



# *Optimal Portfolio Choice under Decision-Based Model Combinations*

*Davide Pettenuzzo, Economics Department, Brandeis University*

*Francesco Ravazzolo, Norges Bank and BI Norwegian Business School*

**Working Paper Series**

---

# Optimal Portfolio Choice under Decision-Based Model Combinations\*

**Davide Pettenuzzo**<sup>†</sup>  
Brandeis University

**Francesco Ravazzolo**<sup>‡</sup>  
Norges Bank, and  
BI Norwegian Business School

October 27, 2014

## Abstract

We propose a novel Bayesian model combination approach where the combination weights depend on the past forecasting performance of the individual models entering the combination through a utility-based objective function. We use this approach in the context of stock return predictability and optimal portfolio decisions, and investigate its forecasting performance relative to a host of existing combination schemes. We find that our method produces markedly more accurate predictions than the existing model combinations, both in terms of statistical and economic measures of out-of-sample predictability. We also investigate the incremental role of our model combination method in the presence of model instabilities, by considering predictive regressions that feature time-varying regression coefficients and volatility. We find that the gains from using our model combination method increase significantly when we allow for instabilities in the individual models entering the combination.

*Key words:* Bayesian econometrics; Time-varying parameters; Model combinations; Portfolio choice.

*JEL classification:* C11; C22; G11; G12

---

\*This Working Paper should not be reported as representing the views of Norges Bank. The views expressed are those of the authors and do not necessarily reflect those of Norges Bank. We would like to thank Blake LeBaron, Allan Timmermann, and Ross Valkanov for helpful comments.

<sup>†</sup>Department of Economics, Brandeis University. Sachar International Center, 415 South St, Waltham, MA. Tel: (781) 736-2834. Fax: +1 (781) 736 2269. Email: [dpettenu@brandeis.edu](mailto:dpettenu@brandeis.edu).

<sup>‡</sup>Norges Bank. Bankplassen 2, P.O. Box 1179 Sentrum, 0107 Oslo, Norway. Tel: +47 22 31 61 72. Fax: +47 22 42 40 62. Email: [francesco.ravazzolo@norges-bank.no](mailto:francesco.ravazzolo@norges-bank.no).

# 1 Introduction

Over the years, the question of whether stock returns are predictable has received considerable attention, both within academic and practitioner circles.<sup>1</sup> However, more than 25 years of research on this topic shows that models allowing for time-varying return predictability often produce worse out-of-sample forecasts than a simple benchmark that assumes a constant risk premium. This finding has led authors such as [Bossaerts and Hillion \(1999\)](#) and [Welch and Goyal \(2008\)](#) to question the economic value of return predictability, and to suggest that there are no out-of-sample benefits to investors from exploiting this predictability when making optimal portfolio decisions.

Forecast combination methods offer a way to improve equity premium forecasts. Since [Bates and Granger \(1969\)](#) seminal paper on forecast combinations, it has been known that combining forecasts across models often produces a forecast that performs better than even the best individual model. [Timmermann \(2006\)](#) offers a compelling explanation for this stylized fact. In a sense, forecast combinations can be thought of a diversification strategy that improves forecast performance, much like asset diversification improves portfolio performance. [Avramov \(2002\)](#), [Aiolfi and Favero \(2005\)](#), [Rapach et al. \(2010\)](#), and [Dangl and Halling \(2012\)](#) confirm this result in the context of stock return predictability, and find that the empirical evidence of out-of-sample predictability improves when using model combinations.

Existing forecast combination methods weight together the individual models according to their statistical performance, without making specific reference to the way the forecasts are used.<sup>2</sup> For example, in [Rapach et al. \(2010\)](#) the individual models are combined according to their relative mean squared prediction error, while [Avramov \(2002\)](#) and [Dangl and Halling \(2012\)](#) use Bayesian Model Averaging (BMA), which weights the individual models according to their marginal likelihoods. In contrast, with stock return forecasts the quality of the individual model predictions depends ultimately on whether such predictions deliver profitable investment decisions, which in turns is directly related to the investor's utility function. This creates an inconsistency between the criterion used to combine the individual predictions and the final use to which the forecasts will be put.

---

<sup>1</sup>The literature on of stock return predictability became particularly active during the 1970s and 1980s. Earlier work in this field include [Fama and Schwert \(1977\)](#), [Keim and Stambaugh \(1986\)](#), [Campbell \(1987\)](#), [Campbell and Shiller \(1988\)](#), [Fama and French \(1988, 1989\)](#), and [Ferson and Harvey \(1991\)](#). More recently, several other authors have suggested new predictor variables, such as the corporate payout and financing activity ([Lamont \(1998\)](#), [Baker and Wurgler \(2000\)](#)), the level of consumption in relation to wealth ([Lettau and Ludvigson \(2001\)](#)), and the relative valuation of low- and high-beta stocks ([Polk et al. \(2006\)](#)).

<sup>2</sup>This is very closely related to the debate between statistical and decision-based approaches to forecast evaluation. The statistical approach focuses on general measures of forecast accuracy intended to be relevant in a variety of circumstances, while the decision-based approach provides techniques with which to evaluate the economic value of forecasts to a particular decision maker or group of decision makers. See [Granger and Machina \(2006\)](#) and [Pesaran and Skouras \(2007\)](#) for comprehensive reviews on this subject.

In this paper, we introduce a novel Bayesian model combination technique where the predictive densities of the individual models are weighted together based on how each model fares relative to the final objective function of the investor. In the spirit of [Pesaran and Skouras \(2007\)](#), we label this new method Decision-Based Density Combination (DB-DeCo), and stress that this new approach combines the entire predictive densities of the individual models, rather than only their point forecasts. Furthermore, our DB-DeCo method features time-varying combination weights, and explicitly factors into the model combination the inherent uncertainty surrounding the estimation of the combination weights.

Our paper adds to a rapidly growing literature developing new and more flexible model combination methods. In particular, our work relates to and extends the contributions of [Geweke and Amisano \(2011\)](#), [Del Negro et al. \(2013\)](#), [Billio et al. \(2013\)](#), and [Fawcett et al. \(2014\)](#). [Geweke and Amisano \(2011\)](#) propose combining a set of individual predictive densities with weights chosen to maximize the predictive log-likelihood of the final model combination, while [Fawcett et al. \(2014\)](#) and [Del Negro et al. \(2013\)](#) generalize their approach to include time-varying weights. On the other hand, [Billio et al. \(2013\)](#) propose a model combination scheme where the individual model weights can change over time, and depend on a learning mechanism based on a squared prediction error function. The approach we propose in this paper shares with the previous papers the feature that the combination weights can change over time. However, differently from these papers, our combination scheme allows for the combination weights to depend on the individual models' past performance in a highly flexible way, through a utility-based objective function.

To test our approach empirically, we evaluate how it fares relative to a host of alternative model combination methods, and consider as the individual models entering the combinations a set of linear predictive regressions for stock returns, each including as regressor one of the predictor variables used by [Welch and Goyal \(2008\)](#). Focusing on linear univariate models and relying on the same set of variables that have been previously studied in the literature allows us to make our results comparable to earlier work. When implemented along the lines proposed in our paper, we find that the DB-DeCo method leads to substantial improvements in the predictive accuracy of the equity premium forecasts. For example, we find that when comparing the DB-DeCo method to BMA, the out-of-sample  $R^2$  improves from 0.39% to 2.32%. Similar differences are found when comparing the DB-DeCo method to other model combination schemes. We also consider the economic value of using the DB-DeCo method. In the benchmark case of an investor endowed with power utility and a relative risk aversion of five, we compare the certainty equivalent return (CER) obtained from using a given model combination method relative to the prevailing mean model. We find that the DB-DeCo method yields an annualized CER of 94 basis points, while BMA delivers a negative annualized CER,  $-5$  basis points, which can be

taken as evidence that the prevailing mean model generates higher economic predictability than BMA. We also compare the economic performance of the DB-DeCo method to that of a simple equal-weighted combination method, proposed in the context of equity premium predictability by [Rapach et al. \(2010\)](#), and find that the DB-DeCo method generates an annualized CER that is 92 basis points higher than the equal-weighted combination method.

We next extend our model combination method by relaxing the linearity assumption on the individual models entering the combination. While it is well known that forecast combination methods can deal with model instabilities and structural breaks and can generate more stable forecasts than those from the individual models (see for example [Hendry and Clements \(2004\)](#) and [Stock and Watson \(2004\)](#)), the joint effect of model instabilities and model uncertainty in the context of equity return forecasts has so far received limited attention. [Dangl and Halling \(2012\)](#) and [Zhu and Zhu \(2013\)](#) are two notable exceptions. [Dangl and Halling \(2012\)](#) model time variation in the conditional mean of stock returns by allowing for gradual changes in the regression coefficients, and find that model combinations featuring these models lead to both statistically and economically significant gains over the standard predictive regressions with constant coefficients. [Zhu and Zhu \(2013\)](#) introduce a regime switching model combination to predict stock returns, and find that it delivers consistent out-of-sample gains relative to traditional model combination methods.<sup>3</sup>

We follow [Johannes et al. \(2014\)](#), and relax the linearity assumption on the individual models entering the model combinations, introducing both time-varying parameters and stochastic volatility (TVP-SV), i.e. allowing both the regression coefficients and the return volatility to change over time. Next, we recompute all model combinations by weighting together the TVP-SV models. Overall we find that controlling jointly for model instability and model uncertainty leads to further improvements in both the statistical and economic predictability of stock returns. In terms of economic predictability, we see improvements in CER for both the individual models and the various model combination methods we entertain. As for the individual models, we find that allowing for instabilities in return prediction models leads to an average increase in CER of almost 100 basis points, under the benchmark case of an investor endowed with power utility and a relative risk aversion of five. This result is in line with the findings of [Johannes et al. \(2014\)](#), but generalize them to a much larger set of predictors than those considered in their study. As for our DB-DeCo method, switching from linear to TVP-SV models produces an improvement in CER that is unrivaled, with an increase in CER of more than 150 basis points, and an absolute CER level of 249 basis points. No other model combination scheme comes close

---

<sup>3</sup>[Johannes et al. \(2014\)](#) generalize the setting of [Dangl and Halling \(2012\)](#) by forecasting stock returns allowing both regression parameters and return volatility to adjust gradually over time. However, their emphasis is not on model combination methods, and focus on a single predictor for stock returns, the dividend yield. Overall, they find that allowing for time-varying volatility leads to both statistically and economically significant gains over simpler models with constant coefficients and volatility.

to this performance.

The remainder of the paper is organized as follows. Section 2 reviews the standard Bayesian framework for predicting stock returns and choosing portfolio allocations in the presence of model and parameter uncertainty. Section 3 introduces the Decision-Based Density Combination method, highlighting the differences from the existing combination methods. Section 4 describes the data and discusses our prior choices, while Section 5 presents empirical results for a wide range of predictor variables and model combination strategies. Next, Section 6 evaluates the economic value of our novel model combination method for a risk averse investor who uses the predictions of the model to form a portfolio of stocks and a risk-free asset. Section 7 extends the linear models to allow for time-varying coefficients and stochastic volatility, and evaluates the joint role of model instabilities and model uncertainty in predicting stock returns. Finally, Section 8 conducts a range of robustness checks, while Section 9 provides some concluding remarks.

