFE of the implied expected return over the past observations. (Time-varying) Risk aversion is then given by $\gamma_t = \phi\lambda_t$, where $\lambda_t = \lambda$ in the constant volatility risk premium specification. A separate multiplication constant is estimated for both the constant and time-varying volatility risk premium specification.

2.6 Forecast Evaluation

We analyze forecasting performance by computing all forecasting performance measures mentioned in this section over the whole sample and in both states of the equity market separately. Moreover, all measures are computed for each industry separately and then averaged over the industries. For a more detailed evaluation of the forecasting performance of implied expected returns in specific industries, refer to Ardia and Boudt [2015] and Romijn [2016].

We analyze forecasting precision, in industry i, using the out-of-sample RMSFE:

$$RMSFE_i = \sqrt{\frac{\sum_{t=1}^{T}(\hat{y_{it}} - y_{it})^2}{n}}, \tag{29}$$

where $\hat{y_{it}}$ and y_{it} are the one-month ahead forecast and the actual value in industry i, respectively. Significance of the difference between the RMSFE of two models will be tested using a Diebold and Mariano [1995]. This leads to loss differentials for ten different industries, which we average when implementing test. Additionally, we consider two more precision measures, namely the hit ratio (HR) and the mean percentage point error(MPPE). The former reports the percentage of times the forecast had the right sign and the latter reports the average percentage point error of the actual value compared to the forecast. This means that if the actual return is 2% and we forecast a 1% return , the MPPE equals (2% - 1%) = 1%.

We choose to use the MPPE in addition to the RMSFE, because the former preserves information on whether forecasts are too high or too low. We find this information important, because we expect implied return forecasts to be on average too high in bear markets and aim to remedy this problem. Moreover, we choose the MPPE instead of the mean percentage error, because our return data has some values close to zero. These values would skew the percentage error, whereas the MPPE does not have this problem.

When we evaluate the hit ratio for standard implied return forecasts, it is merely an indication of the fraction of positive returns in the dataset because the forecasts are almost always positive. However, introducing forecast combinations can lead to negative forecasts which makes investigating the hit ratio meaningful.

Finally, we consider the instability and dispersion of the forecasts. The latter is given by the standard deviation across assets of the return forecasts, while the former is given by the standard deviation across time of the return forecasts. Because of the possible application of these forecasts to portfolio optimization, high dispersion across assets is undesirable since it makes trading more difficult. Furthermore, high instability in the forecasts can lead to excessive transaction costs when using the forecasts in a trading strategy and should therefore be investigated. Higher dispersion and instability are not necessarily bad, but an increase in one of the two should lead to an increase in forecasting accuracy, all else equal. Otherwise, there is no benefit in increasing volatility of the forecasts.

3 Data and Estimation Methods

The monthly returns we wish to forecast in this paper, are the ten US industry portfolios provided by Kenneth French [4]. The dataset contains monthly returns and information about the composition of the industry portfolios, ranging from January 1927 to March 2017. The ten US industries in this dataset are: Non-Durables, Durables, Manufacturing, Energy, Hi-Tech, Telecom, Shops, Utilities, Healthcare and Others. Datasets are available with more industries, however we choose to limit the number of industries to ten to curb the dimensions of the covariance matrix. For the risk-free rate we take the 3-month US T-Bill rate.

To compute the (time-varying) volatility risk premium we require realized volatility, implied volatility and macro-economic state variables. Due to the availability of data, we choose to compute realized volatilities by means of daily returns of the S&P500, obtained from the Chicago Boards of Options Exchange (CBOE) starting at January 1986. As a measure of implied volatility we source the CBOE Volatility Index (VIX) from the CBOE. The VIX reflects the level of implied volatility of liquid options on the S&P500. Observations range from January 1990 to December 2016. The macro-economic state variables are the realized volatility, the AAA bond spread over treasuries, housing starts, the S&P500 price to earnings ratio, industrial production, producer price index and payroll employment. Variables are provided by the Federal Reserve Bank of St.Louis in their FRED[5] and ALFRED[6] databases. Bollerslev et al. [2011] conclude that the aforementioned seven variables, out of 29 potential variables, jointly achieve the highest p-value of the GMM omnibus specification test and are individually significant based on a t-test with a 5% confidence level. We follow their suggestion for including these variables in the specification of the time-varying volatility risk-premium in (25) and verify their results in our sample.

We choose the macro-economic and financial variables considered in our application of regime-switching models, based on prior studies that indicate their success in identifying and forecasting the state of the equity market. We gather data for inflation, industrial production, unemployment, the T-Bill rate, the term and credit spreads and the dividend-to-price ratio from the FRED and ALFRED databases. Additionally, variance based on the RiskMetrics approach is used as a predictor variable as well. Refer to Kole and van Dijk [2017] for an application of these variables to Markov switching models. For variable selection we use a specific-to-general approach, as in Kole and van Dijk [2017]. We

[4] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
[5] https://fred.stlouisfed.org/
[6] https://alfred.stlouisfed.org/

Table 1: Summary Statistics of Return Data

Industries	N	Mean(%)	Vol(%)	Skew	Kurt
Non Durables	220	0.76	3.64	−0.49	3.92
Durables	106	0.57	7.64	0.31	7.95
Manufacturing	534	0.82	5.08	−0.53	4.77
Energy	161	0.86	5.99	−0.00	3.40
Hi-Tech	903	0.82	7.54	−0.33	4.05
Telecom	119	0.53	5.50	−0.21	4.10
Shops	438	0.76	4.53	−0.36	3.86
Healthcare	529	0.71	4.20	−0.36	3.40
Utilities	110	0.76	4.27	−0.61	3.67
Other	1398	0.54	5.44	−0.67	5.29
Average	452	0.71	5.38	−0.32	4.44

Notes: N: Average number of stocks in industry. Skew: Skewness. Kurt: Kurtosis. Average: Average over the column. Statistics are computed from June 1998 until December 2016.

determine which variable combination leads to the largest improvement in the likelihood function and add it when the improvement is significant at the 10% level, based on a likelihood ratio test. We conduct an Augmented Dickey-Fuller (ADF) test , as in Dickey and Fuller [1979], to test for stationarity of the time-series. If the presence of a unit root is not rejected, we transform the corresponding variable by subtracting the yearly moving average. Volatility is I(1) by construction, therefore we do not conduct an ADF test for this predictor variable.

In order to facilitate a comparison between the different forecast methods, we evaluate the forecasts between June 1998 and December 2016 so that the (time-varying) volatility risk premium can be estimated with a sample of no less than 100 observations. This period consists of 223 observations in each industry, of which 48 are observations in a bear market, as classified by the algorithm of Lunde and Timmermann [2004]. In Table 1 we report summary statistics over this period. It shows that the average monthly return is 0.71%, whereas the volatility is 5.38% on average. The monthly return distribution is on average negatively skewed (-0.32) and leptokurtic (4.44). The Durables sector seems most unattractive to invest in, with an average return of 0.57% and volatility of 7.95%. The kurtosis of most industries is below the average, because the Durables and Other industries are outliers with a kurtosis of 7.95 and 5.29, respectively. Summary statistics conditional on the state of the equity market are provided in Section A.4 in the Appendix.

As estimator for the covariance matrix, we take the exponentially weighted moving average model (EWMA) with a decay parameter of $\lambda = 0.97$, as advocated by RiskMetrics Group [1996]. This leads to a memory of 33 months. Romijn [2016] investigates the benefits of employing a multivariate GARCH model (Bollerslev et al. [1988]), instead of EWMA covariance estimator, in the context of implied expected returns. He concludes that the former leads to marginally better forecasting accuracy. However, we choose the EWMA estimator because it is implemented more easily.

4 Results

4.1 Benchmark

We first evaluate the forecasting performance of the benchmark, unconditional on the state of the equity market. Further, we compare the forecasting performance between bull and bear states. This comparison highlights the discrepancy in forecasting accuracy between bull and bear states and shows the shortcomings of the implied expected return methodology in bear markets.

Table 2: Forecast Evaluation for Expected Return Forecasts in Benchmark Case Unconditional on the State of the Equity Market

	RMSFE(‰)	MPPE(%)	HR(%)	$\boldsymbol{\sigma_1}$	$\boldsymbol{\sigma_2}$
Panel A: Implied Return Forecasts					
$\boldsymbol{w_{ew}}$	53.64(0.079)	−0.24	59.37	0.27	0.20
$\boldsymbol{w_{mkt}}$	53.66(0.082)	−0.22	59.42	0.26	0.19
$\boldsymbol{w_{iv}}$	53.67(0.057)	−0.14	59.42	0.21	0.11
$\boldsymbol{w_{erc}}$	53.64(0.047)	−0.15	59.42	0.22	0.12
$\boldsymbol{w_{md}}$	53.67(0.030)	−0.07	59.42	0.18	0.09
Panel B: Time-Series Based Forecasts					
Constant	54.81	−0.43	55.96	0.88	0.55
Fama	54.32	0.29	52.33	1.23	0.66
AR(1)	55.94	−0.48	56.55	1.52	0.96

Notes: This table reports forecast evaluation statistics for the benchmark case and the time-series based forecasts over the whole sample. RMSFE: root-mean-squared forecast error, in per mille with p-values of Diebold-Mariano test in parentheses. This test compares implied return forecasting methods to the Fama and French [1993] three factor model. MPPE: mean percentage point error in percent. HR: hit ratio in percent. $\boldsymbol{\sigma_1}$ is instability over time, in percent and averaged over the ten industries. $\boldsymbol{\sigma_2}$ is dispersion across assets, in percent and averaged over time. Constant: the forecasts under a constant mean model. Fama: three factor model forecasts. AR(1): forecasts of autoregressive model with order one. The evaluation sample is between June 1998 and December 2016.

We report the forecasting performance of the benchmark and time-series based forecasts, unconditional on the state of the equity market, in Table 2. It shows that all five implied return forecasting methods in the benchmark case significantly (10%) improve the forecasting accuracy compared to the Fama and French [1993] three factor model. We choose to report the results for the comparison to the three factor model, because it denotes the lowest RMSFE in Panel B of Table 2. Comparing the implied return forecasts in Panel A to each other, reveals that there is no forecast that significantly outperforms any other forecast, at the 10% level. For the results of this pairwise comparison, refer to Section A.2 in the Appendix.

The hit ratio for every implied return forecast is higher than the time-series based forecasts. This shows that being able to forecast negative returns does not always lead to a higher hit ratio, since incorrectly predicting negative returns is now a problem. Implied expected returns do not have this problem, which apparently improves their hit ratio. The dispersion and instability are lower for implied return forecasts compared to the benchmark case. Increased volatility in the time-series based forecasts does therefore not pay off. Aforementioned results are in line with the results in Ardia and Boudt [2015] and Romijn [2016] whom also conclude that implied expected return forecasts outperform the same time-series based forecasts in their sample.

Table 3: Forecast Evaluation for Expected Return Forecasts in Benchmark Case Conditional on the State of the Equity Market

(a) Bull Market

	RMSFE(‰)	**MPPE(%)**	**HR(%)**	$\boldsymbol{\sigma_1}$	$\boldsymbol{\sigma_2}$
Panel A: Implied Return Forecasts					
$\boldsymbol{w_{ew}}$	46.74 (0.000)	0.79	65.89	0.28	0.19
$\boldsymbol{w_{mkt}}$	46.82 (0.000)	0.81	65.89	0.27	0.18
$\boldsymbol{w_{iv}}$	47.03 (0.000)	0.89	65.89	0.22	0.11
$\boldsymbol{w_{erc}}$	46.96 (0.000)	0.87	65.89	0.23	0.12
$\boldsymbol{w_{md}}$	47.27 (0.000)	0.95	65.89	0.18	0.10
Panel B: Time-Series Based Forecasts					
Constant	48.57	0.52	59.89	0.90	0.55
Fama	48.80	1.28	52.63	1.19	0.60
AR(1)	50.07	0.40	59.54	1.40	0.89

(b) Bear Market

	RMSFE(‰)	**MPPE(%)**	**HR(%)**	$\boldsymbol{\sigma_1}$	$\boldsymbol{\sigma_2}$
Panel A: Implied Return Forecasts					
$\boldsymbol{w_{ew}}$	73.15 (0.999)	-4.01	35.83	0.22	0.25
$\boldsymbol{w_{mkt}}$	73.10 (0.999)	-3.99	35.83	0.21	0.23
$\boldsymbol{w_{iv}}$	72.58 (0.999)	-3.89	35.83	0.17	0.10
$\boldsymbol{w_{erc}}$	72.63 (0.999)	-3.87	35.63	0.20	0.12
$\boldsymbol{w_{md}}$	72.07 (0.999)	-3.77	35.83	0.15	0.07
Panel B: Time-Series Based Forecasts					
Constant	72.90	-3.90	41.67	0.72	0.58
Fama	70.59	-3.32	51.25	1.31	0.71
AR(1)	73.17	-3.67	45.63	1.76	1.20

Notes: Tables 3a and 3b report forecast evaluation statistics for the benchmark case and the time-series based forecasts conditional on the state of the equity market. RMSFE: root-mean-squared forecast error, in per mille with p-values of Diebold-Mariano test in parentheses. This test checks whether implied return forecasting methods outperform the three factor model. MPPE: mean percentage point error in percent. HR: hit ratio in percent. $\boldsymbol{\sigma_1}$ is instability over time, in percent and averaged over the ten industries. $\boldsymbol{\sigma_2}$ is dispersion across assets, in percent and averaged over time. Constant: the forecasts under a constant mean model. Fama: three factor model forecasts. AR(1): forecasts of autoregressive model with order one. The evaluation sample is between June 1998 and December 2016.

Tables 3a and 3b report the forecasting performance of implied expected returns and time-series based forecasts conditional on the state of the equity market. These tables show that in bull markets, implied expected return forecasts significantly outperform the three factor model, at the 1% level. Conversely, the three factor model significantly outperforms the implied expected return forecasts in bear markets, at the 1% level. The second result holds, because the p-values of testing whether the three factor model outperforms the implied return forecasts equal one minus the p-values of the test reported in Table 3b. This is true because the Diebold-Mariano statistic of the first test equals minus one times the Diebold-Mariano statistic of the second test. Because these statistics are both standard normally distributed, their respective p-values equal one minus the other[7].

The three factor model clearly outperforms the two other time-series based forecasts in bear markets. It reports the lowest RMSFE at 70.59‰ and the highest hit ratio at 51.25%. Hence, it is the only time-series based forecast method that significantly outperforms the implied expected returns in bear

[7] Let p_{xy} denote p-value of a Diebold-Marino test, which tests whether forecast method x outperforms forecast method y. Henceforward, we report p-values p_{xy} and let $p_{yx} = 1 - p_{xy}$.

markets. For the results of this and other pairwise comparisons in Table 3, refer to Section A.2 in the Appendix.

In bull markets, implied return forecasts based on $\boldsymbol{w_{ew}}$, significantly (1%) outperform the other implied expected return forecasts. On the other hand, the latter significantly (1%) outperform the implied return forecasts based on $\boldsymbol{w_{ew}}$, in bear markets. This shows that equally weighing assets in the proxy works in bull markets, but it is not optimal in bear markets. In bear markets, it is the $\boldsymbol{w_{md}}$ portfolio that leads to implied expected returns forecasting significantly (1%) better than the other implied expected returns. Moreover, the other two risk-based implied returns also forecast significantly (1%) better in bear markets, than the implied returns based on the $\boldsymbol{w_{ew}}$ and $\boldsymbol{w_{mkt}}$ portfolios. This suggests that accounting for the risk of assets leads to increased performance in bear markets, while this is not the case in bull markets.

When switching from bull to bear markets, time-series based forecasts report an increase in RMSFE from 49.15‰ to 72.22‰. This shows that they also suffer from poor performance in bear markets. However, it is expected that forecast models generally perform worse in bear markets, because the volatility of returns is higher in bear markets. Tables 9a and 9b in the Appendix reveal that the average volatility of monthly returns in our sample is 4.60% in bull markets, whereas it is 6.53% in bear markets. Dividing the RMSFE by the volatility, conditional on the regime, reveals that the RMSFE per unit of volatility is 1.07 in bull markets and 1.11 in bear markets for time-series based forecasts. This shows that the discrepancy between forecasting performance in bull and bear markets is relatively small when correcting for volatility. However, since optimal forecasts should be able to capture the time variation in actual returns, the difference in RMSFE between bull and bear markets is still relevant.