## 2 Return predictability in the presence of parameter and model uncertainty

It is common practice in the literature on return predictability to assume that stock returns, measured in excess of a risk-free rate,  $r_{\tau+1}$ , are a linear function of a lagged predictor variable,  $x_\tau$ :

$$\begin{aligned} r_{\tau+1} &= \mu + \beta x_\tau + \varepsilon_{\tau+1}, \quad \tau = 1, \dots, t-1, \\ \varepsilon_{\tau+1} &\sim N(0, \sigma_\varepsilon^2). \end{aligned} \tag{1}$$

This is the approach followed by, among others, [Welch and Goyal \(2008\)](#) and [Bossaerts and Hillion \(1999\)](#). See also [Rapach and Zhou \(2013\)](#) for an extensive review of this literature.

The linear model in (1) is simple to interpret and only requires estimating two mean parameters,  $\mu$  and  $\beta$ , which can readily be accomplished by OLS. Despite its simplicity, it has been shown empirically that the model in (1) fails to provide convincing evidence of out-of-sample return predictability. [Welch and Goyal \(2008\)](#) provide a comprehensive review on this issue, and conclude that stock return predictability is mostly an in-sample or *ex-post* phenomenon, disappearing once the prediction models are used to form forecasts on new, out-of-sample, data. One possible explanation for the results of [Welch and Goyal \(2008\)](#) is that the true data-generating process of stock returns is highly uncertain and constantly evolving, and the model in (1) is too simple for that.<sup>4</sup> In this context, the Bayesian methodology offers a valuable alternative. For one, it allows to incorporate parameter and model uncertainty into the estimation and inference steps and, compared to (1), should be more robust to model misspecifications. More

---

<sup>4</sup>See for example [Stock and Watson \(2006\)](#), and [Ang and Timmermann \(2012\)](#).

specifically, the Bayesian approach assigns posterior probabilities to a wide set of competing return-generating models. It then uses the probabilities as weights on the individual models to obtain a composite-weighted model. For example, suppose that at time  $t$  the investor wants to predict stock returns at time  $t + 1$ , and for that purpose has available  $N$  competing models  $(M_1, M_2, \dots, M_N)$ . After eliciting prior distributions on the parameters of each model, she can derive posterior estimates on all such parameters, and ultimately obtain  $N$  distinct predictive distributions, one for each model entertained. We denote with  $\{p(r_{t+1}|M_i, \mathcal{D}^t)\}_{i=1}^N$  the  $N$  predictive densities for  $r_{t+1}$ , where  $\mathcal{D}^t$  stands for the information set available at time  $t$ , i.e.  $\mathcal{D}^t = \{r_{\tau+1}, x_\tau\}_{\tau=1}^{t-1} \cup x_t$ . Next, using Bayesian Model Averaging (BMA, henceforth) the individual predictive densities are combined into a composite-weighted predictive distribution  $p(r_{t+1}|\mathcal{D}^t)$ , given by

$$p(r_{t+1}|\mathcal{D}^t) = \sum_{i=1}^N P(M_i|\mathcal{D}^t) p(r_{t+1}|M_i, \mathcal{D}^t) \quad (2)$$

where  $P(M_i|\mathcal{D}^t)$  is the posterior probability of model  $i$ , derived by Bayes' rule,

$$P(M_i|\mathcal{D}^t) = \frac{P(\mathcal{D}^t|M_i) P(M_i)}{\sum_{j=1}^N P(\mathcal{D}^t|M_j) P(M_j)}, \quad i = 1, \dots, N \quad (3)$$

and where  $P(M_i)$  is the prior probability of model  $M_i$ , with  $P(\mathcal{D}^t|M_i)$  denoting the corresponding marginal likelihood.<sup>5</sup> [Avramov \(2002\)](#) and [Dangl and Halling \(2012\)](#) apply BMA to forecast stock returns, and find that it leads to out-of-sample forecast improvements relative to the average performance of the individual models as well as, occasionally, relative to the performance of the best individual model.

We note, however, that BMA, as described in equations (2)-(3), suffers some important drawbacks. Perhaps the most important one is that BMA assumes that the true model is included in the model set. Indeed, under such an assumption, it can be shown that the combination weights in (3) converge (in the limit) to select the true model. However, as noted by [Diebold \(1991\)](#), all models could be false, and as a result the model set could be misspecified. [Geweke \(2010\)](#) labels this problem *model incompleteness*. As an alternative to BMA, [Geweke and Amisano \(2011\)](#) propose replacing the averaging as done in (2)-(3) with a linear prediction pool:

$$p(r_{t+1}|\mathcal{D}^t) = \sum_{i=1}^N w_i p(r_{t+1}|M_i, \mathcal{D}^t) \quad (4)$$

where the individual model weights  $w_i$  are computed by maximizing the log predictive likelihood, or log score ( $LS$ ), of the linear prediction pool:<sup>6</sup>

---

<sup>5</sup>See [Hoeting et al. \(1999\)](#) for a review on BMA.

<sup>6</sup>[Mitchell and Hall \(2005\)](#) discuss the analogy of the log score in a frequentistic framework to the log predictive likelihood in a Bayesian framework, and how it relates to the Kullback-Leibler divergence. See also [Hall and Mitchell \(2007\)](#), [Jore et al. \(2010\)](#), and [Geweke and Amisano \(2010\)](#) for a discussion on the use of the log score as a ranking device for the forecast ability of different models.

$$\sum_{\tau=1}^{t-1} \log \left[ \sum_{i=1}^N w_i \times \exp(LS_{i,\tau+1}) \right] \quad (5)$$

with  $LS_{i,\tau+1}$  denoting the recursively computed log score for model  $i$  at time  $\tau + 1$ . [Geweke and Amisano \(2011\)](#) and [Geweke and Amisano \(2012\)](#) show that the model weights, computed in this way, no longer converge to a unique solution, except in the case where there is a dominant model in terms of Kullback-Leibler divergence.

A second issue, common to both BMA and the linear prediction pool of [Geweke and Amisano \(2011\)](#), is the assumption that the model combination weights are constant over time. However, given the unstable and uncertain data-generating process for stock returns, it is conceivable to imagine that the combination weights may change over time. [Waggoner and Zha \(2012\)](#), [Billio et al. \(2013\)](#), and [Del Negro et al. \(2013\)](#) partly address this issue, proposing alternative combination methods featuring time-varying weights. Finally, a third and overarching issue with all the model combination methods described thus far is the presence of a disconnect between the metric according to which the individual forecasts are combined (i.e., either the marginal likelihood in (2) or the log score in (5)), and how ultimately the final combination is used. In particular, all model combination techniques described thus far weight individual models according to their statistical performance. While statistical performance may be the relevant metric to use in some settings, in the context of equity premium predictions this is likely not the case. On the contrary, when forecasting stock returns the quality of the individual model's predictions should be assessed in terms of whether ultimately such predictions lead to profitable investment decisions. This point has been emphasized before by [Leitch and Tanner \(1991\)](#), who show that good forecasts, as measured in terms of statistical criteria, do not necessarily translate into profitable portfolio allocations.

### 3 A novel model combination strategy

To address the limitations of the existing model combination methods discussed above, we introduce a novel model combination method that allows for model incompleteness and features time-varying combination weights, whose dynamics is driven by the profitability of the individual models entering the pool. We label our new approach Decision-Based Density Combination (DB-DeCo), in the spirit of [Pesaran and Skouras \(2007\)](#). In particular, our approach shares with [Billio et al. \(2013\)](#) and [Del Negro et al. \(2013\)](#) the feature that the model combination weights can change gradually over time. However, differently from these papers, we introduce a mechanism that allows the combination weights to depend on the whole history of the individual models' past profitability. We now turn to explaining in more details how our model combination method works.



We continue to assume that at a generic point in time  $t$ , the investor has available  $N$  different models to predict excess returns at time  $t + 1$ , each model producing a predictive distribution  $p(r_{t+1}|M_i, \mathcal{D}^t)$ ,  $i = 1, \dots, N$ . For example, the investor may be considering  $N$  alternative predictors for stock returns, leading to  $N$  univariate models, each one in the form of (1) and including as right-hand-side one of the  $N$  available predictors. To ease the notation, we aggregate the  $N$  predictive distributions  $\{p(r_{t+1}|M_i, \mathcal{D}^t)\}_{i=1}^N$  into the pdf  $p(\tilde{\mathbf{r}}_{t+1}|\mathcal{D}^t)$ . Next, the composite predictive distribution  $p(r_{t+1}|\mathcal{D}^t)$  is given by

$$p(r_{t+1}|\mathcal{D}^t) = \int p(r_{t+1}|\tilde{\mathbf{r}}_{t+1}, \mathbf{w}_{t+1}, \mathcal{D}^t) p(\mathbf{w}_{t+1}|\tilde{\mathbf{r}}_{t+1}, \mathcal{D}^t) p(\tilde{\mathbf{r}}_{t+1}|\mathcal{D}^t) d\tilde{\mathbf{r}}_{t+1} d\mathbf{w}_{t+1} \quad (6)$$

where  $p(r_{t+1}|\tilde{\mathbf{r}}_{t+1}, \mathbf{w}_{t+1}, \mathcal{D}^t)$  denotes the combination scheme based on the  $N$  predictive densities  $\tilde{\mathbf{r}}_{t+1}$  and the combination weights  $\mathbf{w}_{t+1} \equiv (w_{1,t+1}, \dots, w_{N,t+1})'$ , and  $p(\mathbf{w}_{t+1}|\tilde{\mathbf{r}}_{t+1}, \mathcal{D}^t)$  denotes the posterior distribution of the combination weights  $\mathbf{w}_{t+1}$ . Equation (6) generalizes equation (2), taking into account the limitations discussed in the previous section. First, by specifying a stochastic process for the model combination scheme,  $p(r_{t+1}|\tilde{\mathbf{r}}_{t+1}, \mathbf{w}_{t+1}, \mathcal{D}^t)$ , our approach explicitly allows for either model misspecification or model incompleteness to play a role. Second, by introducing a proper distribution for the model combination weights  $\mathbf{w}_{t+1}$ ,  $p(\mathbf{w}_{t+1}|\tilde{\mathbf{r}}_{t+1}, \mathcal{D}^t)$ , we gain two important advantages. On the one hand, our method can allow for time-varying combination weights. On the other hand, we have flexibility in how to model the dependence of the combination weights on the individual models' performance, and are no longer confined to have the weights depend on some measure of the individual models' statistical fit. We note, *inter alia*, that in addition to addressing the limitations discussed above, the combination scheme in (6) allows to factor into the composite predictive distribution the uncertainty over the model combination weights, a feature that should prove useful in the context of excess return predictions, where there is significant uncertainty over the identity of the best model(s) for predicting returns. We now turn to describing in more details how the individual terms in (6) are obtained.