Implied return forecasts report an increase in RMSFE from 46.94‰ in bull markets to 72.70‰in bear markets. This increase is larger than for time-series based forecasts and can be explained by the MPPE and the HR. The latter is on average 35.59% in bear markets, whereas it is 65.89% in bull markets. Moreover, the MPPE is -3.91% in bear markets, while it is 0.86% in bull markets. Time-series based forecasts show a similar drop in MPPE from bull to bear markets, but the HR shows a smaller decrease from 57.35% to 46.18%. The three factor model even correctly predicts the sign in bear markets, more than half of the time. These results show that being able to predict negative returns positively impacts the hit ratio in bear markets. However, it does not mean that it improves the MPPE as well, since the magnitude of the negative forecast can still be incorrectly predicted. Although implied returns report a somewhat lower average instability of the forecasts in bear markets(0.19 vs. 0.23), there is no clear distinction in instability and dispersion between regimes.

In Figure 1 we present the evolution of the RMSFE. It shows that the RMSFE is generally higher during bear markets. Forecasting accuracy in relatively stable markets, such as the period between 2003 and 2007, is noticeably higher than in bear markets. The period from 1998 to 2003 shows the most volatility in the RMSFE and also reports the highest values. This period coincides with the dot-com bubble and this explains the large forecast errors which a model that always predicts positive returns would make.

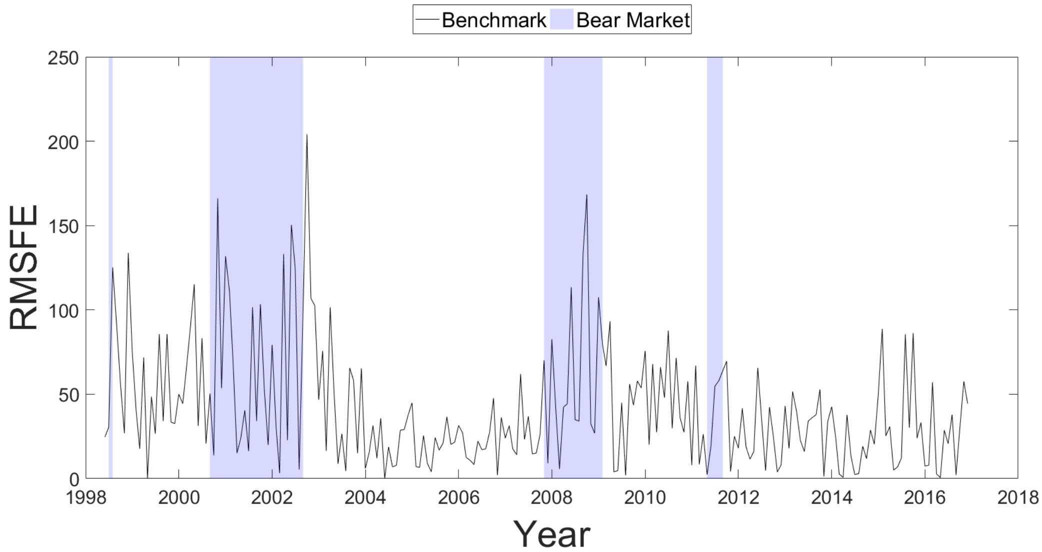


Figure 1: Evolution of RMSFE in the Benchmark Case

Notes: This figure shows the RMSFE for implied expected returns in the benchmark case. RMSFE is computed at each point in time over that particular forecast and is therefore the same as the absolute forecast error. Results are presented for the equally weighted portfolio and the High-Tech industry. Bear markets are classified by the algorithm of Lunde and Timmermann [2004].

Summarizing, we find that implied return forecasts perform relatively poorly in bear markets, compared to bull markets and time-series based forecasts. This is mainly due to the positive nature of implied return forecasts which negatively impacts the RMSFE, hit ratio and MPPE in bear markets, as shown in Tables 3a and 3b. We also see that implied expected returns can forecast the return in bull states relatively accurately. Because the shortcomings of implied return forecasts are so regime-dependent, we expect the dynamic forecast combination approach to improve the forecasting accuracy.

4.2 Risk Aversion

We now investigate the impact of introducing (time-varying) risk aversion computed from the (time-varying) volatility risk premium described in Section 2.5 and compare the forecasting performance to the benchmark case.

Table 4: Estimation Results of Volatility Risk Premium

Variables	Constant	Time-Varying
λ	$-1.810\,(0.000)$	
a		$-0.022\,(0.072)$
b		$0.986\,(0.000)$
c1 - Realized volatility		$-0.056\,(0.096)$
c2 - Credit spread		$-0.113\,(0.022)$
c3 - Housing starts		$-0.095\,(0.002)$
c4 - S&P500 P/E ratio		$0.110\,(0.002)$
c5 - Industrial production		$-0.069\,(0.081)$
c6 - Producer price index		$0.116\,(0.001)$
c7 - Payroll employment		$-0.046\,(0.098)$
$\chi^2(\text{d.o.f} = 1)$	$1.280\,(0.258)$	$0.012\,(0.913)$

Notes: Constant refers to the constant volatility risk premium specification. Time-Varying refers to the time-varying volatility risk premium specification. $\chi^2(\text{d.o.f} = 1)$ reports the Chi-squared statistic of the GMM-omnibus test with the p-value in parentheses. Parentheses in other rows refer to p-values of individual t-tests. Results are obtained over the sample January 1990 to December 2016.

Table 4 reports the GMM estimation results for the constant and time-varying volatility risk premium specifications. All seven variables have been standardized to have mean zero and variance one so that their marginal contributions to the (time-varying) volatility risk premium are comparable. For the constant volatility risk premium specification we find a significant negative volatility risk premium of -1.81. Introducing time variation implies an average risk premium of $a/(1-b) = -1.58$ for the autoregressive part, since there is a high degree of persistence in the time-varying specification (b = 0.986). The GMM omnibus test does not reject the null hypothesis of plausible moment conditions for both specifications. The constant and time-varying specification report p-values of 0.26 and 0.91, respectively. Furthermore, all macro-economic variables are statistically significant at the 10% level based on individual t-tests.

Using that the multiplication constant ϕ is always negative, we can interpret the sign of the coefficient values in Column 3 of Table 4. Moreover, since the risk aversion is given by, $\phi\lambda$, we present marginal contributions to the time-varying risk aversion as if $\phi = -1$. This makes interpretation of the marginal contributions easier. Coefficient values can be multiplied by the multiplication constant at each point in time, to obtain the true marginal contributions.

Realized volatility (-0.056) has a positive impact on the risk aversion, suggesting that risk aversion is higher when volatility is higher. Credit spread (-0.113) has a higher positive impact on the risk aversion than realized volatility, suggesting that a higher premium on highly rated bonds corresponds with higher risk aversion. Moreover, housing starts have a positive impact (-0.095) on risk aversion, since a real estate bubble usually precedes higher risk aversion. The S&P500 P/E ratio is the first variable that negatively impacts the risk aversion (0.110),thus a higher P/E lowers the risk aversion. This makes sense, because the P/E ratio is a reflection of the market's perception of future growth in earnings and the risk thereof. An increase in the P/E ratio should be accompanied with higher future growth or less risk, which are both likely to impact the risk aversion negatively. The growth rate of industrial production positively impacts the risk aversion (-0.069), suggesting that higher growth rates lead to less risk aversion. The Producer Price Index (PPI) impacts the risk aversion negatively (0.116), this shows that higher selling prices lowers the risk aversion. Lastly, payroll employment positively impacts risk aversion (-0.046) but only marginally.

Figure 2 plots the risk aversion computed from the time-varying volatility risk premium for two values of the multiplication constant, ϕ. We include a recession indicator, because the time variation in the volatility risk premium is a consequence of macro-economic state variables. These variables have a closer link to economic recessions than bear markets. This Figure clearly highlights a sharp drop in risk aversion leading up to the dot-com bubble around 2001, while the risk aversion after the recession goes up. The years leading up to the dot-com bubble report a negative risk aversion parameter, implying that the average investor is risk seeking. This risk seeking behavior leads to the possibility of implied expected returns in (3) becoming negative. Wheale and Amin [2003] find evidence of a higher risk tolerance for investors in this period.

The financial crisis in 2008 shows a sharp increase after and during the recession, similar to the dot-com bubble. This is in line with the findings of Guiso et al. [2013], who report an increase in investor risk aversion shortly after the financial crisis of 2008. Figure 2 shows that a lower multiplication constant ϕ leads to more volatility in the risk aversion, which might be more beneficial when using these values in forecasting the returns.

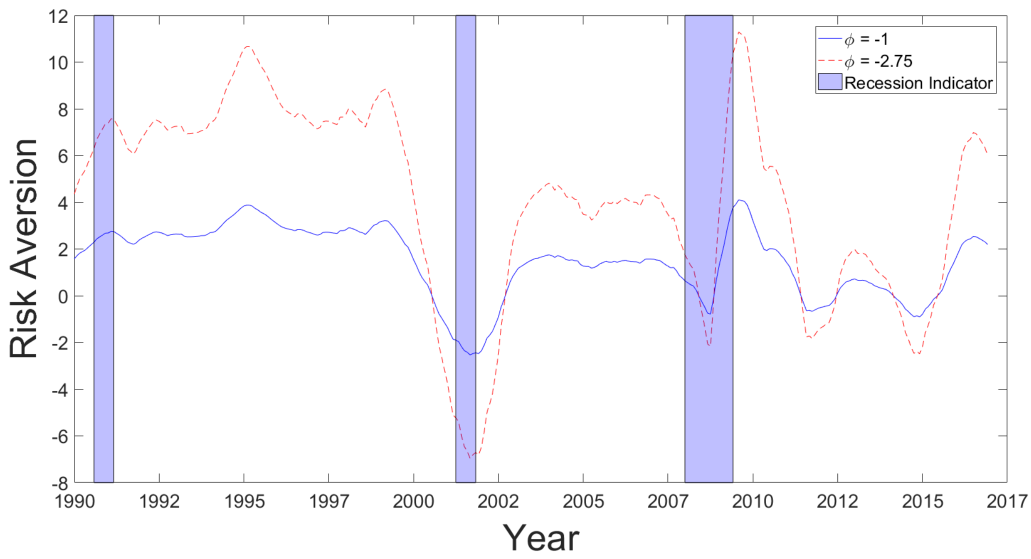


Figure 2: Evolution of Time-Varying Risk Aversion for two Multiplication Constants

Notes: This figure plots the time-varying risk aversion between January 1990 and March 2017 for two multiplication constants. The recession indicator is the NBER recession indicator for the United States for the period following the peak until the trough.

As mentioned before, we search for a multiplication constant ϕ that minimizes the RMSFE over past observations. In Figure 3 we show the evolution of the RMSFE for several multiplication constants which we choose ex-post. These multiplication constants imply values for the (average) time-varying risk aversion and are reported on the x-axis. Figure 3 shows that implied expected returns with an average time-varying risk aversion between one and seven improve the forecast accuracy compared to the benchmark case. We obtain the lowest RMSFE with the average risk aversion at 4.08. The multiplication constant ϕ then equals the average risk aversion divided by the average time-varying volatility risk premium, $4.08/-1.48 = -2.75$. Figure 3 shows that for most reasonable values of average time-varying risk aversion, the RMSFE improves compared to the benchmark case. Additionally, it shows that there is a smooth relation between the RMSFE and the average risk aversion and by extension the multiplication constant. The forecasting performance of implied expected returns with a time-varying volatility risk premium is therefore not overly sensitive to the choice of the multiplication constant.

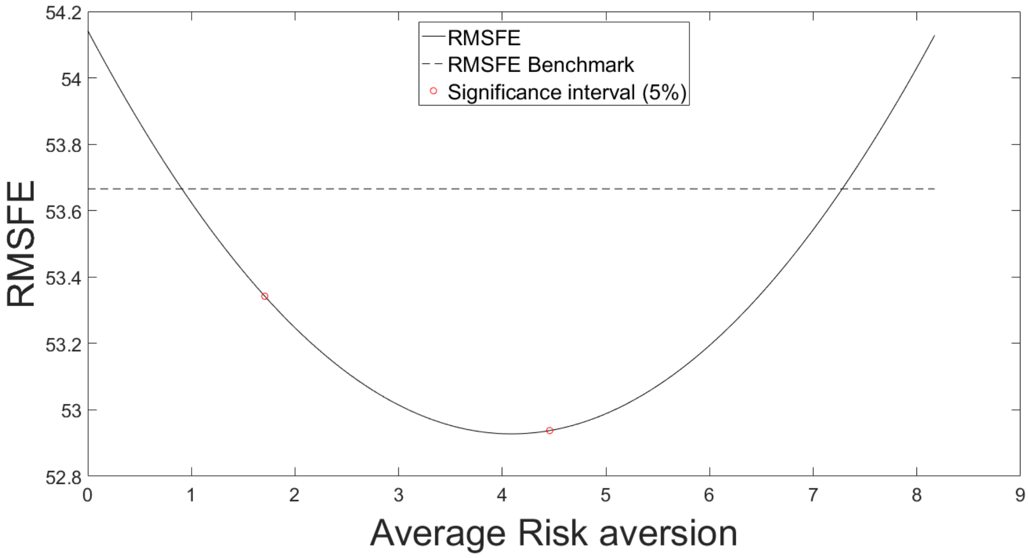
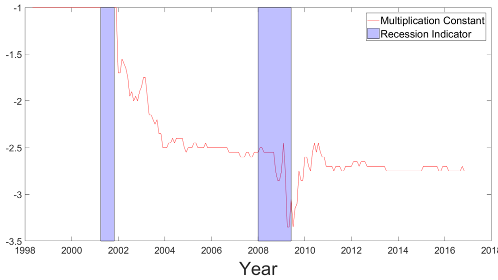


Figure 3: Evolution of RMSFE for Diferent Average Risk Aversions Implied by the Time-Varying Volatility Risk Premium

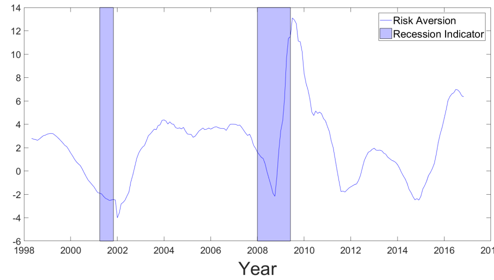
Notes: This Figure plots the RMSFE as a function of the average risk aversion that is implied by the time-varying volatility risk premium for various multiplication constants ϕ. RMSFE (in per mille) is the average RMSFE over the ten industries and the five portfolio proxies considered in this paper. The benchmark refers to the standard implied expected return approach with risk aversion parameter $\gamma = 2.4$. The interval of significance denotes the interval for which the implied expected returns with time-varying risk aversion significantly forecast better than the benchmark case, based on a Diebold-Mariano test with 5% significance level.

To obtain out-of-sample forecasts we now choose the multiplication constant ϕ that minimizes the RMSFE over past forecasts and we re-estimate the (time-varying) volatility risk premium at each point in time, with an expanding window. For the first three years of our sample we set ϕ to -1, as in Bollerslev et al. [2011], to avoid look-ahead bias. The re-estimation of the multiplication constant and volatility risk premium brings this approach of determining the risk aversion closer to the approach in which the risk aversion is determined by matching the data in past periods. However, if the volatility risk premium is time-varying, it is the main component of time-variation in the risk aversion. In this case, the multiplication constant mainly determines scaling, this makes the evolution of the risk aversion more smooth comparatively. On the other hand, if the volatility risk premium is constant, then the multiplication constant is the main source of time-variation in the risk aversion. These conclusion are verified by Figure 4. It plots the evolution of the multiplication constant and resulting risk aversion for both volatility risk premium specifications. Henceforward, we refer to the risk aversion computed from a constant volatility risk premium, as constant risk aversion.

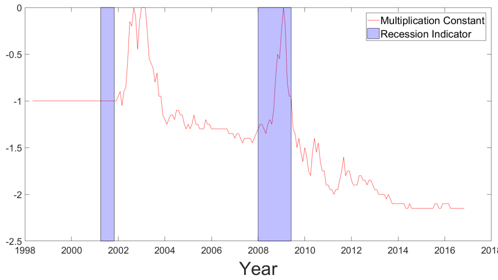
Figure 4: Evolution of Risk Aversion and Multiplication Constant for the Constant and Time-Varying Volatility Risk Premium Specification



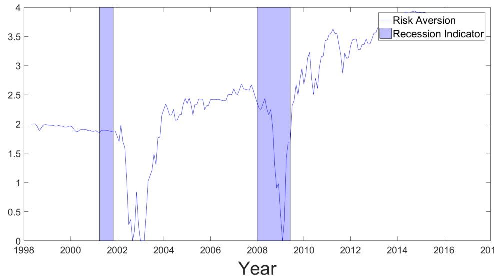
(a) ϕ in Time-varying Specification



(b) Risk Aversion in Time-varying Specification



(c) ϕ in Constant Specification



(d) Risk Aversion in Constant Specification

Notes: Figures 4b and 4a report the multiplication constant and the resulting risk aversion for the time-varying volatility risk premium specification. Figures 4d and 4c report the multiplication constant and the resulting risk aversion for the constant volatility risk premium specification.