### 3.1 Individual models

We begin by explaining how we specify the last term on the right-hand side of (6),  $p(\tilde{\mathbf{r}}_{t+1}|\mathcal{D}^t)$ , which we remind is short-hand for the set of individual predictive distributions  $\{p(r_{t+1}|M_i, \mathcal{D}^t)\}_{i=1}^N$  entering the model combination. As previously discussed, most of the literature on stock return predictability focuses on linear models, so we take this class of models as our starting point. In this way, it will be easier to compare the results of our model combination method with the findings from the existing studies, such as for example [Welch and Goyal \(2008\)](#), [Campbell and Thompson \(2008\)](#), and [Rapach et al. \(2010\)](#).

The linear model projects excess returns  $r_{\tau+1}$  on a lagged predictor,  $x_\tau$ , where  $x_\tau$  can be a

scalar or a vector of regressors<sup>7</sup>

$$\begin{aligned} r_{\tau+1} &= \mu + \beta x_{\tau} + \varepsilon_{\tau+1}, \quad \tau = 1, \dots, t-1, \\ \varepsilon_{\tau+1} &\sim N(0, \sigma_{\varepsilon}^2). \end{aligned} \quad (7)$$

To estimate the model in (7), we rely on a Gibbs sampler, which permit us to form a number of draws from the posterior distributions of  $\mu$ ,  $\beta$ , and  $\sigma_{\varepsilon}^{-2}$ , given the information set available at time  $t$ ,  $\mathcal{D}^t$ . Once draws from the posterior distributions of  $\mu$ ,  $\beta$ , and  $\sigma_{\varepsilon}^{-2}$  are available, we use them to form a predictive density for  $r_{t+1}$  in the following way:

$$p(r_{t+1} | M_i, \mathcal{D}^t) = \int_{\mu, \beta, \sigma_{\varepsilon}^{-2}} p(r_{t+1} | \mu, \beta, \sigma_{\varepsilon}^{-2}, M_i, \mathcal{D}^t) p(\mu, \beta, \sigma_{\varepsilon}^{-2} | M_i, \mathcal{D}^t) d\mu d\beta d\sigma_{\varepsilon}^{-2}. \quad (8)$$

Repeating this process for the  $N$  individual models entering the model combination yields the set of  $N$  individual predictive distributions  $\{p(r_{t+1} | M_i, \mathcal{D}^t)\}_{i=1}^N$ . We refer the reader to [Appendix B](#) for more details on the the Gibbs sampler we implement and on how we compute the integral in equation (8).

### 3.2 Combination weights

We now turn to describing how we specify the conditional density for the combination weights,  $p(\mathbf{w}_{t+1} | \tilde{\mathbf{r}}_{t+1}, \mathcal{D}^t)$ . First, in order to have the weights  $\mathbf{w}_{t+1}$  belong to the simplex  $\Delta_{[0,1]^N}$ , we introduce a vector of latent processes  $\mathbf{z}_{t+1} = (z_{1,t+1}, \dots, z_{N,t+1})'$ , where  $N$  is the total number of models considered in the combination scheme, and we specify the multivariate transform  $\mathbf{g} = (g_1, \dots, g_N)'$ ,<sup>8</sup>

$$\mathbf{g} : \begin{cases} \mathbb{R}^N & \rightarrow \Delta_{[0,1]^N} \\ \mathbf{z}_{t+1} & \mapsto \mathbf{w}_{t+1} = (g_1(z_{1,t+1}), \dots, g_N(z_{N,t+1}))' \end{cases} \quad (9)$$

Next, in order to obtain the combination weights we need to make additional assumptions on how the vector of latent processes  $\mathbf{z}_{t+1}$  evolves over time and how it maps into the combination weights  $\mathbf{w}_{t+1}$ . One possibility is to specify a Gaussian random walk process for  $\mathbf{z}_{t+1}$ ,<sup>9</sup>

$$\begin{aligned} \mathbf{z}_{t+1} &\sim p(\mathbf{z}_{t+1} | \mathbf{z}_t, \Lambda) \\ &\propto |\Lambda|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{z}_{t+1} - \mathbf{z}_t)' \Lambda^{-1} (\mathbf{z}_{t+1} - \mathbf{z}_t) \right\} \end{aligned} \quad (10)$$

<sup>7</sup>In our setting we consider only one predictor at the time, thus  $x_t$  is a scalar. It would be possible to include multiple predictors, but we follow the bulk of the literature on stock return predictability and focus on a single predictor.

<sup>8</sup>Under this convexity constraint, the weights can be interpreted as discrete probabilities over the set of models entering the combination.

<sup>9</sup>We assume that the variance-covariance matrix  $\Lambda$  of the process  $\mathbf{z}_{t+1}$  governing the combination weights is diagonal. We leave for further research the possibility of allowing for cross-correlation between model weights.

with  $\Lambda$  an  $(N \times N)$  diagonal matrix, and have the combination weights computed as

$$w_{i,t+1} = \frac{\exp\{z_{i,t+1}\}}{\sum_{l=1}^N \exp\{z_{l,t+1}\}}, \quad i = 1, \dots, N \quad (11)$$

Effectively, equations (10) and (11) implies time-varying combination weights, where time  $t + 1$  combination weights depend in a non-linear fashion on time  $t$  combination weights. Alternatively, we could allow the combination weights to depend on the past performance of the  $N$  individual prediction models entering the combination. To accomplish this, we modify the stochastic process for  $\mathbf{z}_{t+1}$  in (10) as follows:

$$\begin{aligned} \mathbf{z}_{t+1} &\sim p(\mathbf{z}_{t+1} | \mathbf{z}_t, \Delta\zeta_t, \Lambda) \\ &\propto |\Lambda|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{z}_{t+1} - \mathbf{z}_t - \Delta\zeta_t)' \Lambda^{-1} (\mathbf{z}_{t+1} - \mathbf{z}_t - \Delta\zeta_t) \right\} \end{aligned} \quad (12)$$

where  $\Delta\zeta_t = \zeta_t - \zeta_{t-1}$ , with  $\zeta_t = (\zeta_{1,t}, \dots, \zeta_{N,t})'$  denoting a distance vector, measuring the accuracy of the  $N$  prediction models up to time  $t$ . We opt for an exponentially weighted moving average of the past performance of the  $N$  individual models entering the combination,

$$\zeta_{i,t} = (1 - \lambda) \sum_{\tau=\underline{t}}^t \lambda^{t-\tau} f(r_\tau, \tilde{\mathbf{r}}_{i,\tau}), \quad i = 1, \dots, N \quad (13)$$

where  $\underline{t}$  denotes the beginning of the evaluation period. In other words, we are proposing to have the combination weight of model  $i$  depend on an exponentially weighted sum of the last observed ( $\tau = t$ ) and past history ( $\tau < t$ ) of model  $i$ , where  $\lambda \in (0, 1)$  is a smoothing parameter,  $f(r_\tau, \tilde{\mathbf{r}}_{i,\tau})$  is a measure of the accuracy of model  $i$ , and  $\tilde{\mathbf{r}}_{i,\tau}$  denotes the one-step ahead density forecast of  $r_\tau$  made by model  $i$  at time  $\tau - 1$ .  $\tilde{\mathbf{r}}_{i,\tau}$  is thus short-hand for the  $i$ -th element of  $p(\tilde{\mathbf{r}}_\tau | \mathcal{D}^{\tau-1})$ ,  $p(r_\tau | M_i, \mathcal{D}^{\tau-1})$ .

As for the specific choice of  $f(r_\tau, \tilde{\mathbf{r}}_{i,\tau})$ , given our ultimate interest in the profitability of stock return predictions, we focus on a utility-based measure of predictability, the certainty equivalent return (CER). In the case of a power utility investor who at time  $\tau - 1$  chooses a portfolio by allocating her wealth  $W_{\tau-1}$  between the riskless asset and one risky asset, and subsequently holds onto that investment for one period, her CER is given by

$$f(r_\tau, \tilde{\mathbf{r}}_{i,\tau}) = [(1 - A) U(W_{i,\tau}^*)]^{1/(1-A)} \quad (14)$$

where  $U(W_{i,\tau}^*)$  denotes the investor's realized utility at time  $\tau$ ,

$$U(W_{i,\tau}^*) = \frac{\left[ (1 - \omega_{i,\tau-1}^*) \exp(r_{\tau-1}^f) + \omega_{i,\tau-1}^* \exp(r_{\tau-1}^f + r_\tau) \right]^{1-A}}{1 - A} \quad (15)$$

$r_{\tau-1}^f$  denotes the continuously compounded Treasury bill rate at time  $\tau - 1$ ,  $A$  stands for the investor's relative risk aversion,  $r_\tau$  is the realized excess return at time  $\tau$ , and  $\omega_{i,\tau-1}^*$  denotes the optimal allocation to stocks according to the prediction made for  $r_\tau$  by model  $M_i$ ,

$$\omega_{i,\tau-1}^* = \arg \max_{\omega_{\tau-1}} \int U(\omega_{\tau-1}, r_\tau) p(r_\tau | M_i, \mathcal{D}^{\tau-1}) dr_\tau \quad (16)$$

By replacing equation (10) with (12) and (13), we include the exponentially weighted learning strategy into the weight dynamics and estimate the density of  $\mathbf{z}_{t+1}$  accounting for the whole history of certainty equivalence returns given in Eq. (14). Indeed, note that equation (12) could be rewritten as

$$\mathbf{z}_{t+1} = \mathbf{z}_t + \Delta\zeta_t + \mathbf{v}_{t+1}, \quad (17)$$

where  $\mathbf{v}_{t+1} \sim iid \mathcal{N}(\mathbf{0}, \Lambda)$ . Recursive substitution on (17) all the way to the beginning of the forecast evaluation period  $\underline{t}$  yields

$$z_{i,t+1} = z_{i,\underline{t}} + (1 - \lambda) \sum_{\tau=\underline{t}}^t \lambda^{t-\tau} f(r_\tau, \tilde{\mathbf{r}}_{i,\tau}) + \sum_{\tau=\underline{t}}^t v_{i,\tau+1}, \quad i = 1, \dots, N \quad (18)$$

where  $z_{i,t+1}$ ,  $z_{i,\underline{t}}$  and  $v_{i,\tau+1}$  are the  $i$ -th elements of  $\mathbf{z}_{t+1}$ ,  $\mathbf{z}_{\underline{t}}$  and  $\mathbf{v}_{\tau+1}$ , respectively. Equation (18) clearly conveys the point that  $z_{i,t+1}$  depends on an exponentially weighted sum of the entire past history of model  $i$ 's performance,  $(1 - \lambda) \sum_{\tau=\underline{t}}^t \lambda^{t-\tau} f(r_\tau, \tilde{\mathbf{r}}_{\tau,i})$ , as well as on the whole history of stochastic shocks,  $\sum_{\tau=\underline{t}}^t \mathbf{v}_{i,\tau+1}$ .