Table 5: Forecast Evaluation for Expected Return Forecasts with (Time-Varying) Risk Aversion Estimator Unconditional on the State of the Equity Market

	RMSFE(‰)	**MPPE(%)**	**HR(%)**	$\boldsymbol{\sigma_1}$	$\boldsymbol{\sigma_2}$
Panel A: Constant Risk Aversion					
$\boldsymbol{w_{ew}}$	53.77 (0.759)	-0.19	59.37	0.31	0.17
$\boldsymbol{w_{mkt}}$	53.79 (0.748)	-0.17	59.42	0.30	0.16
$\boldsymbol{w_{iv}}$	53.77 (0.782)	-0.10	59.42	0.25	0.10
$\boldsymbol{w_{erc}}$	53.75 (0.787)	-0.11	59.42	0.26	0.11
$\boldsymbol{w_{md}}$	53.76 (0.833)	-0.04	59.42	0.21	0.08
Panel B: Time-Varying Risk Aversion					
$\boldsymbol{w_{ew}}$	53.07 (0.135)	-0.28	58.88	1.08	0.29
$\boldsymbol{w_{mkt}}$	53.04 (0.098)	-0.25	58.83	1.01	0.26
$\boldsymbol{w_{iv}}$	53.06 (0.062)	-0.18	58.65	0.83	0.17
$\boldsymbol{w_{erc}}$	53.06 (0.089)	-0.21	58.39	0.88	0.19
$\boldsymbol{w_{md}}$	53.15 (0.044)	-0.12	58.03	0.66	0.14

Notes: This table reports forecast evaluation statistics for the constant and time-varying risk aversion specifications unconditional on the state of the equity market. HR is hit ratio in percent. RMSFE: is root-mean-squared forecast error, in per mille with p-values of Diebold-Mariano test in parentheses. This test check if the forecasts in this table outperform the corresponding forecasts (based on portfolio proxy) in the benchmark case in Table 2. $\boldsymbol{\sigma_1}$ is instability over time, in percent and averaged over the ten industries. $\boldsymbol{\sigma_2}$ is dispersion across assets, in percent and averaged over time. The evaluation sample is between June 1998 and December 2016.

Table 5 reports the forecasting performance of implied expected returns with a constant and time-varying risk aversion estimate. This table shows that the constant risk aversion estimate negatively affects forecasting performance compared to the benchmark case. Although the RMSFE is higher for all five portfolio proxies, the differences are not statistically significant at the 10% level. The MPPE is down from an average of 0.17% in the benchmark case to 0.12% in Panel A of Table 5, but this is not a large improvement. Further, the hit ratio is unaffected by the change to the risk aversion. This shows that there are still no positive forecasts when the risk aversion is based on a constant volatility risk premium. However, there is some additional instability in the forecasts in Panel A of Table 5, compared to the benchmark case. This is due to the re-estimation of the multiplication constant ϕ, which introduces some time variation in the estimate, as shown in Figure 4d. The level of instability is still lower than the time-series based forecasts, but it makes an application to portfolio optimization more difficult.

On the other hand, the time-varying risk aversion estimate improves the forecast accuracy compared to the benchmark. It drops the average RMSFE from 53.66‰ to 53.08‰. Hence, all five portfolio proxies report a lower RMSFE, as shown in Panel B of Table 5. These differences in forecasting accuracy are statistically significant (10%) for the $\boldsymbol{w_{mkt}}$, $\boldsymbol{w_{iv}}$,$\boldsymbol{w_{erc}}$ and $\boldsymbol{w_{md}}$ portfolios. However, a 1% reduction in the RMSFE is unlikely to be economically significant. Although the $\boldsymbol{w_{md}}$ portfolio improves the least in terms of RMSFE, it reports the lowest p-value for the Diebold-Mariano test. This is possible, because it also reports the lowest instability in Panel B of Table 5. Further, the MPPE is only marginally affected compared to the benchmark. This indicates that the forecasts are not more biased, even though the risk aversion in Figure 4b leads to more time variation in the forecasts. This additional time variation in the risk aversion does lead to an increase in instability. The increase in dispersion follows from the time variation in the risk aversion as well, but the forecasts are still less dispersed than time-series based forecasts.

A downside of the forecasts in Panel B of Table 5, is that the hit ratio is slightly lower than in the benchmark case. The highest hit ratio is 58.85‰ for the equally weighted portfolio, while the hit ratio in the benchmark case is on average 59.42‰. This shows that while the time-varying risk aversion is sometimes negative, as shown in Figure 4b, the negative forecasts resulting from these periods do not positively affect the hit ratio.

In Section A.2 of the Appendix we show the results of the pairwise comparisons between Table 5 and the time-series based forecasts in Table 2. This comparison reveals that all forecasts with constant risk aversion still significantly (5%) outperform the AR(1) and constant mean model. Moreover, the forecasts in Panel B of Table 5 significantly (5%) outperform the three factor model as well. A pairwise comparison between both panels in Table 5 shows that four out of the five forecasts with time-varying risk aversion, significantly (10%) outperform all forecasts with constant risk aversion. The former consist of the $\boldsymbol{w_{mkt}}$, $\boldsymbol{w_{iv}}$,$\boldsymbol{w_{erc}}$ and $\boldsymbol{w_{md}}$ portfolios.

These portfolios are mean-variance efficient under various assumptions, outlined in Section 2.1. If these assumptions are valid, (3) holds and we can conclude that the estimated time-varying risk aversion better captures the representative investor's risk aversion. However, in practice none of the assumptions in Section 2.1 hold. If the portfolios are not mean-variance efficient, (3) does not hold exactly and we can not conclude from the forecasting performance that one risk aversion estimate is better. Lastly, We find that for both panels in Table 5, there is no portfolio proxy that outperforms any of the other portfolio proxies.

Table 6: Forecast Evaluation for Implied Expected Return Forecasts with (Time-Varying) Risk Aversion Estimator Conditional on the State of the Equity Market

(a) Bull Market

	RMSFE(‰)	**MPPE(%)**	**HR(%)**	$\boldsymbol{\sigma_1}$	$\boldsymbol{\sigma_2}$
Panel A: Constant Risk Aversion					
$\boldsymbol{w_{ew}}$	47.19 (0.937)	0.83	65.83	0.33	0.17
$\boldsymbol{w_{mkt}}$	47.23 (0.939)	0.85	65.89	0.31	0.16
$\boldsymbol{w_{iv}}$	47.38 (0.940)	0.92	65.89	0.26	0.10
$\boldsymbol{w_{erc}}$	47.34 (0.940)	0.90	65.89	0.28	0.11
$\boldsymbol{w_{md}}$	47.51 (0.956)	0.97	65.89	0.22	0.09
Panel B: Time-Varying Risk Aversion					
$\boldsymbol{w_{ew}}$	46.85 (0.581)	0.57	61.43	1.13	0.31
$\boldsymbol{w_{mkt}}$	46.82 (0.525)	0.60	61.43	1.06	0.28
$\boldsymbol{w_{iv}}$	46.80 (0.371)	0.71	61.49	0.88	0.20
$\boldsymbol{w_{erc}}$	46.81 (0.426)	0.68	61.49	0.94	0.22
$\boldsymbol{w_{md}}$	46.98 (0.267)	0.80	61.49	0.71	0.16

(b) Bear Market

	RMSFE(‰)	**MPPE(%)**	**HR(%)**	$\boldsymbol{\sigma_1}$	$\boldsymbol{\sigma_2}$
Panel A: Constant Risk Aversion					
$\boldsymbol{w_{ew}}$	72.54 (0.097)	-3.91	35.83	0.21	0.20
$\boldsymbol{w_{mkt}}$	72.50 (0.096)	-3.89	35.83	0.20	0.18
$\boldsymbol{w_{iv}}$	72.11 (0.107)	-3.81	35.83	0.17	0.08
$\boldsymbol{w_{erc}}$	72.14 (0.110)	-3.80	35.83	0.18	0.09
$\boldsymbol{w_{md}}$	71.78 (0.104)	-3.72	35.83	0.16	0.06
Panel B: Time-Varying Risk Aversion					
$\boldsymbol{w_{ew}}$	70.98 (0.005)	-3.35	49.58	0.44	0.20
$\boldsymbol{w_{mkt}}$	70.95 (0.004)	-3.36	49.38	0.42	0.18
$\boldsymbol{w_{iv}}$	71.04 (0.004)	-3.42	48.33	0.33	0.08
$\boldsymbol{w_{erc}}$	71.05 (0.003)	-3.45	47.08	0.33	0.09
$\boldsymbol{w_{md}}$	70.94 (0.002)	-3.47	45.42	0.25	0.05

Notes: This table reports forecast evaluation statistics for the constant and time-varying risk aversion specifications conditional on the state of the equity market. HR is hit ratio in percent. RMSFE: is root-mean-squared forecast error, in per mille with p-values of Diebold-Mariano test in parentheses. This test check if the forecasts in this table outperform the corresponding forecasts (based on portfolio proxy) in the benchmark case in Tables 3a and 3b. $\boldsymbol{\sigma_1}$ is instability over time, in percent and averaged over the ten industries. $\boldsymbol{\sigma_2}$ is dispersion across assets, in percent and averaged over time. The evaluation sample is between June 1998 and December 2016.

Tables 6a and 6b report the forecasting performance of implied expected returns with different risk aversion estimates in bull and bear markets, respectively. The first table shows that a constant risk aversion estimate negatively affects the forecasting performance in bull markets, compared to the benchmark. The average RMSFE in bull markets goes up from 46.97‰ to 47.33‰ and all individual portfolios significantly (10%) underperform their counterparts in the benchmark. Moreover, the MPPE changes only marginally (0.89% versus 0.86%) and the hit ratio is unaffected by the change in risk aversion. Lastly, the instability is slightly higher than in the benchmark (0.28 versus 0.23) and the dispersion reports a very small decrease.

In bear markets on the other hand, a constant risk aversion estimate leads to improvements in forecast accuracy compared to the benchmark. The average RMSFE drops from 72.71‰ to 72.21‰, but differences are not statistically (10%) significant for most proxies. Interestingly, this change is not accompanied by an increase in the hit ratio or a large increase in MPPE. The former remains unchanged and the latter increases from an average of -3.91% in the benchmark to -3.83%. The hit ratio remains the same because the volatility risk premium is negative for each point in the expanding window. This results in a positive risk aversion parameter, since the multiplication constant ϕ is negative. The risk aversion estimate does therefore not impact the sign of the forecast. Lastly, unlike in bull markets, the instability in bear markets remains unchanged on average.

The time-varying risk aversion estimate does not lead to a significant (10%) decrease or increase in RMSFE in bull markets, compared to the benchmark case. The resulting forecasts report an average RMSFE of 46.85‰ in Panel B of Table 6a, down from 46.96‰ in the benchmark case. Moreover, these forecasts note a reduction in MPPE from 0.86% to 0.67%. Forecasts are therefore close to the actual values and are less too low. This is because the average time-varying risk aversion is 2.20, while the average risk aversion is 2.40 in the benchmark case. The forecasts with a time-varying risk aversion estimate are therefore on average lower than in the benchmark. As a consequence, the average hit ratio is also reduced to 61.47% in Panel B of Table 6a.

The results for the time-varying risk aversion in bear markets, show the benefits of employing this estimate. The forecasts, shown in Panel B of Table 6b, note a reduction in average RMSFE from 72.71‰ in the benchmark to 70.99‰. Increases in forecasting accuracy are statistically significant at the 1% level for all portfolio proxies. Moreover, these forecasts improve the MPPE and the hit ratio compared to the benchmark. The former improves from -3.91% to -3.83% while the latter improves from 35.83% to 47.96%. Interestingly, the large increase in hit ratio does not lead to a similarly large reduction in MPPE. This indicates that the forecasts correctly predict the sign of a negative return more often, but still incorrectly predict the magnitude. Although the performance in bear markets is improved, it is still substantially lower than in bull markets. This shows that the time-varying risk aversion only alleviates the problem of poor performance of implied expected returns in bear markets.

We show the results of a pairwise comparison between the forecasts in Table 6 and the time-series based forecasts in Table 3 in Section A.2 in the Appendix. These results indicate that for the constant risk aversion forecasts in bull markets, the equally weighted portfolio outperforms all other portfolio proxies. In bear markets, the maximum diversification portfolio outperforms all other proxies. These results are in line with the pairwise comparisons in the benchmark. On the other hand, there is no specific portfolio proxy that outperforms all other proxies for the time-varying risk aversion forecasts. This holds for both bull and bear markets. We now compare the forecasts in Table 6 to time-series based forecasts. This comparison reveals that in bull markets, both the constant and time-varying risk aversion estimates lead to significant (10%) outperformance of the time-series based forecasts. However, in bear markets the three factor model significantly (1%) outperforms the forecasts based on the constant risk aversion. On the other hand, the forecasts based on the time-varying risk aversion estimate are not significantly (10%) outperformed by the three factor model.

The static comparison of the forecast accuracy in the tables in this section hide the time variation in the implied return forecasts. Figure 5 displays the time series of monthly implied return forecasts for the constant and time-varying volatility risk premium specifications and the benchmark case. It is interesting to see that the three methods displayed disagree the most after the financial crisis of 2008. The increase in risk aversion as shown in Figure 2leads to higher forecasts for the implied returns with time-varying risk aversion. The other two methods do not have such increases in risk aversion and therefore do not display higher return forecasts.

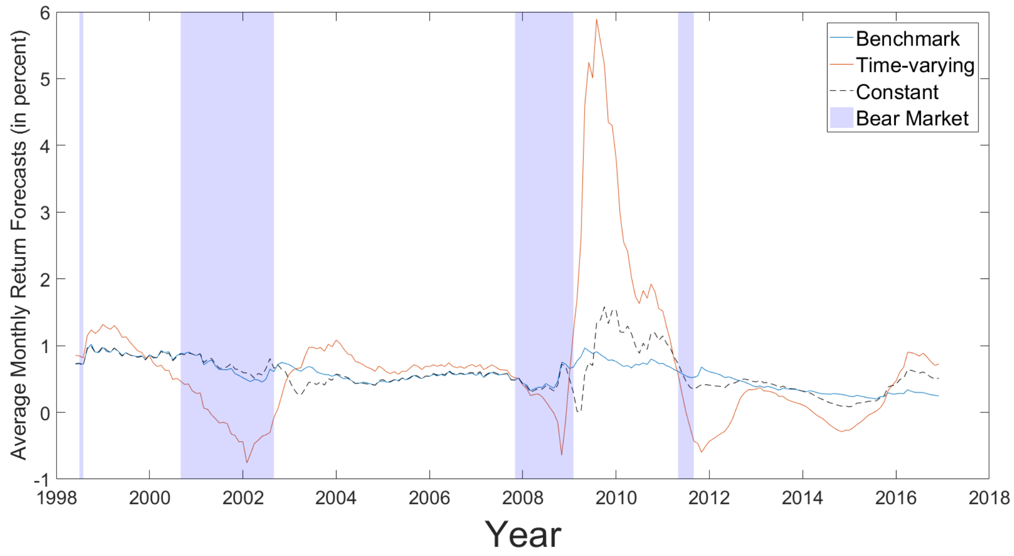


Figure 5: Evolution of Implied Return Forecasts for Different Risk Aversion Specifications and the Benchmark

Notes: This figure shows three different implied expected return forecasts. Time-varying line refers to forecasts based on the time-varying risk aversion. Constant refers to the forecasts based on the constant risk aversion. Returns are presented for the equally weighted portfolio and averaged over all ten industry portfolios, for all methods in this graph. Bear markets are classified by the algorithm of Lunde and Timmermann [2004].