In practice, to estimate  $p(\mathbf{w}_{t+1} | \tilde{\mathbf{r}}_{t+1}, \mathcal{D}^t)$  from (12) and (11), we first need to specify the combination scheme  $p(r_{t+1} | \tilde{\mathbf{r}}_{t+1}, \mathbf{w}_{t+1}, \mathcal{D}^t)$ , so we postpone the discussion on how we estimate  $p(\mathbf{w}_{t+1} | \tilde{\mathbf{r}}_{t+1}, \mathcal{D}^t)$  until the end of the next subsection.

### 3.3 Combination scheme

We now turn to the first term on the right hand side of (6),  $p(r_{t+1} | \tilde{\mathbf{r}}_{t+1}, \mathbf{w}_{t+1}, \mathcal{D}^t)$ , denoting the combination scheme adopted in our model combination. We note that since both the  $N$  original densities  $\{p(r_{t+1} | M_i, \mathcal{D}^t)\}_{i=1}^N$  and the combination weights  $\mathbf{w}_{t+1}$  are in the form of densities, the combination scheme for  $p(r_{t+1} | \tilde{\mathbf{r}}_{t+1}, \mathbf{w}_{t+1}, \mathcal{D}^t)$  is based on a convolution mechanism. Precisely, we follow Billio et al. (2013), and apply a Gaussian combination scheme,

$$p(r_{t+1} | \tilde{\mathbf{r}}_{t+1}, \mathbf{w}_{t+1}, \sigma_\kappa^{-2}) \propto \exp \left\{ -\frac{1}{2} (r_{t+1} - \tilde{\mathbf{r}}_{t+1} \mathbf{w}_{t+1})' \sigma_\kappa^{-2} (r_{t+1} - \tilde{\mathbf{r}}_{t+1} \mathbf{w}_{t+1}) \right\} \quad (19)$$

The combination relationship is assumed to be linear and explicitly allows for model misspecification, possibly because all models in the combination may be false (incomplete model set or open model space). The combination residuals are estimated and their distribution follows a Gaussian process with mean zero and standard deviation  $\sigma_\kappa$ , providing a probabilistic measure of the incompleteness of the model set.<sup>10</sup> In other words, equation (19) can be rewritten as:

---

<sup>10</sup>We note that our method is thus more general than the approach in Geweke and Amisano (2010) and Geweke and Amisano (2011), as it provides as an output a measure of model incompleteness.

$$p(r_{t+1}|\tilde{\mathbf{r}}_{t+1}, \mathbf{w}_{t+1}, \sigma_{\kappa}^{-2}) = \tilde{\mathbf{r}}_{t+1} \mathbf{w}_{t+1} + \kappa_{t+1} \quad (20)$$

with  $\kappa_{t+1} \sim \mathcal{N}(0, \sigma_{\kappa}^2)$ . The convolution mechanism previously described guarantees that the product of the densities  $\tilde{\mathbf{r}}_{t+1}$  and  $\mathbf{w}_{t+1}$  is a proper density. It is also worth pointing out that when the randomness is canceled out by fixing  $\sigma_{\kappa}^2 = 0$  and the weights are derived as in equation (3), the combination in (6) reduces to standard BMA. Hence, one can think of BMA as a special case of the combination approach we propose here. We refer the reader to [Appendix A](#) and [Aastveit et al. \(2014\)](#) for further discussion on convolution and its properties.

We conclude this section by briefly describing how we estimate the posterior distributions  $p(r_{t+1}|\tilde{\mathbf{r}}_{t+1}, \mathbf{w}_{t+1}, \mathcal{D}^t)$  and  $p(\mathbf{w}_{t+1}|\tilde{\mathbf{r}}_{t+1}, \mathcal{D}^t)$ . Equations (6), (11), (12), and (19), as well as the individual model predictive densities  $p(\tilde{\mathbf{r}}_{t+1}|\mathcal{D}^t)$  are first grouped into a non-linear state space model.<sup>11</sup> Because of the non-linearity, standard Gaussian methods such as the Kalman filter cannot be applied. We instead apply a Sequential Monte Carlo method, using a particle filter to approximate the transition equation governing the dynamics of  $\mathbf{z}_{t+1}$  in the state space model, yielding posterior distributions for both  $p(r_{t+1}|\tilde{\mathbf{r}}_{t+1}, \mathbf{w}_{t+1}, \mathcal{D}^t)$  and  $p(\mathbf{w}_{t+1}|\tilde{\mathbf{r}}_{t+1}, \mathcal{D}^t)$ . For additional details, see [Appendix C](#).

## 4 Data and priors

In this section we describe the data used in the empirical analysis and the prior choices we made.

### 4.1 Data

Our empirical analysis uses data on stock returns along with a set of fifteen predictor variables originally analyzed in [Welch and Goyal \(2008\)](#) and subsequently extended up to 2010 by the same authors. Stock returns are computed from the S&P500 index and include dividends. A short T-bill rate is subtracted from stock returns in order to capture excess returns. Data samples vary considerably across the individual predictor variables. To be able to compare results across the individual predictor variables, we use the longest common sample, that is 1927-2010. In addition, we use the first 20 years of data as a training sample. Specifically, we initially estimate our regression models over the period January 1927–December 1946, and use the estimated coefficients to forecast excess returns for January 1947. We next include January 1947 in the estimation sample, which thus becomes January 1927–January 1947, and use the corresponding estimates to predict excess returns for February 1947. We proceed in this recursive fashion

---

<sup>11</sup>The non-linearity is due to the logistic transformation mapping the latent process  $\mathbf{z}_{t+1}$  into the model combination weights  $\mathbf{w}_{t+1}$ .

until the last observation in the sample, thus producing a time series of one-step-ahead forecasts spanning the time period from January 1947 to December 2010.

The identity of the predictor variables, along with summary statistics, is provided in Table 1. Most variables fall into three broad categories, namely (i) valuation ratios capturing some measure of ‘fundamentals’ to market value such as the dividend yield, the earnings-price ratio, the 10-year earnings-price ratio or the book-to-market ratio; (ii) measures of bond yields capturing level effects (the three-month T-bill rate and the yield on long-term government bonds), slope effects (the term spread), and default risk effects (the default yield spread defined as the yield spread between BAA and AAA rated corporate bonds, and the default return spread defined as the difference between the yield on long-term corporate and government bonds); (iii) estimates of equity risk such as the long-term return and stock variance (a volatility estimate based on daily squared returns); (iv) three corporate finance variables, namely the dividend payout ratio (the log of the dividend-earnings ratio), net equity expansion (the ratio of 12-month net issues by NYSE-listed stocks over the year-end market capitalization), and the percentage of equity issuance (the ratio of equity issuing activity as a fraction of total issuing activity). Finally, we consider a macroeconomic variable, inflation, defined as the rate of change in the consumer price index, and the net payout measure of Boudoukh et al. (2007), which is computed as the ratio between dividends and net equity repurchases (repurchases minus issuances) over the last twelve months and the current stock price. Johannes et al. (2014) find that accounting for net equity repurchases in addition to cash payouts produces a stronger predictor for equity returns.<sup>12</sup>

## 4.2 Priors

As described at the outset, we have chosen to adopt a Bayesian approach in this paper, so we briefly describe how the priors are specified. We start with the priors on the parameters of the individual models,  $\mu$ ,  $\beta$ , and  $\sigma_\varepsilon^{-2}$ . Following standard practice, the priors for the parameters  $\mu$  and  $\beta$  in (7) are assumed to be normal and independent of  $\sigma_\varepsilon^2$ ,<sup>13</sup>

$$\begin{bmatrix} \mu \\ \beta \end{bmatrix} \sim N(\underline{b}, \underline{V}), \quad (21)$$

where

$$\underline{b} = \begin{bmatrix} \bar{r}_t \\ 0 \end{bmatrix}, \quad \underline{V} = \underline{\psi}^2 \left[ s_{r,t}^2 \left( \sum_{\tau=1}^{t-1} x_\tau x_\tau' \right)^{-1} \right], \quad (22)$$

---

<sup>12</sup>We follow Welch and Goyal (2008) and lag inflation an extra month to account for the delay in CPI releases.

<sup>13</sup>See for example Koop (2003), Section 4.2.

and data-based moments:

$$\begin{aligned}\bar{r}_t &= \frac{1}{t-1} \sum_{\tau=1}^{t-1} r_{\tau+1}, \\ s_{r,t}^2 &= \frac{1}{t-2} \sum_{\tau=1}^{t-1} (r_{\tau+1} - \bar{r}_t)^2.\end{aligned}$$

Our choice of the prior mean vector  $\underline{b}$  reflects the “no predictability” view that the best predictor of stock excess returns is the average of past returns. We therefore center the prior intercept on the prevailing mean of historical excess returns, while the prior slope coefficient is centered on zero.<sup>14</sup>

In (22),  $\underline{\psi}$  is a constant that controls the tightness of the prior, with  $\underline{\psi} \rightarrow \infty$  corresponding to a diffuse prior on  $\mu$  and  $\beta$ . Our benchmark analysis sets  $\underline{\psi} = 1$ . We assume a standard gamma prior for the error precision of the return innovation,  $\sigma_\varepsilon^{-2}$ :

$$\sigma_\varepsilon^{-2} \sim \mathcal{G}(s_{r,t}^{-2}, \underline{v}_0(t-1)), \quad (23)$$

where  $\underline{v}_0$  is a prior hyperparameter that controls the degree of informativeness of this prior, with  $\underline{v}_0 \rightarrow 0$  corresponding to a diffuse prior on  $\sigma_\varepsilon^{-2}$ .<sup>15</sup> Our baseline analysis sets  $\underline{v}_0 = 1$ .

Moving on to the processes controlling the combination weights and the combination scheme, we need to specify priors for  $\sigma_\kappa^{-2}$  and for the diagonal elements of  $\Lambda$ . The prior for  $\sigma_\kappa^{-2}$ , the precision of our measure of incompleteness in the combination scheme, and the diagonal elements of  $\Lambda^{-1}$ , the precision matrix of the process  $\mathbf{z}_{t+1}$  governing the combination weights  $\mathbf{w}_{t+1}$ , are assumed to be gamma,  $\mathcal{G}(\underline{s}_{\sigma_\kappa}^{-2}, \underline{v}_{\sigma_\kappa}(t-1))$  and  $\mathcal{G}(\underline{s}_\Lambda^{-1}, \underline{v}_\Lambda(t-1))$ , respectively. We set informative values on our prior beliefs regarding the incompleteness and the combination weights. Precisely, we set  $\underline{v}_{\sigma_\kappa} = \underline{v}_{\Lambda_i} = 1$  and set the hyperparameters controlling the means of the prior distributions to  $\underline{s}_{\sigma_\kappa}^{-2} = 1000$ , shrinking the model incompleteness to zero, and to  $\underline{s}_\Lambda^{-1} = 4$ , allowing  $\mathbf{z}_{t+1}$  to evolve freely over time and differ from the initial value  $\mathbf{z}_0$ , set to equal weights.<sup>16</sup>

## 5 Out-of-Sample Predictive Performance

In this section we answer the question of whether the DB-DeCo method produces equity premium forecasts that are more accurate than those obtained from the existing approaches. We compare

<sup>14</sup>It is common to base the priors of some of the hyperparameters on sample estimates—see [Stock and Watson \(2006\)](#) and [Efron \(2010\)](#)—and our analysis can be viewed as an empirical Bayes approach rather than a more traditional Bayesian approach that fixes the prior distribution before any data are observed.

<sup>15</sup>Following [Koop \(2003\)](#), we adopt the Gamma distribution parametrization of [Poirier \(1995\)](#). Namely, if the continuous random variable  $Y$  has a Gamma distribution with mean  $\mu > 0$  and degrees of freedom  $v > 0$ , we write  $Y \sim \mathcal{G}(\mu, v)$ . In this case,  $E(Y) = \mu$  and  $Var(Y) = 2\mu^2/v$ .

<sup>16</sup>In our empirical application,  $N$  is set to 15 therefore  $z_{0,i} = \ln(1/15) = -2.71$  resulting in  $w_{0,i} = 1/15$ . The prior choices we made for the diagonal elements of  $\Lambda$  allow the posterior weights on the individual models to differ substantially from equal weights. See section 8 for alternative prior specifications.

the performance of DB-DeCo to both the fifteen univariate models entering the combination as well as a number of alternative model combination methods. As in [Welch and Goyal \(2008\)](#) and [Campbell and Thompson \(2008\)](#), the predictive performance of each model is measured relative to the prevailing mean (PM) model. One of the advantages of adopting a Bayesian framework in this work is the ability to compute predictive distributions, rather than simple point forecasts, which incorporate parameter uncertainty. Accordingly, to shed light on the predictive ability of the various models, we consider several evaluation statistics for both point and density forecasts.

As for assessing the accuracy of the point forecasts, the first measure we consider is the Cumulative Sum of Squared prediction Error Difference (CSSED) introduced by [Welch and Goyal \(2008\)](#),

$$CSSED_{m,t} = \sum_{\tau=\underline{t}}^t (e_{PM,\tau}^2 - e_{m,\tau}^2) \quad (24)$$

where  $m$  denotes the model under consideration (either univariate or model combination),  $\underline{t}$  denotes the beginning of the forecast evaluation period, and  $e_{m,t}$  ( $e_{PM,\tau}$ ) denotes model  $m$ 's (PM's) prediction error from time  $\tau$  forecasts, obtained by synthesizing the predictive density  $p(r_\tau | M_i, \mathcal{D}^{\tau-1})$  (or  $p(r_\tau | \mathcal{D}^{\tau-1})$  in the case of model combinations) into a point forecast. An increase from  $CSSED_{m,t-1}$  to  $CSSED_{m,t}$  indicates that relative to the benchmark PM model, the alternative model  $m$  predicts more accurately at observation  $t$ . Following [Campbell and Thompson \(2008\)](#), we also summarize the predictive ability of the various models over the whole evaluation sample by reporting the out-of-sample  $R^2$  measure,

$$R_{OoS,m}^2 = 1 - \frac{\sum_{\tau=\underline{t}}^{\bar{t}} e_{m,\tau}^2}{\sum_{\tau=\underline{t}}^{\bar{t}} e_{PM,\tau}^2}. \quad (25)$$

whereby a positive  $R_{OoS,m}^2$  is indicative of some predictability from model  $m$  (again, relative to the benchmark PM model), and where  $\bar{t}$  denotes the end of the forecast evaluation period.

Turning next to the accuracy of the density forecasts, we consider two different metrics of predictive performance. First, following [Amisano and Giacomini \(2007\)](#), [Geweke and Amisano \(2010\)](#), and [Hall and Mitchell \(2007\)](#), we consider the average log score differential,

$$LSD_m = \frac{\sum_{\tau=\underline{t}}^{\bar{t}} (LS_{m,\tau} - LS_{PM,\tau})}{\sum_{\tau=\underline{t}}^{\bar{t}} LS_{PM,\tau}} \quad (26)$$

where  $LS_{m,\tau}$  ( $LS_{PM,\tau}$ ) denotes model  $m$ 's (PM's) log predictive score computed at time  $\tau$ . If  $LSD_m$  is positive, this indicates that on average the alternative model  $m$  produces more accurate density forecasts than the benchmark prevailing mean model (PM). We also consider using the recursively computed log scores as inputs to the period  $t$  difference in the cumulative log score



differential between the PM model and the  $m$ th model,

$$CLSD_{m,t} = \sum_{\tau=\underline{t}}^t (LS_{m,\tau} - LS_{PM,\tau}) \quad (27)$$

An increase from  $CLSD_{m,t-1}$  to  $CLSD_{m,t}$  indicates that relative to the benchmark PM model, the alternative model  $m$  predicts more accurately at observation  $t$ . Next, we follow [Gneiting and Raftery \(2007\)](#), [Gneiting and Ranjan \(2011\)](#) and [Groen et al. \(2013\)](#), and consider the average continuously ranked probability score differential (CRPSD),

$$CRPSD_m = \frac{\sum_{\tau=\underline{t}}^{\bar{t}} (CRPS_{PM,\tau} - CRPS_{m,\tau})}{\sum_{\tau=\underline{t}}^{\bar{t}} CRPS_{PM,\tau}} \quad (28)$$

where  $CRPS_{m,\tau}$  ( $CRPS_{PM,\tau}$ ) measures the average distance between the empirical cumulative distribution function (CDF) of  $r_\tau$  (which is simply a step function in  $r_\tau$ ), and the empirical CDF that is associated with model  $m$ 's (PM's) predictive density. [Gneiting and Raftery \(2007\)](#) explain how the CRPSD measure circumvents some of the problems of the logarithmic score, most notably the fact that the latter does not reward values from the predictive density that are close but not equal to the realization. Finally, we consider using the recursively computed continuously ranked probability score as inputs to the period  $t$  difference in the cumulative continuously ranked probability score differential between the PM model and the  $m$ th model,

$$CCRPSD_{m,t} = \sum_{\tau=\underline{t}}^t (CRPS_{PM,\tau} - CRPS_{m,\tau}) \quad (29)$$

An increase from  $CCRPSD_{m,t-1}$  to  $CCRPSD_{m,t}$  indicates that relative to the benchmark PM model, the alternative model  $m$  predicts more accurately at observation  $t$ .

## 5.1 Empirical results

Table 2 reports  $R_{OoS}^2$ -values for both the fifteen univariate models (top panel) and a variety of model combination methods, including the DB-DeCo approach introduced in Section 3 (bottom panel). Positive values suggest that the alternative models perform better than the PM model. We also report stars to summarize the statistical significance of the  $R_{OoS}^2$ -values, where the underlying p-values are based on the [Diebold and Mariano \(1995\)](#) t-statistics for equality of the root mean squared forecast errors (RMSFE) of the competing models and are computed with a serial correlation-robust variance, using the pre-whitened quadratic spectral estimator of [Andrews and Monahan \(1992\)](#). We begin by focusing on the results under the column header "Linear". We will return later to the remaining half of the table. Starting with the top panel, the results for the individual models are reminiscent of the findings of [Welch and Goyal \(2008\)](#), where the  $R_{OoS}^2$ -values are negative for 13 out of the 15 predictor variables. Moving on to bottom panel

of the table, we find that with the exception of the optimal prediction pool method of [Geweke and Amisano \(2011\)](#), accounting for model uncertainty leads to positive  $R_{OoS}^2$ -values. We note in particular that the DB-DeCo method yields the largest improvement in forecast performance among all model combination methods, with an  $R_{OoS}^2$  of 2.32%, statistically significant at the 1% level. This is almost two percentage points higher than all other model combination methods.

To shed light on the sources of such improvement in predictability, we also compute a version of DB-DeCo where we suppress the learning mechanism in the weight dynamics (that is, we replace equation (12) with (10)). We label this combination scheme “Density Combination”. A quick look at the comparison between DB-DeCo and the Density Combination method in Table 2 reveals that the learning mechanism introduced via equations (12)-(14) explains a large fraction of the increase in performance we see for the DB-DeCo method.

We next turn to Table 3, which reports the density forecast performance for the same set of models listed in Table 2. Focusing on the columns under the header “Linear”, we find that the DB-DeCo method is the only model combination method that yields positive and statistically significant results (as in Table 2, the underlying p-values are based on the [Diebold and Mariano \(1995\)](#) t-statistics). This is true for both measures of density forecast accuracy, the average log score differential and the average CRPS differential. Finally, Figures 1-3 plot the  $CSSED_t$  (Figure 1),  $CLSD_t$  (Figure 2), and  $CCRPD_t$  (Figure 3) for the various model combination methods considered in this study. These plots show periods where the various models perform well relative to the PM model - periods where the lines are increasing and above zero - and periods where the models underperform against this benchmark - periods with decreasing lines. All three figures show that the DB-DeCo model consistently outperforms the benchmark model as well as all the alternative model combination methods over the whole out-of-sample period. Once again, the effect of learning is quite evident, as shown by the gaps between the two lines in the fourth panel of each figure.

## 6 Economic Performance

So far we have focused on the statistical performance of the forecasts from the various models. We next evaluate the economic significance of these return forecasts by considering the optimal portfolio choice of an investor who uses the return forecasts to guide her investment decisions. As mentioned earlier, one advantage of adopting a Bayesian approach is that it yields predictive densities that account for parameter estimation error.<sup>17</sup> Another related point is that having available the full predictive densities means that we are not reduced to considering only mean-variance utility but can use utility functions with better properties.

---

<sup>17</sup>The importance of controlling for parameter uncertainty in investment decisions has been emphasized by [Kandel and Stambaugh \(1996\)](#) and [Barberis \(2000\)](#). [Klein and Bawa \(1976\)](#) were among the first to note that using estimates for the parameters of the return distribution to construct portfolios induces an estimation risk.

Having computed the optimal asset allocation weights for both the individual models ( $M_1, \dots, M_N$ ) and the various model combinations, we assess the economic predictability of all models by computing their implied (annualized) CER. Under power utility, the investor's annualized CER is given by

$$CER_m = 12 \times \left[ (1 - A) \frac{1}{t^*} \sum_{\tau=\underline{t}}^{\bar{t}} U(W_{m,\tau}^*) \right]^{1/(1-A)} - 1 \quad (30)$$

where  $m$  denotes the model under consideration (either univariate or model combination), and  $t^* = \bar{t} - \underline{t} + 1$ . We next define the differential certainty equivalent return of model  $m$ , relative to the benchmark prevailing mean model  $PM$ ,

$$\Delta CER_m = CER_m - CER_{PM}. \quad (31)$$

A positive  $\Delta CER_m$  can be interpreted as evidence that model  $m$  generates a higher (certainty equivalent) return than the benchmark model.