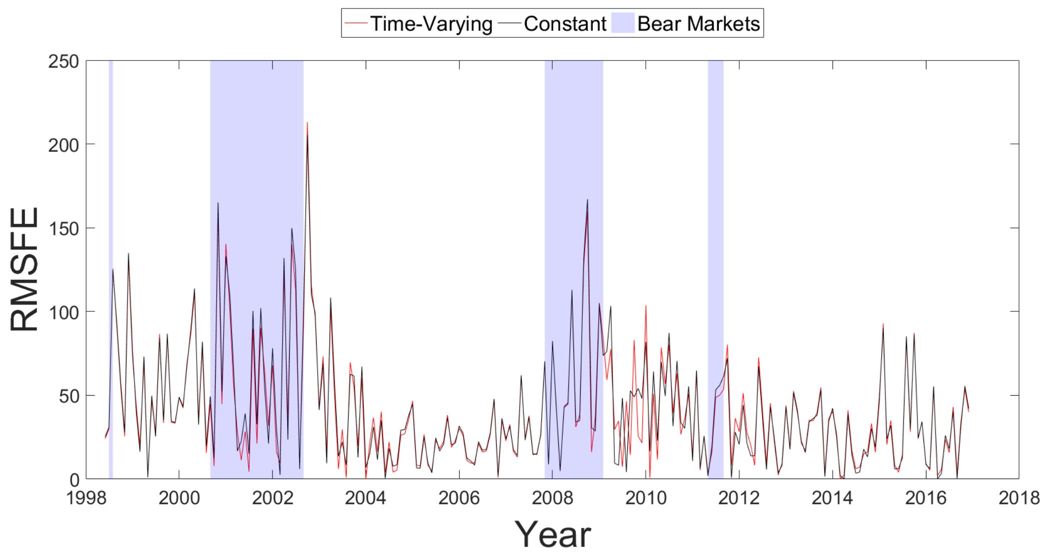


Figure 6: Evolution of RMSFE for Forecasts with Different Risk Aversion Specifications

Notes: This figure shows the RMSFE for implied expected returns with both risk aversion specifications. RMSFE are computed at each point in time over that particular forecast and are therefore the same as the absolute forecast error. Time-varying line refers to forecasts based on the time-varying risk aversion. Constant refers to the forecasts based on the constant risk aversion. Results are presented for the equally weighted portfolio and averaged over all ten portfolios. Bear markets are classified by the algorithm of Lunde and Timmermann [2004].

We now show the time variation in the RMSFE, distinguished between bull and bear states, in Figure

6. We choose not to include the realized values of the returns in either Figure 5 or 6, because they are a great deal more instable than our forecasts[8]. This makes it hard to compare the forecasts to the realized values by means of a graph. Since both forecast methods are relatively close most of the time, they also report similar values of the RMSFE. However, the forecasting methods disagree during the bear market between 2001 and 2003. This leads to periods, in which the time-varying risk aversion lowers the RMSFE in bear markets. The difference between the two method differs the most, shortly after the financial crisis. The forecasts based on time-varying risk aversion are clearly more volatile during this period. This is because the observations from the financial crisis are now incorporated in the estimation. However, it is unclear which method performs best. To examine this, we compute the average RMSFE for this period separately and we present them in Table 7. The results show that even though the time-varying forecasts are much larger, they report a lower RMSFE during this period. More stable forecasts, such as the forecasts in the benchmark and with constant risk aversion are therefore worse during this period.

Table 7: RMSFE between March 2009 and April 2011

	Benchmark	Constant	Time-varying
RMSFE(‰)	53.67	53.79	53.08

Notes: Constant refers to forecasts resulting from the constant risk aversion estimate. Time-varying refers to forecasts resulting from the time-varying risk aversion estimate. RMSFE is averaged over ten industries and five portfolio proxies.

Summarizing, we firstly find that the constant risk aversion estimate positively impacts forecasting performance in bear markets, but it comes at a cost of negatively impacting the performance in bull markets. Since the equity market is more frequently in bull states, this results in lower forecasting accuracy over the whole sample. The increase in performance in bear markets likely follows from the periods of low risk aversion, shown in Figure 4d. Secondly, we find that the time-varying risk aversion estimate not only improves forecasting performance in bear markets, but it also marginally improves the performance in bull markets. This leads to the implied expected returns outperforming their counterparts in the benchmark. We therefore prefer the time-varying risk aversion to the constant risk aversion and $\gamma = 2.4$, in both bull and bear markets.

4.3 Forecast Combination Approaches

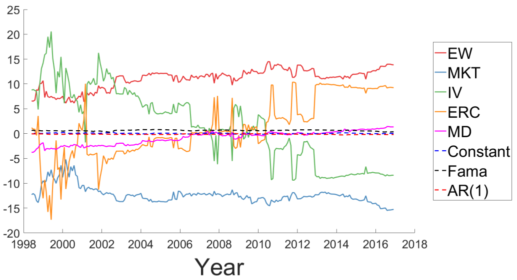
Next, we investigate the impact of introducing dynamic forecast combinations with Markov Switching models as described in Section 2.3 and 2.4. We study the impact of five types of forecast combination models over: (i) all five implied return forecasts, (ii) all five implied return forecasts and all time-series based forecasts and (iii) risk-based implied return forecasts. The latter consist of the w_{iv}, w_{erc} and w_{md} portfolios.

Throughout this section, Model 1 refers to the switching regression forecast combination in (11), Model 2 refers to the constant switching regression forecast combination in (13), Model 3 refers to the switching regression model based on volatility in (15) , Model 4 refers to the smooth transition regression model in (14) and Model 5 refers to the simple regression model in (16). We obtain results by employing a Markov switching model with constant transition probabilities for identifying and forecasting the bull and bear markets. Results for the Markov switching model are presented in Figure 10a. We evaluate alternative procedures for identifying and forecasting the state of the equity market in our sensitivity

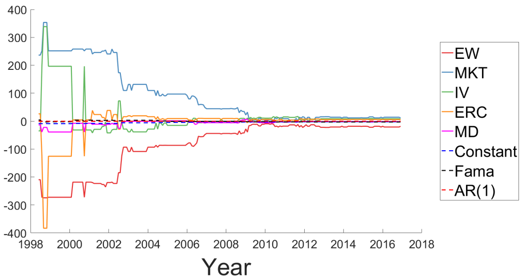
[8]The benchmark is not included because it is very similar to the Constant risk aversion specification.

study. Forecast combination weights are estimated with a moving window of ten years for bull market observations and an expanding window for bear market observations. We choose different estimation windows, because there are periods with very few bear markets, which makes a moving window not suitable for estimating the forecast combination weights in bear markets. The sample over which the forecast combination weights of Models 1,2 and 5 are estimated starts on January 1970. Because the VIX is only available from January 1990, the estimation sample for Models 3 and 4 starts later.

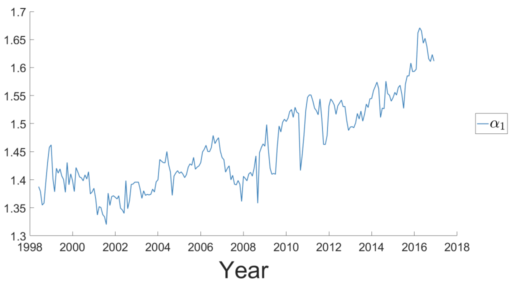
(a) Model 1 Bull



(b) Model 1 Bear



(c) Model 2 Bull



(d) Model 2 Bear

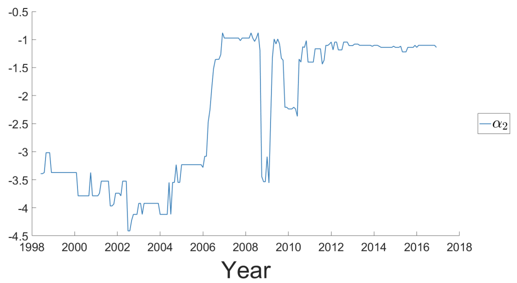


Figure 7: Forecast Combination Weights

Notes: This Figure shows the forecast combination weights for Model 1 and 2 in bull and bear markets. Dashed lines represent time-series based forecasts. First five entries in the legend stand for implied expected returns for various portfolio proxies. EW: equally weighted portfolio. MKT: market capitalization portfolio. IV: inverse volatility portfolio. ERC: equal risk contribution portfolio. MD: maximum diversification portfolio. Constant: Constant mean model. Fama: three factor model.

We first discuss the weights of the forecast combinations of Model 1 and 2 over all return forecasts in Figure 7. The results for the other models are presented in Section A.3 in the Appendix. Figure 7a shows that implied expected return forecasts get much higher weights than time-series based forecasts. However, this is likely due to the high correlation present between the different implied expected return forecasts. Ardia and Boudt [2015] find that the correlation between various implied expected return forecasts is between 0.85 and 0.95 in their sample and we obtain similar values. This leads to multicollinearity in the estimation of the forecast combination weights, which can be clearly seen in Figures 7a and 7b. We therefore avoid making inferences on the individual forecast combination weights. However, the multicollinearity issues does not affect the overall fit of the model in (12) and since we incorporate more information by means of the Markov switching models we can still expect an increase in forecasting accuracy. Model 5, which does not incorporate the state of the equity market is unlikely to yield an improvement in forecasting accuracy.

The relation between the forecast combination weights of implied expected returns is also evidence of high correlation between the forecasts. In both Figures 7a and 7b, EW weights move in the opposite direction of MKT weights, whereas IV weights move in the opposite direction of ERC weights. Summing the weights of the five implied return forecasts in Figure 7a returns an average of 0.48, while the time-series based forecasts return an average of 0.53. This shows that although the time-series based forecasts get a relatively low weight in the forecast combination, they still contribute to it.

The scaling factor we give to the average forecast in Model 2 shows less time variation in Figures 7c and 7d. However, there is a big difference in interpretation of the scaling factors in bull and bear markets. The former is on average -2.29, whereas the latter is 1.45. This shows that the average forecast in bull markets should be slightly higher according to Model 2. On the other hand, the scaling factor in bear markets shows that the average forecast in bear markets is way too high and should be lowered more.

Table 8: Forecast Evaluation for Expected Return Forecasts with Dynamic Forecast Combinations Unconditional on the State of the Equity Market

	RMSFE	**MPPE(%)**	**HR(%)**	$\boldsymbol{\sigma_1}$	$\boldsymbol{\sigma_2}$
Panel A: Forecast Combination of five Implied Return Forecasts					
Model 1	51.21 (0.021)	0.39	63.54	1.53	0.65
Model 2	51.97 (0.070)	0.18	61.43	0.94	0.14
Model 3	54.76 (0.975)	−0.65	58.65	1.20	0.80
Model 4	54.21 (0.614)	0.45	50.31	2.49	0.81
Model 5	53.72 (0.611)	−0.45	59.55	0.31	0.16
Panel B: Forecast Combination of all Return Forecasts					
Model 1	51.23 (0.033)	0.44	60.67	1.47	0.65
Model 2	52.28 (0.114)	0.10	60.94	1.05	0.19
Model 3	55.33 (0.993)	−0.45	56.32	1.29	0.81
Model 4	54.58 (0.691)	0.48	50.90	2.19	0.73
Model 5	53.73 (0.561)	−0.47	59.42	0.36	0.18
Panel C: Forecast Combination of Risk-Based Forecasts					
Model 1	51.99 (0.049)	0.29	62.51	1.35	0.40
Model 2	51.97 (0.063)	0.18	61.57	0.95	0.10
Model 3	54.47 (0.997)	−0.48	58.79	0.50	0.23
Model 4	52.09 (0.132)	−0.14	57.76	1.85	0.60
Model 5	53.69 (0.587)	−0.46	59.19	0.27	0.14

Notes: This table reports forecast evaluation statistics over four types of forecast combination models and for four sets of forecasts on which these forecast combinations are based. RMSFE is root-mean-squared forecast error, in per mille with p-values of Diebold-Mariano test in parentheses. This test checks whether forecasts in this table outperform the equally weighted portfolio in the benchmark. HR is hit ratio in percent. We choose to compare to the equally weighted portfolio, because it is the best performing forecast in the benchmark.MPPE is mean percentage point error. $\boldsymbol{\sigma_1}$ is instability over time, in percent and averaged over the ten industries. $\boldsymbol{\sigma_2}$ is dispersion across assets, in percent and averaged over time. The evaluation sample is between June 1998 and December 2016.

We present the forecast evaluation of the various forecast combination approaches unconditional on the state of the equity market in Table 8. It clearly highlights the benefits of incorporating the state of the equity market and forecast combinations in implied expected return forecasts, although not all approaches perform equally well. The best performing forecast combination approach is Model 1 in Panel A. It reports an RMSFE of 51.21‰ which is significantly (5%) lower than in the benchmark case. The hit ratio is also increased from 59.42% to 63.54%. Forecast combinations can therefore better predict the sign of the monthly return. Further, the MPPE increases from -0.24% in the benchmark to 0.39% in Panel A. Thus, forecasts are now on average 0.24 percentage points too high, instead of 0.39 percentage points too low. We compare all models in Table 8 pairwise in Section A.3 in the Appendix. A comparison of Model 1 and Model 2, shows that the RMSFE of Model 1 is significantly (5%) lower than the RMSFE of Model 2 in Panels A and B, while Model 2 only slightly outperforms Model 1 in Panel C. This suggests that estimating the forecast combination weights with a switching linear regression model is better than only estimating a constant, as in Model 2.

Adding time-series based forecasts to the forecast combination does not improve the RMSFE over the whole sample. This holds true for all five models, although Model 1 and 5 only note a small increase in RMSFE. However, Model 1 also reports a decrease in the hit ratio from 63.54% to 60.67%. Further, the MPPE increases slightly from 0.39% to 0.44%. This shows that forecasts are too low on average, while still incorrectly predicting the sign more often. This is a consequence of the model incorrectly predicting negative returns.

Furthermore, the forecasting performance of Model 2 becomes worse when time-series based forecasts are included. This makes sense, because time-series based return forecasts are able to predict negative returns and their bad performance in bear markets is thus not necessarily due to the sign of the forecast. Moreover, implied expected return forecasts are more stable than time-series based forecasts, as shown in Table 2. If we then group implied return and time-series based forecasts together, the average forecast has less regime dependent shortcomings while it also performs worse overall due to the time-series based forecasts. This reduces the effectiveness of Model 2.

Even in this scenario, Model 2 still improves the forecast accuracy compared to the benchmark. This shows that estimating a vector of forecast combination weights is not necessary. Simply estimating a constant to multiply the average forecast with dependent on the regime is enough to improve forecasting accuracy. This can be seen by the RMSFE of Model 2 in the different Panels of Table 8. Because they are all lower than the benchmark case, although not all statistically significant at the 10% level. Over the whole sample, Model 2 reports an RMSFE of 51.97‰ , 52.28‰ and 51.97‰ in Panels A, B and C, respectively. A benefit of Model 2 is that the forecasts are less instable and much less dispersed than the forecasts of Model 1, this makes sense because the estimation of additional parameters in the vector of forecast combination weights in Model 1 lead to additional instability in the forecast combination compared to Model 2. In Panel C, Model 2 even outperforms Model 1, but only slightly.

In both Panel A and B, Model 3 and 4, perform worse than the other forecast combination models and the benchmark case. The smooth transition model performs somewhat better than the switching regression model (54.72‰ vs. 54.21‰), but it is also more volatile and the difference in forecasting performance is not statistically significant at the 10% level. However, if we only incorporate risk-based implied returns, then the smooth transition approach has a RMSFE of 52.09% down from 54.21% in Panel A. Introducing time-series based forecasts does not achieve the same effect and increases the RMSFE to 54.58%. Therefore, Model 4 can improve the forecasting accuracy compared to the benchmark case when only risk-based portfolio proxies are considered. Additionally, the instability and dispersion of the forecast combination is lowest in Panel C. This shows that a smooth transition forecast combination approach based on volatility shows the best results when forecasts are based on volatility as well. However, even then does this approach not lead to an improvement compared to Model 1 and 2, while it is still more unstable and dispersed.

On the other hand, Model 3 shows no such increase in forecasting accuracy when only considering risk-based portfolio proxies. It is the worst performing forecast combination model in all Panels of Table 8. The only upside to Model 3, is that is is less unstable than the smooth transition alternative, especially so in Panel C (0.50 vs 1.85). The poor performance of Model 3 shows that grouping forecasts together based on volatility is not effective. This makes sense, because although bear markets are usually characterized as periods with high volatility, bull markets can also show high volatility. The different mean returns in these two states then make it difficult to properly combine the forecasts. These results indicate that the forecast combination approaches of (14) and (15) that only incorporate volatility are inferior to the forecast combination approaches of (11) and (13) that do not.

In terms of instability and dispersion, all methods and different forecast combinations in Table 8 suffer from the introduction of Markov Switching models and the dynamic forecast combination approach. However, this increase in instability and dispersion seems worthwhile for some models when we look at the decrease in RMSFE compared to the benchmark case. Model 4 reports the highest instability in all Panels of Table 8. This is likely due to the re-estimation of the slope parameter, which affects the forecast combination estimates as well.

To distinguish the effect that accounting for the state of the equity market has on the forecasting performance, we compare Model 1 and 2, to Model 5. This comparison reveals that, Models 1 and 2 report an average RMSFE in Table 8, of 51.48‰ and 52.07‰, respectively. On the other hand, Model 5 reports an average RMSFE of 53.71‰. In all Panels of Table 8, Model 1 and 2 significantly (10%) outperform Model 5. A comparison to the benchmark, reveals that Models 1 and 2 reduce the RMSFE, whereas Model 5 has approximately the same RMSFE as the benchmark. These results show that accounting for the state of the equity market is the main source of improvement in the forecast combinations of Model 1 and 2 in Table 8.

Table 9: Forecast Evaluation for Expected Return Forecasts with Dynamic Forecast Combinations Conditional on the State of the Equity Market

(a) Bull Market

	RMSFE	**MPPE(%)**	**HR(%)**	$\boldsymbol{\sigma_1}$	$\boldsymbol{\sigma_2}$
Panel A: Forecast Combination of five Implied Return Forecasts					
Model 1	47.40 (0.922)	1.01	63.66	0.71	0.46
Model 2	48.53 (0.987)	0.98	62.46	0.61	0.14
Model 3	47.16 (0.849)	0.30	63.83	1.08	0.68
Model 4	51.00 (0.997)	1.28	48.97	2.29	0.84
Model 5	46.83 (0.467)	0.56	65.77	0.27	0.13
Panel B: Forecast Combination of all Return Forecasts					
Model 1	48.24 (0.986)	1.18	60.40	0.93	0.55
Model 2	48.86 (0.989)	0.88	61.54	0.73	0.19
Model 3	48.65 (0.997)	0.53	60.74	1.23	0.73
Model 4	50.89 (0.997)	1.42	48.91	2.20	0.78
Model 5	46.88 (0.450)	0.53	65.31	0.31	0.14
Panel C: Forecast Combination of Risk-Based Forecasts					
Model 1	48.06 (0.987)	0.97	62.69	0.55	0.21
Model 2	48.56 (0.987)	0.97	62.46	0.61	0.11
Model 3	47.21 (0.967)	0.54	65.31	0.34	0.17
Model 4	48.06 (0.989)	0.58	59.83	1.23	0.55
Model 5	46.71 (0.298)	0.55	65.77	0.19	0.10

(b) Bear Market

	RMSFE	**MPPE(%)**	**HR(%)**	$\boldsymbol{\sigma_1}$	$\boldsymbol{\sigma_2}$
Panel A: Forecast Combination of five Implied Return Forecasts					
Model 1	62.31 (0.002)	-1.85	63.13	2.54	1.33
Model 2	62.41 (0.000)	-2.74	57.71	1.48	0.14
Model 3	75.66 (0.970)	-4.09	39.79	1.31	1.23
Model 4	63.98 (0.028)	-2.60	55.21	2.91	0.74
Model 5	73.25 (0.698)	-4.13	36.88	0.43	0.30
Panel B: Forecast Combination of all Return Forecasts					
Model 1	60.10 (0.001)	-2.25	61.67	2.41	1.03
Model 2	62.71 (0.000)	-2.74	58.75	1.54	0.21
Model 3	74.20 (0.646)	-3.99	40.21	1.35	1.10
Model 4	65.79 (0.041)	-2.91	58.13	2.05	0.57
Model 5	73.18 (0.638)	-4.12	37.92	0.48	0.34
Panel C: Forecast Combination of Risk-Based Forecasts					
Model 1	63.51 (0.002)	-2.18	61.87	2.32	1.07
Model 2	62.35 (0.000)	-2.72	58.33	1.48	0.08
Model 3	74.71 (0.984)	-4.19	35.00	0.79	0.44
Model 4	64.11 (0.032)	-2.76	50.21	2.96	0.76
Model 5	73.41 (0.804)	-4.15	35.21	0.41	0.27

Notes: This table reports forecast evaluation statistics over four types of forecast combination models and for four sets of forecasts on which these forecast combinations are based. RMSFE is root-mean-squared forecast error, in per mille with p-values of Diebold-Mariano test in parentheses. This test checks whether forecasts in this table outperform the equally weighted portfolio in the benchmark. We choose to compare to the equally weighted portfolio, because it is the best performing forecast in the benchmark. HR is hit ratio in percent. MPPE is mean percentage point error. $\boldsymbol{\sigma_1}$ is instability over time, in percent and averaged over the ten industries. $\boldsymbol{\sigma_2}$ is dispersion across assets, in percent and averaged over time. The evaluation sample is between June 1998 and December 2016.

We now compare the forecasting accuracy of the forecast combinations between bull and bear markets. We present these results in Table 9 and provide p-values of all pairwise comparisons in Section A.3 in the Appendix. The first table shows that combining implied expected return forecasts according to Model 1 leads to a significant (1%) reduction of the RMSFE in bear markets to 62.31‰ , compared to 72.07‰ in the benchmark. On the other hand, this forecast combination significantly (10%) increases the RMSFE in bull markets, from 46.74‰ to 47.40‰[9]. The gain in forecasting accuracy over the whole sample is therefore due to the increased performance in bear markets. The forecast combination approach of Model 1 also improves the hit ratio in bear markets from 35.83% in the benchmark to 63.13% in Panel A of Table 9b. Moreover, it improves the MPPE from -4.01% to -1.85%. This forecast combination therefore more often correctly predicts the sign and magnitude of negative returns. This shows that the improved forecasting accuracy in bear markets is a consequence of the forecast combination predicting negative returns.

The reason that forecasting performance in bull markets is negatively affected is because forecasts are slightly lower than in the benchmark. This can be seen by the MPPE (1.01%) and hit ratio (63.66%) in Panel A of Table 9a. The former is higher than in the benchmark, whereas the latter is lower. The reason that forecasts in bull markets are lower, is because the probability of a bear market is frequently between 10% and 20%, as shown in Figure 10a. This results in the forecast combination weights in bear markets lowering the forecast.

Model 2 is one of the simplest forecast combination approaches considered in this paper and aims to remedy the bad performance of implied return forecasts in bear markets. It succeeds in Panel A of Table 9b, since the RMSFE in bear markets is significantly (1%) lower than in the benchmark, at 62.41‰. However, the RMSFE in bull markets takes an even bigger hit than Model 1 and increases to 48.51‰. The MPPE and hit ratio of Model 2 are -2.74% and 57.71% in bear markets, respectively. Because Model 2 improves all forecast precision measures in bear markets relative to the benchmark, it is effective at improving the forecasting performance of implied return forecasts in bear markets. Therefore, multiplying the average forecast by a constant depending on the regime is enough to remedy the poor performance in bear markets.

Although the performance in bear markets of two of the three time-series based forecasts is on-par with implied return forecasts, they do possess additional information for the forecast combination. This can be seen by comparing the performance of Model 1 in Panel B of Table 9b to the performance in Panel A. The inclusion of time-series based forecasts reduces the RMSFE from 62.31‰ to 60.10‰. This makes sense, because the three factor model, which is now included, reports the lowest RMSFE in bear markets in the benchmark. In bull markets on the other hand, the inclusion of time-series based forecasts increases the RMSFE from 47.40‰ to 48.24‰. This is likely because all time-series based forecast underperform all implied expected returns in bull markets. These results indicate that time-series based forecasts should be included in the forecast combination in bear markets, whereas they should not be included in bull markets.

The two volatility based forecast combination models show big differences between their forecasting performance in bull and bear markets. Model 3 increases the RMSFE compared to the benchmark in both bull and bear markets, to 47.16 ‰ and 75.77‰ ,respectively. Model 4 on the other hand, reports an RMSFE of 51.00‰ in bull markets and 63.98‰ in bear markets. This shows that Model 4 is effective at improving the RMSFE in bear markets, whereas Model 3 is not. This is because the bear market forecasts of Model 3 are on average still too high, as shown by the MPPE and hit ratio, which are

[9] This increase is only significant at the 10% level, when comparing to the w_{ew} portfolio.

-4.09% and 39.79%, respectively. Model 4 does improve the MPPE to -2.91% and hit ratio to 55.21%. Consequently, the resulting forecasts are less too high. These results indicate that a threshold model is preferable to a smooth transition transition model in bull markets, while the opposite is true in bear markets. Interestingly, only incorporating risk-based implied returns reduces the RMSFE of Model 4 in bull markets to 48.06‰. On the other hand, the RMSFE in bear markets does not see a similar decrease. This shows the reason risk-based implied returns work better for the smooth transition regression model, is because they improve the forecasting accuracy in bull markets.

All methods and different forecast combinations in Table 9 report higher levels of instability in bear markets than in bull markets. This makes sense, because bear markets usually have higher volatility which results in more time variation in the forecast combination weights and thus in the forecast. Moreover, the instability in either state is also higher than in the benchmark. This is a benefit for Models 1,2 and 4 in bear markets, since the forecasts now more closely match the actual returns. In bull markets, the increased instability does not lead to a closer match to the actual returns for Models 1,2,3 and 4. It is therefore a downside for these models, in bull markets.

Finally, we consider the benefits of incorporating the state of the equity market in the forecast combination for bull and bear markets. We do this by comparing Model 1 and 2 to Model 5. We find that Model 5 significantly (10%) outperforms both Model 1 and 2 in bull markets. Hence, Model 5 reports a lower average RMSFE (46.81‰) than both Model 1 (47.90‰) and 2 (48.65‰) in Table 9a. Model 5 is so successful in bull markets, because the MPPE is 0.56 in Panel A which is lower compared to Model 1. This shows that forecasts are on average higher and closer to the actual values. Forecasts are relatively higher, because the probability of a bear market is not weighing down the forecasts, as in Model 1. This results in Model 5 being the best performing forecast combination in Table 9a. It even reports an RMSFE slightly lower than the benchmark in Panel C (46.71‰). On the other hand, the average RMSFE of Model 5 in bear markets is 73.28‰. This is slightly higher than in the benchmark, but very much higher than the RMSFE of Model 1 (61.97‰). This shows that accounting for the state of the equity market in the forecast combination is the source of very big improvements to the forecast accuracy.

To highlight the benefits of implementing the dynamic forecast combinations, we now examine the evolution over time of the forecast combinations of Model 1 and 2 over the five implied return forecasts. We choose this set of forecasts to show that it is not necessary to include time-series based forecasts to obtain negative forecasts. Figure 8 also highlights the periods which the algorithm of Lunde and Timmermann classifies as bear states. Model 3 and 4 are not included in this figure because of their relative poor performance. Figures of all methods and all forecast combinations models are available in the Appendix.

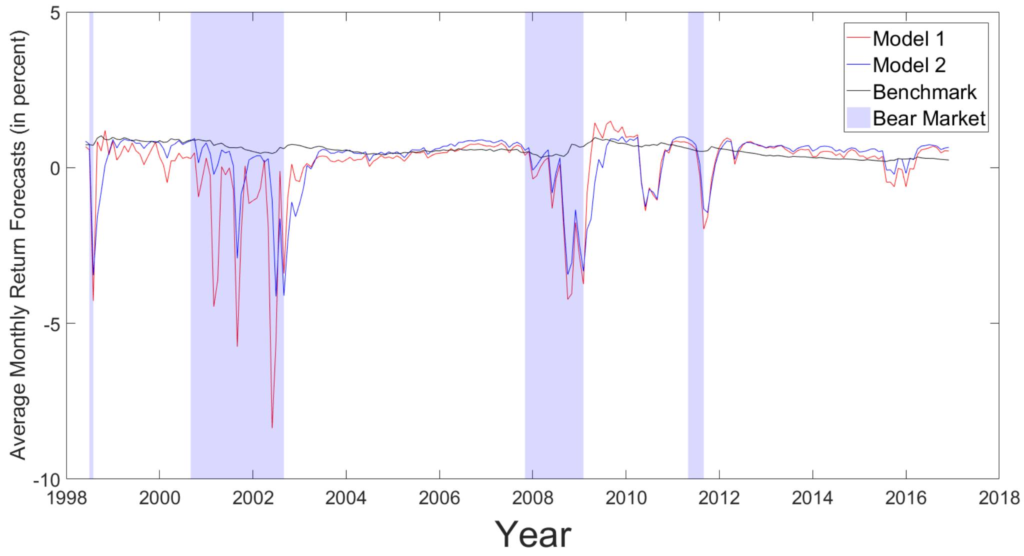


Figure 8: Evolution of Forecast Combinations of Implied Return Forecasts

Notes: This figure shows two different forecast combinations and the implied expected return forecasts in the benchmark. Forecast combinations are made over all five implied return forecasts and the benchmark refers to the w_{ew} portfolio. All forecasts are averaged over the ten industries.

As can be seen in Figure 8 the forecast combinations of both Model 1 and 2 can predict large negative returns. A consequence of being able to predict negative returns is that for large periods of time, the forecast combination of Models 1 and 2 are lower than the benchmark case. This is because in these periods, the uncertainty surrounding the state of the equity market in the next period leads to the forecast combination weights in a bear market getting relatively more weight. Hence, the forecast combinations of Model 1 and 2 are on average more frequently too low when compared to the benchmark. It is interesting to see that negative returns do not always correspond to bear states. This is possible because the Markov switching model can predict a bear market, while the rule-based identification method does not identify a bear state. Such a mismatch is visible around 2010 and shortly after most bear states. The largest negative return is predicted after the dot-com bubble. This can be explained by looking at the bear market probabilities in Figure 10a, since this period notes the highest probability of a bear market in our sample.

We now investigate if the RMSFE exhibits excessive time variation due to the introduction of the forecast combinations. We compare the RMSFE of Model 1 to the benchmark in Figure 9, because the other forecast combination models either show poor performance or are relatively similar to Model 1[10]. This figure shows that the forecast combination of Model 1, leads to additional time variation in the RMSFE in the period between 1998 and 2003. The bear market in this period exhibits more volatility in the RMSFE in the first half, but less so in the last half. In the bear market that starts around 2008, the forecast combination clearly reduces the time variation in the RMSFE. Therefore, in the two longest bear markets, Model 1 exhibits less clustering of the forecast errors than in the benchmark case. This shows that the forecast combination better matches the return data in bear markets and that there is no excessive forecast error clustering in the forecast combination during these periods.

Figure 8 reveals that biggest disagreement between the benchmark and Model 1 are between 1998 and 2003. Consequently, the RMSFE also differs the most in this period. Because it hard to see which

[10]Models 3,4 and 5 show poor performance and Model 2 is relatively similar to Model 1 in Table 8.

forecasting method outperforms the other, we compute the RMSFE for this period separately and report it in Table 10. This table shows that the RMSFE between June 1998 and December 2003 is lower for Model 1, over the whole sample and in bear markets. On the other hand, the benchmark performs better in bull markets. Because a different method is better dependent on the state of the equity market, the time variation in the RMSFE is larger during this period. All things considered, we find that there is no excessive time variation in the forecasting accuracy of the forecast combinations, compared to the benchmark.

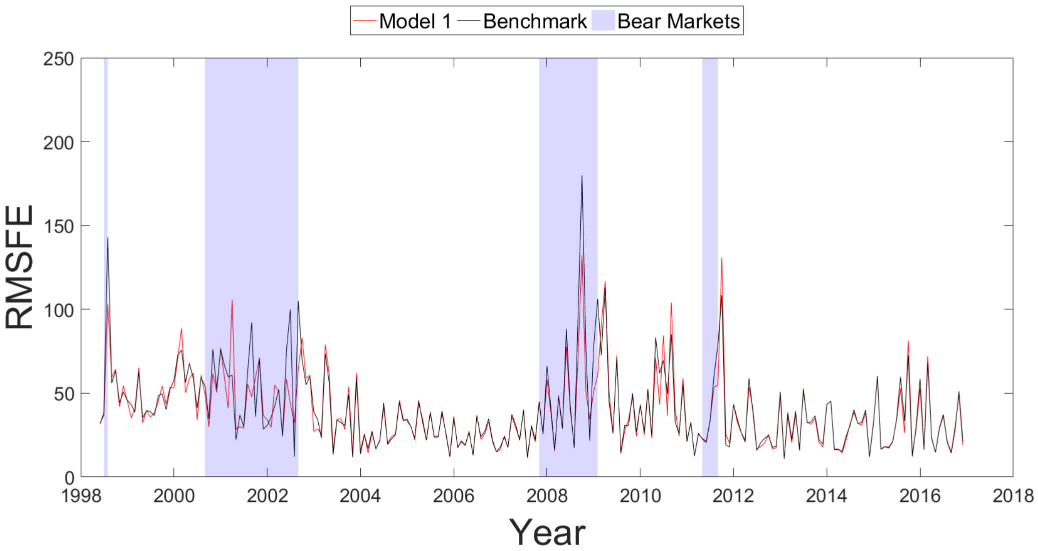


Figure 9: Evolution of RMSFE for Forecast Combination Approach

Notes: This figure shows the RMSFE for Model 1 and the benchmark. The RMSFE is computed at each point in time, over one forecast, and is therefore the same as the absolute forecast error. Forecasts are averaged over all ten portfolios and the benchmark represents forecasts with the w_{ew} portfolio. Bear markets are classified by the algorithm of Lunde and Timmermann [2004].