## 6.1 Empirical results

Table 4 shows annualized CER values for the same models listed in Tables 2 and 3, assuming a coefficient of relative risk aversion of  $A = 5$ . Positive values suggest that the alternative model (either the individual models in the top panels or the model combinations in the bottom panel) performs better than the PM model. Once again, we focus on the column under the header ‘‘Linear’’. An inspection of the bottom panel of Table 4 reveals that the statistical gains we saw for the DB-DeCo approach in Tables 2 and 3 translate into CER gains of almost 100 basis points, relative to the PM model. No other combination scheme provides gains of a magnitude comparable to the DB-DeCo scheme. A comparison between the Decision-Based Density Combination and the Density Combination methods reveals that it is the learning mechanism introduced via equations (12)-(14) that drives this result. Turning to the top panel of Table 4, it appears that some of the individual models provide positive CER figures, but in terms of magnitude these gains are at least 50 basis points smaller than the DB-DeCo method. Figure 4 plots the cumulative CER values, computed relative to the PM benchmark. These plots parallel the cumulated differential plots of Figures 1-3, the key difference being that Figure 4 shows the cumulative risk-adjusted return from using a particular model combination method, relative to the PM model. The figure shows how the economic performance of the DB-DeCo model is not the result of any specific and short-lived episode, but rather it is built gradually over the entire out-of-sample period, as indicated by the the constantly increasing red dashed line in the fourth panel of Table 4. The only exception is during the second part of the 1990s, where the PM benchmark appears to outperform the DB-DeCo model. Figure 4 also indicates that the DB-DeCo cumulated CER value at the end of the sample exceeds 40 percent, while it is

negative for all other model combination methods. In addition, the effect of learning (obtained by comparing the heights of the CERs lines belonging to the Density Combination method and the DB-DeCo method at the end of the sample) appears to be around 55 percentage points.

## 7 Capturing Parameter Instability: A Time-Varying Parameter Stochastic Volatility Model

Parameter instability is a very important issue in the context of equity premium predictions and several studies have found a distinctly time-varying and unstable nature of the return predictability. See for example [Henkel et al. \(2011\)](#), [Paye and Timmermann \(2006\)](#), and [Pettenuzzo and Timmermann \(2011\)](#). While it is well known that forecast combination methods can deal with model instabilities and structural breaks and can generate more stable forecasts than those from the individual models (see for example [Hendry and Clements \(2004\)](#), and [Stock and Watson \(2004\)](#)), the impact of the linearity assumption on the individual models entering the model combination is an aspect that has not yet been thoroughly investigated.

There are many reasons why one may suspect that the linear model in (7) could be misspecified. For one, it is very likely that the regression parameters in (7) may vary over time. Parameter instability is present in a wide array of macroeconomic and financial time series (see, e.g., the comprehensive analyses of [Stock and Watson \(2006\)](#), and [Ang and Timmermann \(2012\)](#)), and there is no reason to believe that this should not represent an issue with inferences and forecasting in the setting of return predictability where, due to a particularly low signal-to-noise ratio of the predictive regressions, researchers often prefer to employ data spanning several decades in order to extract more precise parameter estimates. Similarly, the baseline model in (7) assumes that return volatility is constant over time, while the empirical literature agrees that return volatility clusters over time, to the point that time-varying return volatility is by now widely considered a stylized fact (see, e.g., [Andersen et al. \(2006\)](#)). Indeed, recent contributions to the literature on stock return predictability have found that it is important to account for both of these features; see [Dangl and Halling \(2012\)](#), [Johannes et al. \(2014\)](#) and [Pettenuzzo et al. \(2013\)](#).

In this section, we extend the model in (7) along both of these dimensions, and introduce a time-varying parameter, stochastic volatility (TVP-SV) model, where both the regression coefficients and the return volatility are allowed to change gradually over time:

$$r_{\tau+1} = (\mu + \mu_{\tau+1}) + (\beta + \beta_{\tau+1}) x_{\tau} + \exp(h_{\tau+1}) u_{\tau+1}, \quad \tau = 1, \dots, t-1, \quad (32)$$

where  $h_{\tau+1}$  denotes the (log of) stock return volatility at time  $\tau + 1$ , and  $u_{\tau+1} \sim \mathcal{N}(0, 1)$ . As for the time-varying parameters  $\theta_{\tau+1} = (\mu_{\tau+1}, \beta_{\tau+1})'$ , we assume that the intercept and slope

parameters follow a zero-mean stationary and mean-reverting process:

$$\begin{bmatrix} \mu_{\tau+1} \\ \beta_{\tau+1} \end{bmatrix} = \begin{bmatrix} \gamma_\mu & 0 \\ 0 & \gamma_\beta \end{bmatrix} \begin{bmatrix} \mu_\tau \\ \beta_\tau \end{bmatrix} + \begin{bmatrix} \eta_{1,\tau+1} \\ \eta_{2,\tau+1} \end{bmatrix}, \quad (33)$$

where  $(\mu_1, \beta_1) = (0, 0)'$ ,  $|\gamma_\mu| < 1$ ,  $|\gamma_\beta| < 1$ , and  $\eta_{\tau+1} \equiv (\eta_{1,\tau+1}, \eta_{2,\tau+1})'$  is a bi-variate normal random variable, independent of  $u_s$  for all  $t$  and  $s$ , and with  $\eta_{\tau+1} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$ .<sup>18</sup> As for the log-volatility  $h_{\tau+1}$ , it is assumed to evolve as a stationary and mean reverting process,

$$h_{\tau+1} = \lambda_0 + \lambda_1 h_\tau + \xi_{\tau+1} \quad (34)$$

where  $|\lambda_{1r}| < 1$ ,  $\xi_{\tau+1} \sim \mathcal{N}(0, \sigma_\xi^2)$  and  $u_\tau$ ,  $\eta_t$  and  $\xi_s$  are mutually independent for all  $\tau$ ,  $t$ , and  $s$ .

We now turn to describing our prior choices for the TVP-SV specification. As for  $(\mu, \beta)'$  we follow the same prior choices made for the linear model:

$$\begin{bmatrix} \mu \\ \beta \end{bmatrix} \sim \mathcal{N}(\underline{b}, \underline{V}), \quad (35)$$

Next, we note that in addition to specifying prior distributions and hyperparameters for  $[\mu, \beta]'$ , the TVP-SV model in (32)-(34) requires eliciting priors for the sequence of time-varying parameters,  $\theta^t = \{\theta_2, \dots, \theta_t\}$  and its variance covariance matrix  $\mathbf{Q}$ , the sequence of log return volatilities,  $h^t = \{h_2, \dots, h_t\}$  and its error precision  $\sigma_\xi^{-2}$ , and finally the parameters  $\gamma_\mu$ ,  $\gamma_\beta$ ,  $\lambda_0$ , and  $\lambda_1$ . Beginning with  $\theta^t$ ,  $\gamma_\mu$ ,  $\gamma_\beta$ , and  $\mathbf{Q}$ , we first write  $p(\theta^t, \gamma_\mu, \gamma_\beta, \mathbf{Q}) = p(\theta^t | \gamma_\mu, \gamma_\beta, \mathbf{Q}) p(\gamma_\mu, \gamma_\beta) p(\mathbf{Q})$ , and note that (33) along with the assumption that  $\theta_1 = (0, 0)'$  implies

$$p(\theta^t | \gamma_\mu, \gamma_\beta, \mathbf{Q}) = \prod_{\tau=1}^{t-1} p(\theta_{\tau+1} | \gamma_\mu, \gamma_\beta, \theta_\tau, \mathbf{Q}), \quad (36)$$

with  $\theta_{\tau+1} | \gamma_\mu, \gamma_\beta, \theta_\tau, \mathbf{Q} \sim \mathcal{N}(\text{diag}\{\gamma_\mu, \gamma_\beta\} \times \theta_\tau, \mathbf{Q})$ , for  $\tau = 1, \dots, t-1$ . To complete the prior elicitation for  $p(\theta^t, \gamma_\mu, \gamma_\beta, \mathbf{Q})$ , we need to specify priors for  $\mathbf{Q}$ ,  $\gamma_\mu$ , and  $\gamma_\beta$ . As for  $\mathbf{Q}$ , we choose an Inverted Wishart distribution

$$\mathbf{Q} \sim \mathcal{IW}(\underline{\mathbf{Q}}, t-2), \quad (37)$$

with

$$\underline{\mathbf{Q}} = \underline{k}_Q (t-2) \left[ s_{r,t}^2 \left( \sum_{\tau=1}^{t-1} x_\tau x_\tau' \right)^{-1} \right]. \quad (38)$$

The constant  $\underline{k}_Q$  controls the degree of variation in the time-varying regression coefficients  $\theta_\tau$ , where larger values of  $\underline{k}_Q$  imply greater variation in  $\theta_\tau$ .<sup>19</sup> We set  $\underline{k}_Q = 0.01$  to limit the extent

<sup>18</sup>Note that this is equivalent to writing  $r_{\tau+1} = \tilde{\mu}_{\tau+1} + \tilde{\beta}_{\tau+1} x_\tau + \exp(h_{\tau+1}) u_{\tau+1}$ , where  $(\tilde{\mu}_1, \tilde{\beta}_1)$  is left unrestricted.

<sup>19</sup>In this way, the scale of the Wishart distribution for  $\mathbf{Q}$  is specified to be a fraction of the OLS estimates of the variance covariance matrix  $s_{r,t}^2 (\sum_{\tau=1}^{t-1} x_\tau x_\tau')^{-1}$ , multiplied by the degrees of freedom,  $t-2$ , since for the inverted-Wishart distribution the scale matrix has the interpretation of the sum of squared residuals. This approach is consistent with the literature on TVP-VAR models; see, e.g., Cogley et al. (2005) and Primiceri (2005).

to which the parameters can change over time. As for  $\gamma_\mu, \gamma_\beta$ , we choose independent normal distributions,

$$\begin{aligned}\gamma_\mu &\sim \mathcal{N}(\underline{m}_\gamma, \underline{V}_\gamma), \quad \gamma_\mu \in (-1, 1) \\ \gamma_\beta &\sim \mathcal{N}(\underline{m}_\gamma, \underline{V}_\gamma), \quad \gamma_\beta \in (-1, 1).\end{aligned}\tag{39}$$

and we set  $\underline{m}_\gamma = 0.95$ , and  $\underline{V}_\gamma = 1.0e - 6$ , implying a high autocorrelation in both  $\mu_\tau$  and  $\beta_\tau$ .