Table 10: RMSFE between June 1998 and December 2003

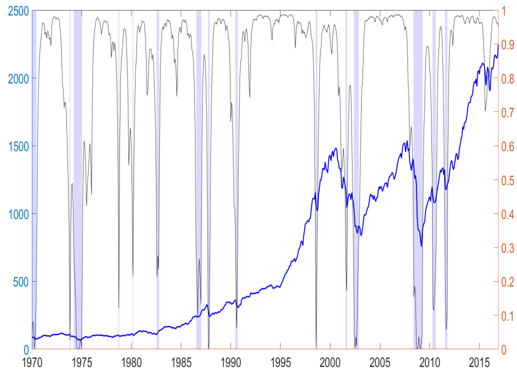
	Model 1	Benchmark
Overall	61.19	63.88
Bull	59.81	58.58
Bear	62.37	70.73

Notes: This Table reports the RMSFE in per mille for Model 1 and the Benchmark, between June 1998 and December 2003. The forecast combination is made over all five implied return forecasts and the benchmark refers to the forecasts with the w_{ew} portfolio. RMSFE is averaged over all ten industries.

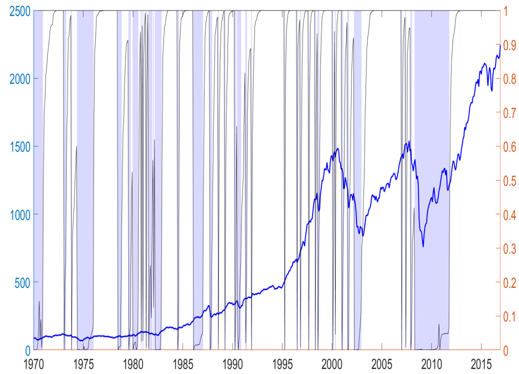
4.4 Sensitivity Study

The results of the forecast combination approach presented in Section 4.3 may be sensitive towards the choice of procedure we use for identifying and forecasting the state of the equity market. In order to investigate this we firstly compare the different identification approaches to each other. Subfigures 10a and 10b report the smoothed probabilities of being in a bull and bear state for Markov switching models with constant and time-varying transition probabilities. Subfigure 10c highlights the periods in which the algorithm of Lunde and Timmermann [2004] classifies the market as being in a bear state. The two

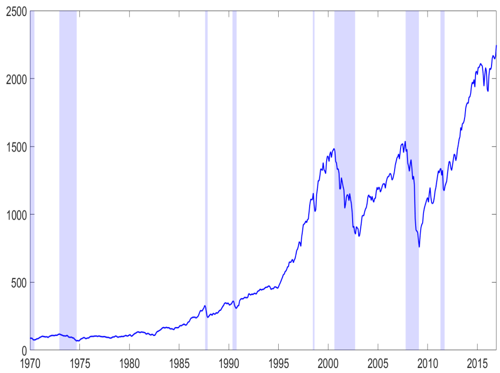
Markov switching models disagree about the probability of being in a bull and bear market for large periods of time. The Markov switching model with time-varying transition probabilities is more volatile in the classification of the regimes and reports more periods with high probability of being in a bear market. The last bear market it reports seems to also capture long periods of positive returns which shows that the model does not always switch rapidly.



(a) Constant Transition Probabilities



(b) Time-Varying Transition Probabilities



(c) Lunde and Timmermann [2004]

Figure 10: Identification of

Notes: The black lines in Subfigures 10a and 10b represent the smoothed probability of being in a bull state. The thick blue line represents the S&P 500 price index and bear markets are indicated by blue areas. Bear markets in Subfigures 10a and 10b are identified when the smoothed probability of a bear market exceeds 70%.

In Table 11 we report the mean and volatility of returns in the bull and bear markets identified by the various methods. It shows that the Markov switching model with time-varying transition probabilities results in the lowest difference between means in bull and bear states (1.12 vs. -0.50). This is a consequence of the higher proportion of bear markets in Figure 10b. The algorithm of Lunde and Timmermann [2004] reports a mean return of 1.64% in bull markets and -2.69% in bear markets, which is the biggest difference. All three identification methods show higher volatility in bear states than in bull states and individual differences are bigger than for mean returns.

Table 11: Sample Moments of Returns in Bull and Bear Markets

Regime		Lunde	Markov	Markovtv
Bull	Mean(%)	1.64	1.30	1.12
	Vol.(%)	3.39	2.90	2.76
Bear	Mean	−2.69	−2.42	−0.50
	Vol(%)	5.13	7.61	6.03

Notes: This Table reports the mean and volatilities (in % per month) of the returns based on the different identification methods in the columns. Bull(bear) states are identified for Markov switching models if the smoothed probability exceeds 70%(60%). Lunde: refers to the algorithm of Lunde and Timmermann [2004]. Markov and Markovtv refer to Markov switching models with constant and time-varying transition probabilities, respectively. In all three cases we condition on the state of the equity market and estimate the mean and volatilities of the returns.

Secondly, we compare the forecasting accuracy for all identification and forecasting models considered. We investigate whether a mis-match between the identification procedure and the forecasting procedure has an impact on the forecasting accuracy of the forecast combination. In Table 12 we report the results for the forecast combination over all implied return forecasts. We use the same identification method for estimating the RMSFE conditional on the regime, as for the estimation of the forecast combination weights. Results conditional on the regime are therefore not directly comparable across different identification method, because the methods classify different periods as bull or bear states. However, for a given identification method, results can be compared across forecasting methods.

Table 12: Root-Mean-Squared Forecast Error of Forecast Combinations for Various Identification and Forecasting Procedures

			Forecasting								
			Lunde			Markov			Markovtv		
			All	Bull	Bear	All	Bull	Bear	All	Bull	Bear
Identification	Lunde	Model 1	71.32	71.26	71.47	53.74	50.04	64.77	60.58	58.82	65.92
		Model 2	65.92	64.54	70.00	52.56	48.39	65.04	55.53	52.32	65.42
	Markov	Model 1	67.44	64.70	84.52	51.21	42.92	77.86	60.98	51.24	88.27
		Model 2	65.20	59.47	89.21	52.44	43.47	83.18	55.99	47.35	86.23
	Markovtv	Model 1	56.53	47.15	71.41	52.16	41.35	68.29	52.22	41.47	68.46
		Model 2	56.07	45.13	72.62	52.87	41.25	69.68	52.96	41.60	69.50

Notes: Lunde refers to the algorithm of Lunde and Timmermann [2004]. Markov refers to a Markov switching model with constant transition probabilities. Markovtv refers to a Markov switching model with time-varying transition probabilities. Bull(bear) bear states are identified by both markov models when the smoothed probability of a (bull)bear state exceeds 70%(60%). All refers to the RMSFE over all observations in the sample. Bull(Bear) refers to the observations classified as being in a bull(bear) state by the corresponding identification method. Forecasting combinations are made over all five implied return forecasts.

When we identify the regime by the algorithm of Lunde and Timmermann [2004] (LT, henceforward), we see that forecasting with a Markov switching model with constant transition probabilities (MC, henceforward) results in the lowest RMSFE over the whole sample. In second place comes the Markov switching model with time-varying transition probabilities (MTV, henceforward), while the highest RMSFE is reported by the forecasting method of LT. This shows that choosing the corresponding forecasting method does not lead to superior forecasting accuracy for the algorithm of LT. When we forecast based on this algorithm, we find that the preferred identification method is the MTV model. Comparing results over the whole sample, reveals that forecasting with the algorithm of LT leads to the highest RMSFE for all identification methods. Hence, we prefer the two Markov switching models to the rule-based algorithm when forecasting the regime. The Markov switching models perform better when forecasting the regime, because they use both the mean and variance of returns. The algorithm of LT only incorporates the mean and is therefore slower in picking up changes between regimes.

Identifying and forecasting the regime with a MC model, as in Section 4.3, leads to the lowest RMSFE over the whole sample (51.21‰). However, choosing the algorithm of LT as the forecasting method increases the RMSFE to 67.44‰. This big increase in RMSFE is due to the poor forecasting performance, rather than poorly identifying states. This holds, because the sample moments in Table 11 are relatively similar for the MC model and the algorithm of LT. Interestingly, switching to the MTV model as forecasting method also substantially increases the RMSFE to 60.98‰. This shows that identifying the regime based on a MC model only shows good results, when the same method is used for forecasting the regime.

When we choose the MTV model as method for identifying the regime, the differences in RMSFE between the forecasting methods are the lowest. This is related to the increased proportion of bear markets in Figure 10b and the sample moments in Table 11. The relatively low difference in the mean returns between bull and bear markets, combined with Figure 10b, suggest that this method wrongly classifies the regime more often. This leads to estimation error in the forecast combination weights and limits the importance of accurately forecasting the state of the equity market in the next period. Furthermore, the higher proportion of bear markets, leads to less time variation in the estimation of the forecast combination weights and less extreme weights. This explains why the RMSFE of forecast combinations which use a MTV model as identification method are more robust to the choice of a forecasting method.

We now compare the RMSFE over the whole sample of all forecasting methods for the regime. This comparison reveals that forecast combinations with the MTV model as forecasting method for the regime, perform relatively well for all identification methods. Aforementioned forecast combinations are therefore robust to the identification procedure used for identifying the regime. This is interesting, because the sample over which the forecast combination weights are estimated differs across identification methods. For each of these samples, the Markov switching model with constant transition probabilities leads to the lowest RMSFE.

Comparing Model 1 to Model 2 shows that the latter is more robust to the choice of the identification and forecasting procedures. This makes sense, because in (13) we have to estimate only one parameter in each regime while Model 1 requires the estimation of a vector of weights. Therefore, the impact of wrongly classifying a period as a bull or bear market is less compared to Model 1.

Summarizing, the results in Table 12 show that mis-matching the identification and forecasting performance drops the forecasting performance in most cases. The decrease in forecasting performance is larger when switching to the algorithm of LT as identification or forecasting method. This makes sense, because this method incorporates different information than the Markov switching models. It is therefore not problematic that mis-matching the identification and forecasting method leads to a drop in performance, because the models all use different information sets. Furthermore, the results show that forecasting the regime with a MC model is preferable. Also identifying the regime with this model leads to the best results. However, the MTV model makes the forecasting performance more robust to the choice of a forecasting method for the regime.

Next, we examine the impact of changing the threshold in the threshold model based on the VIX index. As a starting point we chose the average value of the VIX over the sample. In Table 13 we present the forecasting accuracy for various other thresholds. It shows that the RMSFE is 54.19‰, when we choose the median as threshold. This value improves the RMSFE compared to Table 8, in which we choose the mean as threshold. Increasing the threshold to the 70th percentile of the VIX in our sample decreases the RMSFE to 54.90‰. Further increasing it to 90%, lowers the RMSFE to 54.30‰, but this is still higher than the RMSFE with the median as threshold. Both the 70th and 90th percentile do not improve the RMSFE in either bull or bear markets. This shows that grouping high volatility states together is not better than splitting the sample according to the median. The former likely performs bad, because it contains both bull and bear market observations which makes estimating the forecast combination weight difficult.

We now lower the threshold to the 10th and 30th percentile, this reduces the RMSFE to 54.17‰ and 54.13‰ ,respectively. However, this is only a marginal improvement compared to the median threshold. Both threshold slightly improve the RMSFE in bear markets. This is because grouping low volatility states together, mostly filters out bull market observations. Hence, estimating the forecast combination weight becomes slightly easier and this increases the effectiveness in bear markets.

The RMSFE for all thresholds in Table 13 are higher than in the benchmark. Additionally, the differences between the RMSFE of the various threshold are not very big. We therefore find that the threshold model is not overly sensitive to the choice of a threshold and that this threshold is not the reason that it performs poorly in Table 8.

Table 13: RMSFE of Threshold Model for Various Thresholds

	10%	30%	Median	70%	90%
Overall	54.17	54.13	54.19	54.90	54.30
Bull	47.01	46.94	46.82	47.28	46.96
Bear	74.24	74.25	74.74	75.87	74.83

Notes: This table reports RMSFE values for different threshold values in (15). Columns represent various percentiles of the VIX over our sample. For example, 30% corresponds to the 30th percentile. Forecast combinations are computed over the three risk-based implied expected returns, because Model 3 shows the best performance there.

5 Conclusion

In this paper, we propose various procedures to improve the forecasting accuracy of implied return forecasts. Because these forecasts are almost always positive, the difference in forecasting performance between bull and bear markets is substantial. We therefore compare the forecasting performance between these two states of the equity market.

First, we try to improve the risk aversion estimate. We do this by estimating a constant and time-varying volatility risk premium. Subsequently, we link this (time-varying) volatility risk premium to the risk aversion by searching for a constant to multiply the premium with. Secondly, to improve the poor forecasting performance in bear markets, we implement dynamic forecast combination approaches. These approaches are based on either a Markov switching model or the level of the CBOE Volatility Index (VIX). The Markov switching model identifies and forecasts whether the market is in a bull or bear state. We incorporate this information in the forecast combination, by estimating separate weights in each regime. As for the forecast combination models based on volatility, we implement a smooth transition regression model which allows for less sudden changes in the forecast combination weights. Additionally, we apply a forecast combination model based on volatility that uses cut-off points in the VIX to estimate forecast combination weights in high- and low-volatility states. We compare aforementioned improvements to implied return forecasts to a benchmark case with a constant parameter of risk aversion at 2.4, as in Ardia and Boudt [2015]

Examining implied expected returns without aforementioned additions reveals that the forecasting performance shows substantial differences between bull and bear markets. The RMSFE in bear markets is on average 72.71‰, whereas it is 46.96‰ in bull markets. Comparing implied expected returns to time-series based forecasts, shows that the former outperforms the Fama and French [1993] three factor model in bull markets, whereas the opposite is true in bear markets.

We find that a constant volatility risk premium only marginally affects the forecasting accuracy compared to the benchmark. It increases the RMSFE from 53.65‰ in the benchmark, to 53.77‰. This shows that the little time variation introduced by the multiplication constant and the re-estimation of the constant volatility risk premium is not enough to improve forecasting accuracy. When we incorporate time variation in the volatility risk premium, by means of macro-economic variables, it improves the forecasting accuracy of the resulting forecasts. The average RMSFE of these forecasts drops to 53.08‰ and improvements are statistically significant (10%) for the $\boldsymbol{w_{mkt}}$, $\boldsymbol{w_{iv}}$, $\boldsymbol{w_{erc}}$ and $\boldsymbol{w_{md}}$ portfolio proxies. This improvement in forecasting accuracy is a consequence of the drop in RMSFE in bear markets to 70.99‰. Moreover, the time-varying volatility risk premium slightly lowers the RMSFE in bull markets from 46.96‰to 46.85‰. These results show that capturing investor risk aversion by estimating a time-varying volatility risk premium improves the forecasting performance of the resulting implied expected return forecasts. However, the difference in forecasting accuracy between bull and bear markets is still substantial.

Secondly, we find that the dynamic forecast combination approach based on a Markov switching model shows the largest improvement in RMSFE, to 51.21‰. This improvement follows from the significant (1%) decrease in bear markets to 62.31‰. However, the decrease in bear markets comes at a cost of significantly (10%) increasing the RMSFE in bull markets to 47.40‰. Forecast combination approaches based on the level of volatility perform relatively poorly in most circumstances, expect when the only return forecasts in the forecast combination are risk-based implied returns. However, even then is the difference in forecasting accuracy not statistically significant at the 10% level and do these forecasts not outperform the forecast combination approaches based on a Markov switching model. In bear markets, the latter benefit from including time-series based forecasts in the forecast combination, whereas their inclusion reduces the forecasting accuracy in bull markets. Consequently, the inclusion of time-series based forecasts marginally decreases the RMSFE from 51.21‰ to 51.23‰. However, the hit ratio shows a larger decrease from 63.54% to 60.67%. In general, forecast combinations based on a Markov switching model improve the the hit ratio. Combining this with the evolution of these forecasts, shows that they can now predict negative returns and as shown before, this improves the forecasting accuracy, especially in bear markets. Forecast combination approaches based on the state of the equity market are therefore a suitable solution to the problem of poor performance of implied expected returns in bear markets.

Because of data availability, we choose to compute the realized volatility over a given month by means of daily returns, instead of higher frequency returns. Bollerslev et al. [2011] note that this results in poorer finite-sample performance of the estimator in (24). The estimate of the (time-varying) volatility risk premium in our results is therefore likely biased. Because our main focus is on the effect this volatility risk premium has on the forecasting performance of implied expected returns, this is less of a problem. It is also somewhat mitigated we do not directly link the volatility risk premium to the risk aversion. Instead, we first estimate a multiplication constant to multiply the volatility risk premium with.

Another complication is that the VIX index starts in January 1990. Hence, because we want to estimate the volatility risk premium with a sample of no less than 100 observations, the evaluation sample starts on June 1998. This reduces the number of monthly observations we wish to forecast to 223 in each industry of which 48 are in bear markets. This is not a long period of time. Future research could be conducted on whether our conclusions hold in other or larger datasets. Moreover, investigating whether implied expected returns also outperform time-series based forecasts in emerging markets could prove to be insightful. This is because finding a portfolio that has no unsystematic risk is likely harder in emerging markets. This means that investors are not fully compensated for the risk they take with this portfolio. Implied expected returns are therefore more likely to be inaccurate.

Our attempt to improve the forecasting performance of implied expected returns in bear markets, focuses on forecast combinations. Investigating whether choosing a forecast method, instead of combining them, results in a similar increase in forecasting accuracy can provide insight into what the shortcomings of individual forecasting methods are. If this choice involves implied expected returns, forecasts probably have to be multiplied by a constant to reduce the bias in bear markets. Furthermore, in (11) we use the probability of being in a bear or bull state to weigh the forecast combination weights. It is also possible to choose a cut-off point where either the weights in bear or bull markets are selected. This does not incorporate the uncertainty surrounding the state of the equity market in the next period, which likely affects forecasting performance.

References

David Ardia and Kris Boudt. Implied expected returns and the choice of a mean–variance efficient portfolio proxy. *Journal of Portfolio Managment*, 41(4):68–81, 2015.

Fischer Black and Robert Litterman. Asset allocation: Combining investor views with market equilibrium. *The Journal of Fixed Income*, 1(2):7–18, 1991.

Tim Bollerslev and Hao Zhou. Estimating stochastic volatility diffusion using conditional moments of integrated volatility. *Journal of Econometrics*, 109(1):33 – 65, 2002. ISSN 0304-4076. doi: http://dx.doi.org/10.1016/S0304-4076(01)00141-5. URL `http://www.sciencedirect.com/science/article/pii/S0304407601001415`.

Tim Bollerslev and Hao Zhou. Volatility puzzles: a simple framework for gauging return-volatility regressions. *Journal of Econometrics*, 131(1-2):123–150, 2006. URL `http://EconPapers.repec.org/RePEc:eee:econom:v:131:y:2006:i:1-2:p:123-150`.

Tim Bollerslev, Robert F. Engle, and Jeffrey M. Woolridge. A capital asset pricing model with time-varying covariances. *Journal of Political Economy*, 96(1), 1988.

Tim Bollerslev, Michael Gibson, and Hao Zhou. Dynamic estimation of volatility risk premia and investor risk aversion from option-implied and realized volatilities. *Journal of Econometrics*, 160(1):235–245, 2011.

Mark Britten-Jones and Anthony Neuberger. Option prices, implied price processes, and stochastic volatility. *The Journal of Finance*, 55(2):839–866, 2000. ISSN 00221082, 15406261. URL `http://www.jstor.org/stable/222524`.

John Y. Campbell and John H. Cochrane. By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy*, 107(2):205–251, 1999. ISSN 00223808, 1537534X. URL `http://www.jstor.org/stable/10.1086/250059`.

Rachel Campbell, Kees Koedijk, and Paul Kofman. Increased correlation in bear markets. *Financial Analysts Journal*, 58(1):87–94, 2002. ISSN 0015198X. URL `http://www.jstor.org/stable/4480371`.

Yves Choueifaty and Yves Coignard. Toward maximum diversification. *Journal of Portfolio Management*, 35(1):40–51, 2008.

Victor DeMiguel, Lorenzo Garlappi, and Raman Uppal. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial Studies*, 22(5):1915–1953, 2009.

Melinda Deutsch, Clive W.J Granger, and Timo Teräsvirta. The combination of forecasts using changing weights. *International Journal of Forecasting*, 10(1):47–57, 1994.

David A. Dickey and Wayne A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366):427–431, 1979. ISSN 01621459. URL `http://www.jstor.org/stable/2286348`.

Francis X. Diebold and Robert S. Mariano. Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3):253–263, 1995.

Zhuanxin Ding. An implementation of markov regime switching model with time varying transition probabilities in matlab. 2012.

Eugene F. Fama and Kenneth R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3 – 56, 1993. ISSN 0304-405X. doi: https://doi.org/10.1016/0304-405X(93)90023-5. URL `http://www.sciencedirect.com/science/article/pii/0304405X93900235`.

Clive W. J. Granger and Ramu Ramanathan. Improved methods of combining forecasts. *Journal of Forecasting*, 3(2):197–204, 1984. ISSN 1099-131X. doi: 10.1002/for.3980030207. URL `http://dx.doi.org/10.1002/for.3980030207`.

Luigi Guiso, Paola Sapienza, and Luigi Zingales. Time varying risk aversion. (19284), August 2013. doi: 10.3386/w19284. URL `http://www.nber.org/papers/w19284`.

James D. Hamilton. Analysis of time series subject to changes in regime. *Journal of Econometrics*, 45 (1):39–70, 1990.

Roger Hartley and Lisa Farrell. Can expected utility theory explain gambling? *American Economic Review*, 92(3):613–624, June 2002. doi: 10.1257/00028280260136426. URL `http://www.aeaweb.org/articles?id=10.1257/00028280260136426`.

Ulf Herold. Computing implied returns in a meaningful way. *Journal of Asset Management*, 6(1):53–64, 2005.

Erik Kole and Dick van Dijk. How to identify and forecast bull and bear markets? *Journal of Applied Econometrics*, 32(1):120–139, 2017.

R Leote De Carvalho, X Lu, and P Moulin. Demystifying equity risk-based strategies: A simple alpha plus beta description. *The Journal of Portfolio Management*, 38(3):56–70, 2012.

Chien-Fu Jeff Lin and Timo Teräsvirta. Testing the constancy of regression parameters against continuous structural change. *Journal of Econometrics*, 62(2):211 – 228, 1994. ISSN 0304-4076. doi: https://doi.org/10.1016/0304-4076(94)90022-1. URL `http://www.sciencedirect.com/science/article/pii/0304407694900221`.

Asger Lunde and Allan Timmermann. Duration dependence in stock prices: an analysis of bull and bear markets. *Journal of Business and Economic Statistics*, 22(3):253–273, 2004.

Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

Rajnish Mehra and Edward C. Prescott. The equity premium: A puzzle. *Journal of Monetary Economics*, 15(2):145–161, 1985.

Marcelo Perlin. Ms regress - the matlab package for markov regime switching models. 2015.

Sietse Romijn. Implied expected return forecasts using a bekk model as covariance estimator. 2016.

William Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3):425–442, 1964.

William F. Sharpe. Imputing expected security returns from portfolio composition. *The Journal of Financial and Quantitative Analysis*, 9(3):463–472, 1974. ISSN 00221090, 17566916. URL `http://www.jstor.org/stable/2329873`.

Allan Timmermann. *Forecast Combinations*, volume 1 of *Handbook of Economic Forecasting*, chapter 4, pages 135–196. Elsevier, 2006. URL `https://ideas.repec.org/h/eee/ecofch/1-04.html`.

Peter Robert Wheale and Laura Heredia Amin. Bursting the dot.com "bubble': A case study in investor behaviour. *Technology Analysis & Strategic Management*, 15(1):117–136, 2003. doi: 10.1080/0953732032000046097. URL `http://dx.doi.org/10.1080/0953732032000046097`.

A Appendix

A.1 Lunde and Timmermann [2004]

The algorithm of Lunde and Timmermann [2004] requires input for the initial state of the market. To determine this, we initialize the minimum and the maximum at the first observation and count how many times each has to be adjusted. If the maximum(mimimum) is the first that needs to be adjusted three times, the initial state is a bull(bear) market. Provided with this initial state and a time series of index prices P_t, the algorithm can be summarized, similar to Kole and van Dijk [2017], as follows:

1. Initialize the last extreme depending on the initial state of the market.
2. If the last extreme before time t was a peak with value P_{max}.
   (a) If $P_t > P_{max}$, the maximum is updated $P_{max} = P_t$ and the new extreme is a peak.
   (b) If $P_t < \lambda_2 P_{max}$ for some fraction λ_2, a new trough has been found and $P_{min} = P_t$.
   (c) If neither of these conditions hold, no update takes place and we move to $t + 1$.
3. If the last extreme before t was a peak with value P_{min}.
   (a) If $P_t < P_{min}$, the minimum is updated $P_{min} = P_t$ and the new extreme is a trough.
   (b) If $P_t > \lambda_1 P_{min}$ for some fraction λ_1, a new peak has been found and $P_{max} = P_t$.
   (c) If neither of these conditions hold, no update takes place and we move to $t + 1$.
4. Classify the market in a bull state for the period following the trough until the peak.
5. Classify the market in a bear state for the period following the peak until the trough.

As for the values λ_1 and λ_2 we choose 1.2 and 0.85 respectively.

A.2 Benchmark Case and Risk Aversion Additional Results

Table 14: Diebold-Mariano Tests for Benchmark Case Unconditional on the State of the Equity Market

	w_{ew}	w_{mkt}	w_{iv}	w_{erc}	w_{md}	**Constant**	**Fama**	**AR(1)**
w_{ew}	-	0.112	0.380	0.592	0.464	0.039	0.080	0.022
w_{mkt}		-	0.497	0.735	0.527	0.039	0.082	0.022
w_{iv}			-	0.944	0.546	0.034	0.057	0.020
w_{erc}				-	0.366	0.029	0.047	0.019
w_{md}					-	0.025	0.030	0.017
Constant						-	0.719	0.085
Fama							-	0.092
AR(1)								-

Notes: This table reports p-values of the Diebold-Mariano test. This test compares whether the forecasting method in a given rowen outperforms the forecasting method in a given column. Forecasting methods refer to the benchmark case and are unconditional on the state of the equity market. P-values are only reported for the upper triangular of the matrix, since the lower triangular can be derived from the reported values.

Table 15: Diebold-Mariano Tests for Benchmark Case in Bull Markets

	w_{ew}	w_{mkt}	w_{iv}	w_{erc}	w_{md}	**Constant**	**Fama**	**AR(1)**
w_{ew}	-	0.001	0.001	0.001	0.001	0.007	0.000	0.003
w_{mkt}		-	0.001	0.004	0.001	0.008	0.000	0.004
w_{iv}			-	1.000	0.002	0.016	0.000	0.006
w_{erc}				-	0.001	0.013	0.000	0.005
w_{md}					-	0.032	0.000	0.009
Constant						-	0.267	0.050
Fama							-	0.193
AR(1)								-

Notes: This table reports p-values of the Diebold-Mariano test. This test compares whether the forecasting method in a given rowen outperforms the forecasting method in a given column. Forecasting methods refer to the benchmark case and are conditional on being in a bull state. P-values are only reported for the upper triangular of the matrix, since the lower triangular can be derived from the reported values.

Table 16: Diebold-Mariano Tests for Benchmark Case in Bear Markets

	w_{ew}	w_{mkt}	w_{iv}	w_{erc}	w_{md}	**Constant**	**Fama**	**AR(1)**
w_{ew}	-	0.978	1.000	0.997	1.000	0.708	1.000	0.663
w_{mkt}		-	1.000	0.996	1.000	0.685	1.000	0.648
w_{iv}			-	0.419	0.999	0.491	0.999	0.539
w_{erc}				-	0.999	0.498	1.000	0.542
w_{md}					-	0.235	0.996	0.392
Constant						-	0.959	0.570
Fama							-	0.146
AR(1)								-

Notes: This table reports p-values of the Diebold-Mariano test. This test compares whether the forecasting method in a given rowen outperforms the forecasting method in a given column. Forecasting methods refer to the benchmark case and are conditional on being in a bear state. P-values are only reported for the upper triangular of the matrix, since the lower triangular can be derived from the reported values.

Table 17: Diebold-Mariano Tests for Implied Expected Returns and Time-series Based Forecasts Unconditional on the State of the Equity Market

		Constant					TV					TS		
		w_{ew}	w_{mkt}	w_{iv}	w_{erc}	w_{md}	w_{ew}	w_{mkt}	w_{iv}	w_{erc}	w_{md}	**Constant**	**Fama**	**AR(1)**
Constant	w_{ew}	-	0.239	0.432	0.631	0.557	0.904	0.927	0.959	0.947	0.980	0.045	0.106	0.023
	w_{mkt}		-	0.496	0.722	0.601	0.906	0.929	0.960	0.949	0.982	0.045	0.105	0.023
	w_{iv}			-	0.927	0.682	0.903	0.926	0.959	0.947	0.983	0.041	0.081	0.021
	w_{erc}				-	0.471	0.896	0.920	0.955	0.941	0.980	0.036	0.069	0.020
	w_{md}					-	0.887	0.912	0.948	0.933	0.977	0.033	0.046	0.019
TV	w_{ew}						-	0.774	0.560	0.530	0.402	0.036	0.039	0.011
	w_{mkt}							-	0.470	0.402	0.341	0.029	0.029	0.009
	w_{iv}								-	0.374	0.264	0.019	0.015	0.006
	w_{erc}									-	0.326	0.023	0.020	0.007
	w_{md}										-	0.012	0.006	0.005
TS	**Constant**											-	0.719	0.085
	Fama												-	0.092
	AR(1)													-

Note: This table reports p-values of the Diebold-Mariano test. This test compares whether the forecasting method in a given rowen outperforms the forecasting method in a given column. Constant: to implied expected returns with a constant volatility risk premium. TV: implied expected returns with a time-varying volatility risk premium. TS: time-series based forecasts. All results are unconditional on the state of the equity market. P-values are only reported for the upper triangular of the matrix, since the lower triangular can be derived from the reported values.

Table 18: Diebold-Mariano Tests for Implied Expected Returns and Time-series Based Forecasts in Bull States

		Constant					TV					TS		
		$\boldsymbol{w_{ew}}$	$\boldsymbol{w_{mkt}}$	$\boldsymbol{w_{iv}}$	$\boldsymbol{w_{erc}}$	$\boldsymbol{w_{md}}$	$\boldsymbol{w_{ew}}$	$\boldsymbol{w_{mkt}}$	$\boldsymbol{w_{iv}}$	$\boldsymbol{w_{erc}}$	$\boldsymbol{w_{md}}$	**Constant**	**Fama**	**AR(1)**
Constant	$\boldsymbol{w_{ew}}$	-	0.015	0.007	0.008	0.012	0.679	0.708	0.751	0.734	0.702	0.029	0.000	0.008
	$\boldsymbol{w_{mkt}}$		-	0.008	0.012	0.013	0.696	0.725	0.770	0.753	0.733	0.031	0.000	0.009
	$\boldsymbol{w_{iv}}$			-	0.993	0.027	0.760	0.788	0.838	0.819	0.839	0.049	0.000	0.012
	$\boldsymbol{w_{erc}}$				-	0.017	0.742	0.771	0.820	0.802	0.813	0.044	0.000	0.011
	$\boldsymbol{w_{md}}$					-	0.799	0.826	0.875	0.857	0.892	0.072	0.000	0.015
TV	$\boldsymbol{w_{ew}}$						-	0.697	0.564	0.580	0.372	0.100	0.020	0.018
	$\boldsymbol{w_{mkt}}$							-	0.509	0.508	0.320	0.087	0.014	0.015
	$\boldsymbol{w_{iv}}$								-	0.491	0.208	0.066	0.005	0.010
	$\boldsymbol{w_{erc}}$									-	0.256	0.074	0.008	0.012
	$\boldsymbol{w_{md}}$										-	0.060	0.002	0.009
TS	**Constant**											-	0.267	0.050
	Fama												-	0.193
	AR(1)													-

Note: This table reports p-values of the Diebold-Mariano test. This test compares whether the forecasting method in a given rowen outperforms the forecasting method in a given column. Constant: to implied expected returns with a constant volatility risk premium. TV: implied expected returns with a time-varying volatility risk premium. TS: time-series based forecasts. Results are conditional on the equity market being in a bull state. P-values are only reported for the upper triangular of the matrix, since the lower triangular can be derived from the reported values.