Turning next to the sequence of log return volatilities,  $h^t$ , the error precision,  $\sigma_\xi^{-2}$ , and the parameters  $\lambda_0$  and  $\lambda_1$ , we write  $p(h^t, \lambda_0, \lambda_1, \sigma_\xi^{-2}) = p(h^t | \lambda_0, \lambda_1, \sigma_\xi^{-2}) p(\lambda_0, \lambda_1) p(\sigma_\xi^{-2})$ . As for  $p(h^t | \lambda_0, \lambda_1, \sigma_\xi^{-2})$ , it follows from (34) that

$$p(h^t | \lambda_0, \lambda_1, \sigma_\xi^{-2}) = \prod_{\tau=1}^{t-1} p(h_{\tau+1} | \lambda_0, \lambda_1, h_\tau, \sigma_\xi^{-2}) p(h_1),\tag{40}$$

with  $h_{\tau+1} | \lambda_0, \lambda_1, h_\tau, \sigma_\xi^{-2} \sim \mathcal{N}(h_\tau, \sigma_\xi^2)$ . To complete the prior elicitation for  $p(h^t, \lambda_0, \lambda_1, \sigma_\xi^{-2})$ , we need to specify priors for  $\lambda_0, \lambda_1$ , the initial log volatility  $h_1$ , and  $\sigma_\xi^{-2}$ . We choose these from the normal-gamma family as follows:

$$h_1 \sim \mathcal{N}(\ln(s_{r,t}), \underline{k}_h),\tag{41}$$

$$\begin{bmatrix} \lambda_0 \\ \lambda_1 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \underline{m}_{\lambda_0} \\ \underline{m}_{\lambda_1} \end{bmatrix}, \begin{bmatrix} \underline{V}_{\lambda_0} & 0 \\ 0 & \underline{V}_{\lambda_1} \end{bmatrix}\right), \quad \lambda_1 \in (-1, 1),\tag{42}$$

and

$$\sigma_\xi^{-2} \sim \mathcal{G}(1/\underline{k}_\xi, 1).\tag{43}$$

We set  $\underline{k}_\xi = 0.01$  and choose the remaining hyperparameters in (41) and (42) to imply uninformative priors, allowing the data to determine the degree of time variation in the return volatility. Specifically, we set  $\underline{k}_h = 0.01$ ,  $\underline{m}_{\lambda_0} = 0$ , and  $\underline{V}_{\lambda_0} = 10$ . As for the hyperparameters controlling the degree of mean reversion in  $h_\tau$ , we set  $\underline{m}_{\lambda_1} = 0.95$ , and  $\underline{V}_{\lambda_1} = 1.0e^{-06}$ , which implies a high autocorrelation in  $h_{\tau+1}$ .

To estimate the model in (32), we rely on a Gibbs sampler, which allows us to obtain a sequence of posterior draws for all the parameters of the model:  $\mu, \beta, \theta^t, h^t, \mathbf{Q}, \sigma_\xi^{-2}, \gamma_\mu, \gamma_\beta, \lambda_0$ , and  $\lambda_1$ . Finally, once such draws are available, we use them to form a density forecast for  $r_{t+1}$  in the following way:

$$\begin{aligned}p(r_{t+1} | M'_i, \mathcal{D}^t) &= \int_{\Theta, \theta^{t+1}, h^{t+1}} p(r_{t+1} | \theta_{t+1}, \Theta, h_{t+1}, \theta^t, h^t, M'_i, \mathcal{D}^t) \\ &\quad \times p(\theta_{t+1}, h_{t+1} | \Theta, \theta^t, h^t, M'_i, \mathcal{D}^t) \\ &\quad \times p(\Theta, \theta^t, h^t | M'_i, \mathcal{D}^t) d\Theta d\theta^{t+1} dh^{t+1}.\end{aligned}\tag{44}$$

where, to ease the notation, we have grouped all the time invariant parameters into the matrix  $\Theta$ , with  $\Theta = (\mu, \beta, \mathbf{Q}, \sigma_\xi^{-2}, \gamma_\mu, \gamma_\beta, \lambda_0, \lambda_1)$ . Repeating this process over the  $N$  individual models entering the model combination yields the set of  $N$  individual predictive distributions for the TVP-SV models,  $\{p(r_{t+1}|M'_i, \mathcal{D}^t)\}_{i=1}^N$ . [Appendix B](#) contains additional details on the Gibbs sampler and on how we compute the integral in (44).

Having produced the set of predictive densities  $\{p(r_{t+1}|M'_i, \mathcal{D}^t)\}_{i=1}^N$ , we next recompute all model combinations, including the DB-DeCo method described in [Section 3](#), by weighting together the TVP-SV models as specified in (32).

## 7.1 Empirical results

$R_{OoS}^2$ -values and density forecast performance for the individual TVP-SV models as well as a variety of model combinations based on such models are reported in the second halves of [Tables 2 and 3](#), under the header “TVP-SV”. Starting with the top panel of [Table 2](#), we note how allowing for instabilities in the individual models’ coefficients and volatilities leads to improvements in forecasting ability for ten of the fifteen predictors. However, for the most part, the resulting  $R_{OoS}^2$  are still largely negative, thus suggesting that in terms of point-forecast accuracy it remains very hard to beat the benchmark model. These results are consistent with the findings of [Cenesizoglu and Timmermann \(2012\)](#). Moving on to the bottom panel of [Table 2](#), we find positive  $R_{OoS}^2$  for all model combinations methods, with the exception of the Optimal prediction pool of [Geweke and Amisano \(2011\)](#). In particular, the  $R_{OoS}^2$  of the DB-DeCo method remains large and significant, though we note a marginal decrease from the results based on the linear models. We turn next to investigating the results for the density forecast comparisons in [Table 3](#). Starting with the individual models in the top panel, we find that allowing for instabilities in the individual models’ coefficients and volatilities leads in all cases to better out-of-sample density forecasts than the PM model, with all these comparisons being significant at the 1% critical level. In particular, we find average gains of approximately 8% for the log score measure and of approximately 10% for the CRPS measure. Moving on to the bottom panel of the table, we find a similar pattern for all model combinations. Once again, the DB-DeCo model stands out with the largest gains among all model combination methods considered, both in terms of CRPS and log score metrics.<sup>20</sup> The stark contrast between the results in [Table 2](#) and [Table 3](#) is suggestive of the importance of also looking at metrics summarizing the forecast performance of the density forecasts, rather than focusing only on the performance based on point forecasts. This point has been previously emphasized by [Cenesizoglu and Timmermann](#)

---

<sup>20</sup>Interestingly, we note that for the LSD metric, the individual model based on the Stock variance predictor yields a log score differential value of 11.86%, higher than the DB-DeCo. In a non-reported set of results, we find that if the learning mechanism in equations (12)-(14) is modified to produce combination weights for DB-DeCo that rely on the individual histories of log scores (i.e.  $f(r_\tau, \tilde{\mathbf{r}}_{i,\tau}) = LS_{i,\tau}$ ), the DB-DeCo LSD increases from 11.75% to 12.24% (and from 0.26% to 0.38% in the linear case).

(2012) in a similar setting.

Moving on to Table 4, the CER results appear to be largely consistent with the findings of Table 3. In particular, it appears that in all cases switching from linear to TVP-SV models produces large improvements in CERs. This is true for the individual models, whose CER values relative to the linear case increase on average by 96 basis points, and for the model combinations, whose CER values increase on average by 146 points. As for the individual models, this result is in line with the findings of Johannes et al. (2014), but generalize them to a much larger set of predictors than those considered in their study. As for the model combinations, we note that the DB-DeCo model produces the largest CER, with a value of 249 basis points. This CER value is more than twice the average CER generated by the individual TVP-SV models entering the combination. These results suggest that when controlling for model instabilities it is particularly useful to allow the combination weights to change over time and adjust to the past level of profitability of the individual models entering the combination, as our DB-DeCo method does.

Finally, Figures 5 and 6 offer a graphical illustration of the results summarized in Tables 2 and 4 for the TVP-SV based model combinations, showing over time the statistical and economic performance of the TVP-SV combination methods, relative to the PM benchmark. In particular, the cumulated CER value at the end of the sample for the DB-DeCo is approximately equal to 200%, as evidenced by the fourth panel of Figure 6. This exceeds by approximately 40% all other model combination methods.

## 8 Robustness analyses

In this section we consider several robustness checks. First, we investigate the effect on the profitability analysis presented in section 6 of altering the investor’s relative risk aversion coefficient  $A$ . Next, we conduct a subsample analysis to shed light on the robustness of our results to the choice of the forecasting evaluation period. We also consider altering the benchmark PM model by introducing time-varying volatility, to investigate the effect of pure volatility timing in our results, as well as altering the parameter  $\lambda$  controlling the degree of learning in the model combination weights. Next, we explore the sensitivity of the results to the particular choice we made with respect to the investor’s preferences, by replacing the investor’s power utility with a mean variance utility. Finally, we conduct an extensive prior sensitivity to ascertain the role of our baseline prior choices on the overall results.

### 8.1 Sensitivity to risk aversion

The economic predictability analysis we reported in sections 6 and 7 assumed a coefficient of relative risk aversion  $A = 5$ . To explore the sensitivity of our results to this value, we also



consider lower ( $A = 2$ ) and higher ( $A = 10$ ) values of this parameter. Results are shown in Table 5. First consider the case with  $A = 2$ , i.e., lower risk aversion compared to the baseline case. Under this scenario, the DB-DeCo scheme generates CERs that are above 200 basis points for both the linear and TVP-SV cases. No other model combination method comes close to these values, even though, relative to the baseline case of  $A = 5$ , we see on average an increase in all model combinations' CERs. As for the individual models, an interesting pattern emerges. Relative to the baseline case of  $A = 5$ , we find that when lowering the risk aversion to  $A = 2$ , the average CER of the linear models decreases from -0.17% ( $A = 5$ ) to -0.40% ( $A = 2$ ); in contrast, for the TVP-SV models we see that the average CER increases from 0.79% ( $A = 5$ ) to 1.13% ( $A = 2$ ). Thus, lowering the risk aversion coefficient from  $A = 5$  to  $A = 2$  has the effect of boosting the economic performance of the individual TVP-SV models, while decreasing the CER of the linear models.

We next consider the case with  $A = 10$ . In this case we find an overall decrease in CER values, both for the individual models and the model combinations. However, the DB-DeCo combination scheme continues to dominate all the other specifications. This is true for both the linear and the TVP-SV models. In particular, the CER for the DB-DeCo combination scheme averaging across the TVP-SV models is still quite large, equal to 125 basis points.

## 8.2 Subsample analysis

We next consider the robustness of our results to the choice of the forecast evaluation period. Columns two to five of Table 6 show CER results separately for recession and expansion periods, as defined by the NBER indicator. This type of analysis has been proposed by authors such as Rapach et al. (2010) and Henkel et al. (2011). When focusing on the linear models (columns two and four), we find higher economic predictability in recessions than in expansions. This results is consistent with the findings in these studies. For the TVP-SV models (column three and five), the story is however different. There we find the largest economic gains during expansions. This holds true both for the individual models and the various model combinations. This finding is somewhat surprising, since we would expect time-varying models to help when entering recessions; on the other hand, stochastic volatility might reduce the return volatility during long expansionary periods, having important consequences in the resulting asset allocations. Clark and Ravazzolo (2014) document a similar pattern in forecasting macroeconomic variables. Interestingly, the DB-DeCo scheme continue to provide positive and large economic gains in both expansions and recessions, and for both linear and TVP-SV models.