Table 19: Diebold-Mariano Tests for Implied Expected Returns and Time-series Based Forecasts in Bear States

		Constant					TV					TS		
		w_{ew}	w_{mkt}	w_{iv}	w_{erc}	w_{md}	w_{ew}	w_{mkt}	w_{iv}	w_{erc}	w_{md}	**Constant**	**Fama**	**AR(1)**
Constant	w_{ew}	-	0.961	1.000	0.993	1.000	0.972	0.979	0.989	0.990	0.999	0.445	0.998	0.515
	w_{mkt}		-	1.000	0.993	1.000	0.970	0.978	0.989	0.989	0.999	0.420	0.998	0.502
	w_{iv}			-	0.552	0.999	0.944	0.957	0.975	0.974	0.997	0.265	0.996	0.412
	w_{erc}				-	0.999	0.950	0.961	0.978	0.976	0.998	0.261	0.997	0.410
	w_{md}					-	0.878	0.900	0.929	0.923	0.988	0.132	0.987	0.313
TV	w_{ew}						-	0.813	0.499	0.419	0.558	0.050	0.722	0.174
	w_{mkt}							-	0.390	0.328	0.521	0.043	0.711	0.165
	w_{iv}								-	0.288	0.599	0.035	0.750	0.163
	w_{erc}									-	0.665	0.039	0.768	0.169
	w_{md}										-	0.017	0.756	0.143
TS	**Constant**											-	0.959	0.570
	Fama												-	0.146
	AR(1)													-

Note: This table reports p-values of the Diebold-Mariano test. This test compares whether the forecasting method in a given rowen outperforms the forecasting method in a given column. Constant: to implied expected returns with a constant volatility risk premium. TV: implied expected returns with a time-varying volatility risk premium. TS: time-series based forecasts. Results are conditional on the equity market being in a bear state. P-values are only reported for the upper triangular of the matrix, since the lower triangular can be derived from the reported values.

Table 20: Diebold-Mariano Tests for Forecast Combinations Unconditional on the State of the Equity Market

		All-ier					All					Risk-Based				
		Model 1	Model 2	Model 3	Model 4	Model 5	Model 1	Model 2	Model 3	Model 4	Model 5	Model 1	Model 2	Model 3	Model 4	Model 5
All-ier	Model 1	-	0.05	0.00	0.02	0.02	0.51	0.03	0.00	0.01	0.02	0.02	0.05	0.00	0.24	0.02
	Model 2		-	0.02	0.08	0.06	0.89	0.02	0.00	0.05	0.07	0.71	0.34	0.02	0.51	0.07
	Model 3			-	0.56	0.96	0.99	0.96	0.11	0.49	0.96	0.99	0.98	0.69	0.96	0.97
	Model 4				-	0.60	0.97	0.89	0.33	0.27	0.61	0.94	0.92	0.48	0.98	0.60
	Model 5					-	0.97	0.90	0.00	0.32	0.60	0.96	0.94	0.00	0.88	0.60
All	Model 1						-	0.07	0.00	0.02	0.04	0.15	0.11	0.01	0.28	0.03
	Model 2							-	0.00	0.07	0.11	0.88	0.98	0.03	0.60	0.11
	Model 3								-	0.60	0.99	1.00	1.00	0.93	0.99	1.00
	Model 4									-	0.68	0.97	0.95	0.55	0.99	0.68
	Model 5										-	0.95	0.93	0.00	0.87	0.48
Risk-based	Model 1											-	0.28	0.01	0.44	0.05
	Model 2												-	0.02	0.51	0.07
	Model 3													-	0.94	1.00
	Model 4														-	0.13
	Model 5															-

This table reports p-values of the Diebold-Mariano test. This test compares whether the forecasting method in a given rowen outperforms the forecasting method in a given column. All-ier: All five implied expected returns. All: All five implied expected returns and all time-series based forecasts. Risk-based: Risk-based implied expected returns. Results are unconditional on the state of the equity market. P-values are only reported for the upper triangular of the matrix, since the lower triangular can be derived from the reported values.

Table 21: Diebold-Mariano Tests for Forecast Combinations in Bull Markets

		All-ier					All					Risk-Based				
		Model 1	Model 2	Model 3	Model 4	Model 5	Model 1	Model 2	Model 3	Model 4	Model 5	Model 1	Model 2	Model 3	Model 4	Model 5
All-ier	Model 1	-	0.03	0.67	0.01	0.92	0.08	0.02	0.10	0.01	0.92	0.06	0.03	0.79	0.26	0.94
	Model 2		-	0.91	0.06	0.99	0.63	0.04	0.55	0.07	0.99	0.94	0.12	0.97	0.70	0.99
	Model 3			-	0.01	0.86	0.14	0.07	0.01	0.01	0.86	0.15	0.09	0.59	0.06	0.89
	Model 4				-	1.00	0.96	0.91	0.95	0.58	1.00	0.97	0.93	0.99	0.98	1.00
	Model 5					-	0.02	0.01	0.00	0.00	0.55	0.01	0.01	0.00	0.00	0.85
All	Model 1						-	0.26	0.44	0.04	0.98	0.68	0.36	0.94	0.59	0.98
	Model 2							-	0.69	0.09	0.99	0.97	0.95	0.98	0.79	1.00
	Model 3								-	0.06	1.00	0.72	0.44	1.00	0.76	1.00
	Model 4									-	1.00	0.97	0.93	0.99	0.98	1.00
	Model 5										-	0.01	0.01	0.01	0.01	0.72
Risk-based	Model 1											-	0.06	0.96	0.50	0.99
	Model 2												-	0.97	0.70	0.99
	Model 3													-	0.03	1.00
	Model 4														-	1.00
	Model 5															-

This table reports p-values of the Diebold-Mariano test. This test compares whether the forecasting method in a given rowen outperforms the forecasting method in a given column. All-ier: All five implied expected returns. All: All five implied expected returns and all time-series based forecasts. Risk-based: Risk-based implied expected returns. Results are for bull states. P-values are only reported for the upper triangular of the matrix, since the lower triangular can be derived from the reported values.

Table 22: Diebold-Mariano Tests for Forecast Combinations in Bear Markets

		All-ier					All					Risk-Based				
		Model 1	Model 2	Model 3	Model 4	Model 5	Model 1	Model 2	Model 3	Model 4	Model 5	Model 1	Model 2	Model 3	Model 4	Model 5
All-ier	Model 1	-	0.35	0.00	0.33	0.00	0.84	0.29	0.00	0.18	0.00	0.07	0.35	0.00	0.33	0.00
	Model 2		-	0.00	0.39	0.00	0.92	0.12	0.00	0.21	0.00	0.29	0.75	0.00	0.38	0.00
	Model 3			-	0.99	0.94	1.00	1.00	0.95	0.99	0.93	1.00	1.00	0.70	0.99	0.93
	Model 4				-	0.03	0.79	0.59	0.01	0.06	0.03	0.55	0.62	0.01	0.44	0.02
	Model 5					-	1.00	1.00	0.43	0.96	0.58	1.00	1.00	0.04	0.97	0.08
All	Model 1						-	0.07	0.00	0.11	0.00	0.06	0.08	0.00	0.20	0.00
	Model 2							-	0.00	0.23	0.00	0.39	0.91	0.00	0.40	0.00
	Model 3								-	0.98	0.58	1.00	1.00	0.20	0.98	0.52
	Model 4									-	0.04	0.72	0.79	0.02	0.97	0.03
	Model 5										-	1.00	1.00	0.05	0.97	0.23
Risk-based	Model 1											-	0.72	0.00	0.44	0.00
	Model 2												-	0.00	0.37	0.00
	Model 3													-	0.98	0.94
	Model 4														-	0.03
	Model 5															-

This table reports p-values of the Diebold-Mariano test. This test compares whether the forecasting method in a given rowen outperforms the forecasting method in a given column. All-ier: All five implied expected returns. All: All five implied expected returns and all time-series based forecasts. Risk-based: Risk-based implied expected returns. Results are for bear states. P-values are only reported for the upper triangular of the matrix, since the lower triangular can be derived from the reported values.

A.3 Forecast Combination Approaches

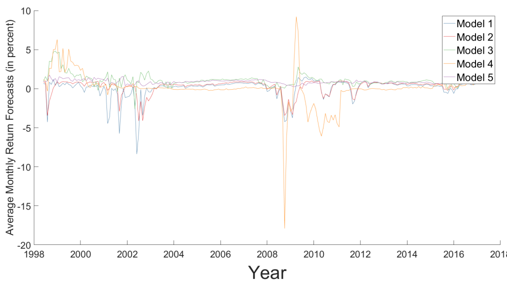
We now illustrate how the forecast combination of implied expected return forecasts can be seen as an implied expected return itself. Let μ_1 and μ_2 be implied expected return forecasts with mean-variance efficient portfolio proxies w_1 and w_2, respectively. Then, according to (3):

$$\mu_1 = \ell\iota + \gamma\Sigma w_1$$
$$\mu_2 = \ell\iota + \gamma\Sigma w_2,$$

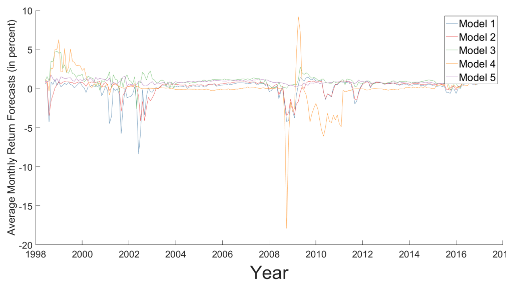
A linear combination $\lambda_1\mu_1 + \lambda_2\mu_2$ of the two implied return forecasts equals:

$$\lambda_1\mu_1 + \lambda_2\mu_2 = (\lambda_1 + \lambda_2)\ell\iota + \gamma\Sigma(\lambda_1 w_1 + \lambda_2 w_2).$$

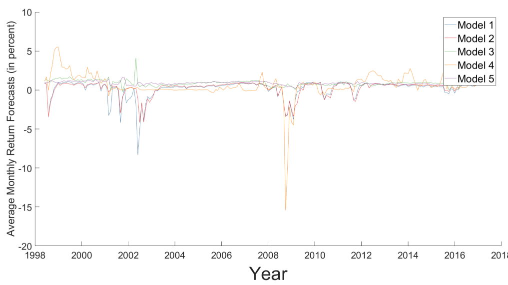
This shows that a linear combination of implied return forecasts can be seen as a implied return forecast with intercept parameter scaled by $\lambda_1 + \lambda_2$ and a mean-variance efficient portfolio proxy that is a linear combination of the proxies w_1 and w_2. This linear combination can be seen as a mean-variance efficient portfolio proxy as well. Not imposing the restriction that $\lambda_1 + \lambda_2 = 1$ loosens the budget restriction, such that it is now possible for the mean-variance efficient portfolio proxy to not be fully invested. Consequently, the intercept parameter changes as well. The argument presented above generalizes to linear combinations of more implied return forecasts and to all forecast combination models in Section 2.4.



(a) All Implied Return Forecasts



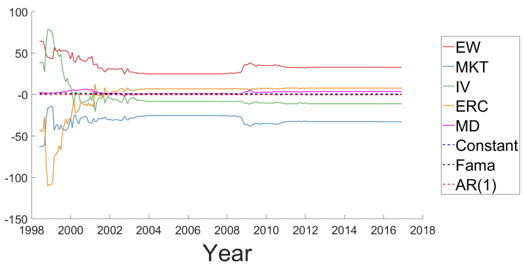
(b) All Implied and Time-Series Based Return Forecasts



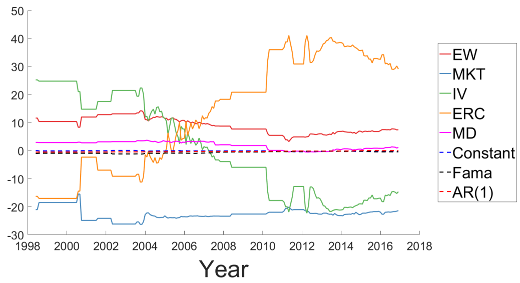
(c) Risk-Based Implied Return Forecasts

Figure 11: Evolution of Forecast Combinations of Implied Return Forercasts

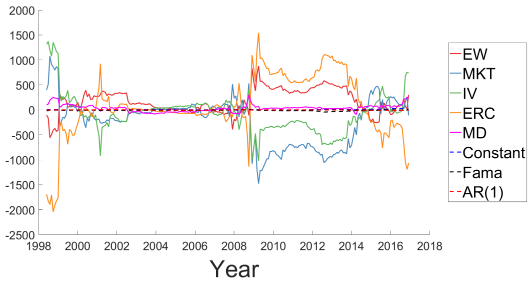
(a) Model 3 Bull



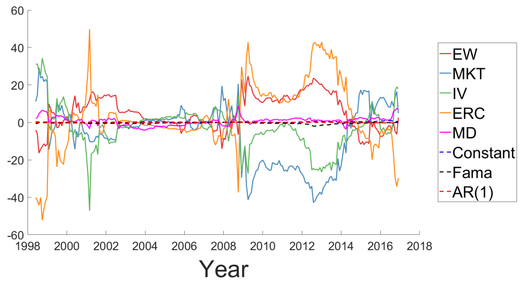
(b) Model 3 Bear



(c) Model 4 Bull



(d) Model 4 Bear



(e) Model 5

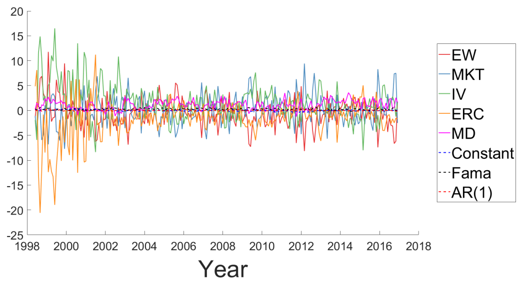


Figure 12: Forecast Combination Weights

Notes: This Figure shows the forecast combination weights for Models 3,4 and 5 in bull and bear markets. Dashed lines represent time-series based forecasts. First five entries in the legend stand for implied expected returns for various portfolio proxies. EW: equally weighted portfolio. MKT: market capitalization portfolio. IV: inverse volatility portfolio. ERC: equal risk contribution portfolio. MD: maximum diversification portfolio.

A.4 Summary Statistic Conditional on the State of the Equity Market

Table 23: Summary Statistics of Return Data in Bull Markets

Industries	Mean(%)	Vol(%)	Skew	Kurt
Non Durables	1.22	3.27	0.02	2.81
Durables	1.83	6.76	1.37	9.87
Manufacturing	1.79	4.17	0.38	3.75
Energy	1.71	5.48	0.33	3.36
Hi-Tech	2.35	5.89	0.35	3.63
Telecom	1.72	4.55	0.39	4.53
Shops	1.39	4.02	0.06	3.43
Healthcare	1.40	3.83	-0.16	3.33
Utilities	1.36	3.63	-0.29	2.97
Other	1.70	4.40	0.33	3.73
Average	1.65	4.60	0.28	4.14

Notes: Skew: Skewness. Kurt: Kurtosis. Average: Average over the columns. Statistics are computed from June 1998 to December 2016.

Table 24: Summary Statistics of Return Data in Bear Markets

Industries	Mean(%)	Vol(%)	Skew	Kurt
Non Durables	-0.91	4.42	-0.78	3.31
Durables	-4.02	8.90	-0.55	3.99
Manufacturing	-2.71	6.44	-0.50	2.92
Energy	-2.26	6.76	-0.10	2.52
Hi-Tech	-4.77	9.97	0.32	3.02
Telecom	-3.81	6.46	0.16	2.86
Shops	-1.53	5.50	-0.40	2.90
Healthcare	-1.82	4.54	-0.35	2.66
Utilities	-1.41	5.58	-0.26	2.70
Other	-3.69	6.68	-0.74	3.43
Average	-2.69	6.53	-0.32	3.03

Notes: Skew: Skewness. Kurt: Kurtosis. Average: Average over the columns. Statistics are computed from June 1998 to December 2016.