The last four columns of Table 6 show CER results separately for two out-of-sample periods, 1947-1978 and 1979-2010. Welch and Goyal (2008) argue that the predictive ability of many predictor variables deteriorates markedly after the 1973-1975 oil shock, so we are particularly interested in whether the same holds true here. The results of Table 6 are overall consistent with

this pattern, as we observe smaller gains during the second subsample, both for the individual models and the various model combinations. However, the DB-DeCo CERs are still fairly large, as high as 87 basis points in the case of linear models, and as high as 177 basis points in the TVP-SV case.

### 8.3 Sensitivity to the benchmark model

Table 7 explores the sensitivity of our main results to the choice of the benchmark model. In particular, our economic analysis in Sections 6 and 7 relied on the choice of the prevailing mean (PM) model as our benchmark. Johannes et al. (2014) point out that such choice does not allow one to isolate the effect of volatility timing from the effect of jointly forecasting expected returns and volatility. To address this point, we follow Johannes et al. (2014), and modify our benchmark model to include stochastic volatility. We label this new benchmark Prevailing Mean with Stochastic Volatility, or PM-SV in short. A quick comparison between Tables 4 and 7 reveals that switching benchmark from the PM to the PM-SV model produces a marked decrease in economic predictability, both for the individual models and the various model combinations. This comparison shows the important role of volatility timing, something that can be directly inferred by comparing the TVP-SV results across the two tables. Most notably, the DB-DeCo results remain quite strong even after replacing the benchmark model, especially for the case of TVP-SV models. In particular, when  $A = 5$  the DB-DeCo CER under the TVP-SV models is as high as 159 basis points. Altering the relative risk aversion coefficient  $A$  has some effect on the results, but the overall CERs remain quite large, ranging from 120 basis points ( $A = 2$ ) to 79 basis points ( $A = 10$ ).

### 8.4 Sensitivity to the learning dynamics

When specifying the learning mechanism for the DB-DeCo in equations (12)-(14), we introduced the smoothing parameter  $\lambda$ , where  $\lambda \in (0, 1)$ . Our main analysis of the economic value of equity premium forecasts in Sections 6 and 7 relied on  $\lambda = 0.95$ , which implies a monotonically decreasing impact of past forecast performance in the determination of the model combination weights. Several studies, such as Stock and Watson (1996) and Stock and Watson (2004) support such value. A larger or smaller discount factor is, however, possible and we investigate the sensitivity of our results to using  $\lambda = 0.9$ .<sup>21</sup> Table 8 reports the results of this sensitivity analysis where, to ease the comparison with the benchmark results based on  $\lambda = 0.95$ , we reproduce those as well. We explore the impact of altering the value of the smoothing parameter  $\lambda$  by investigating the economic impact of such choice across different risk aversion coefficients

---

<sup>21</sup>As for the case of a larger discount factor, note that when  $\lambda = 1$  equation (13) implies that the DB-DeCo scheme simplifies to the Density Combination scheme we investigated earlier, where the combination weights no longer depend on the past performance of the individual models entering the combination.

( $A = 2, 5, 10$ ) and across four different subsamples (NBER expansions and recessions, 1947-1978, and 1979-2010). Overall we find very similar results along all dimensions, with DB-DeCo models based on  $\lambda = 0.95$  generating, on average, slightly higher CERs.

## 8.5 Mean variance utility preferences

As a robustness to the particular choice of the utility function for our investor, we consider replacing the power utility function with mean variance preferences. Under mean variance preferences, at time  $\tau - 1$  the investor's utility function takes the form

$$U(W_{i,\tau}) = E[W_{i,\tau} | \mathcal{D}^{\tau-1}] - \frac{A}{2} \text{Var}[W_{i,\tau} | \mathcal{D}^{\tau-1}] \quad (45)$$

with  $W_{i,\tau}$  denoting the investor's wealth at time  $\tau$  implied by model  $M_i$ ,

$$W_{i,\tau} = (1 - \omega_{i,\tau-1}) \exp(r_{\tau-1}^f) + \omega_{i,\tau-1} \exp(r_{\tau-1}^f + r_\tau) \quad (46)$$

Next, it can be shown that the optimal allocation weights  $\omega_{i,\tau-1}^*$  are given by the solution of

$$\omega_{i,\tau-1}^* = \frac{\exp\left(\hat{\mu}_{i,\tau} + \frac{\hat{\sigma}_{i,\tau}^2}{2}\right) - 1}{A \exp(r_{\tau-1}^f) \exp\left(2\hat{\mu}_{i,\tau} + \hat{\sigma}_{i,\tau}^2\right) \left(\exp(\hat{\sigma}_{i,\tau}^2) - 1\right)}. \quad (47)$$

where  $\hat{\mu}_{i,\tau}$  and  $\hat{\sigma}_{i,\tau}^2$  are shorthands for the mean and variance of  $p(r_\tau | M_i, \mathcal{D}^{\tau-1})$ , the predictive density of  $r_\tau$  under model  $M_i$ . It is important to note that altering the utility function of the investor will have repercussions not only on the profitability of the individual models  $M_1, \dots, M_N$ , but also on the overall statistical and economic predictability of the DB-DeCo combination scheme. In fact, as we have discussed in Section 3.2, the combination weight conditional density at time  $\tau$ ,  $p(\mathbf{w}_\tau | \tilde{\mathbf{r}}_\tau, \mathcal{D}^{\tau-1})$ , depends on the history of profitability of the individual models  $M_1$  to  $M_N$  through equations (13) and (14).

Note next that in the case of a mean variance investor, time  $\tau$  CER is simply equal to the investor's realized utility  $W_{i,\tau}^*$ , hence equation (13) is replaced by

$$f(r_\tau, \tilde{\mathbf{r}}_{i,\tau}) = U(W_{i,\tau}^*), \quad (48)$$

where  $W_{i,\tau}^*$  denotes time  $\tau$  realized wealth, and is given by

$$W_{i,\tau} = (1 - \omega_{i,\tau-1}^*) \exp(r_{\tau-1}^f) + \omega_{i,\tau-1}^* \exp(r_{\tau-1}^f + r_\tau). \quad (49)$$

Having computed the optimal allocation weights for both the individual models  $M_1$  to  $M_N$  and the various model combinations, we assess the economic predictability of all such models

by computing their implied (annualized) CER, which in the case of mean variance preferences is computed simply as the average of all realized utilities over the out-of-sample period,

$$CER_m = 12 \times \frac{1}{t^*} \sum_{\tau=\underline{t}}^{\bar{t}} U(W_{m,\tau}^*) \quad (50)$$

where  $m$  denotes the model under consideration (either univariate or model combination), and  $t^* = \bar{t} - \underline{t} + 1$ . Table 9 presents differential certainty equivalent return estimates, relative to the benchmark prevailing mean model  $PM$ ,

$$\Delta CER_m = CER_m - CER_{PM} \quad (51)$$

whereby a positive entry can be interpreted as evidence that model  $m$  generates a higher (certainty equivalent) return than the benchmark model. A quick comparison between Tables 4 and 9 reveals that the economic gains for power utility and mean variance utility are quite similar in magnitude, and the overall takeaways from section 6 remain unchanged. In particular, the DB-DeCo combination scheme generates sizable CERs, especially when combining TVP-SV models. For the benchmark case of  $A = 5$ , the CER is as high as 227 basis points. Altering the risk aversion coefficients produces CERs for the DB-DeCo model ranging from 112 basis points ( $A = 10$ ) to 446 basis points ( $A = 2$ ).

## 8.6 Sensitivity to priors

As a final sensitivity, we test the robustness of our results to alternative prior assumptions and perform a sensitivity analysis in which we experiment with different values for some of the key prior hyperparameters. Given the more computational demanding algorithm required to estimate the TVP-SV models, we focus our attention on the linear models, and investigate the effectiveness of the DB-DeCo combination scheme as the key prior hyperparameters change.

First, we investigate the impact of changing the prior hyperparameter  $\underline{s}_\Lambda^{-1}$  in (12) controlling the degree of time variation in the DB-DeCo combination weights, which was set to  $\underline{s}_\Lambda^{-1} = 4$  in our baseline results. As sensitivities, we experiment with  $\underline{s}_\Lambda^{-1} = 0.2$  and  $\underline{s}_\Lambda^{-1} = 1000$ , which imply more volatile combination weights (in the case of  $\underline{s}_\Lambda^{-1} = 0.2$ ), or smoother combination weights (in the case of  $\underline{s}_\Lambda^{-1} = 1000$ ). In the former case, the annualized CER of the DB-DeCo combination scheme decreases to 0.80%, only a marginal reduction from its baseline 0.94%. Hence, it appears that having more volatile combination weights does not hinder the overall performance of DB-DeCo. On the other hand, setting  $\underline{s}_\Lambda^{-1} = 1000$  yields a much larger reduction in the DB-DeCo CER, which decreases to 0.27%. It thus appears that too large a value for  $\underline{s}_\Lambda^{-1}$  produces combination weights that are far too smooth, affecting the economic performance of DB-DeCo.<sup>22</sup>

---

<sup>22</sup>We also investigate the sensitivity of our baseline results to the choice of  $\underline{s}_{\sigma_\kappa}^{-2}$ , the prior hyperparameter

Next, we study the impact of changing the prior hyperparameters  $\underline{\psi}$  and  $\underline{v}_0$ . As discussed in Subsection 4.2, the hyperparameter  $\underline{\psi}$  plays the role of a scaling factor controlling the informativeness of the priors for  $\mu$  and  $\beta$ , and our baseline results are based on  $\underline{\psi} = 1$ . As sensitivities, we experiment with  $\underline{\psi} = 10$  and  $\underline{\psi} = 0.01$ , which imply more dispersed prior distributions (in the case of  $\underline{\psi} = 10$ ) or more concentrated prior distributions (in the case of  $\underline{\psi} = 0.01$ ) for  $\mu$  and  $\beta$ . Similarly, the prior hyperparameter  $\underline{v}_0$  controls the tightness of the prior on  $\sigma_\varepsilon^{-2}$ , and our baseline results are based on  $\underline{v}_0 = 1$ , which correspond to an hypothetical prior sample size as large as the actual sample size. As sensitivities, we experiment with  $\underline{v}_0 = 0.1$  and  $\underline{v}_0 = 100$ , which imply, respectively, an hypothetical prior sample as large as 10% of the underlying sample size (in the case of  $\underline{v}_0 = 0.1$ ) or as large as 100 times the underlying sample size (in the case of  $\underline{v}_0 = 100$ ). Table 10 summarizes the relative economic performances of both the individual linear models and the various combination schemes under these two alternative prior choices, over the whole forecast evaluation period, 1947-2010. A comparison with Table 4 reveals that relying on more dispersed prior distributions (the case of  $\underline{\psi} = 10$ ,  $\underline{v}_0 = 0.1$ ) has only minor consequences on the overall results. In particular, the economic performance of the DB-DeCo combination scheme remains unaffected by the prior change. As for the more concentrated prior distributions (the case of  $\underline{\psi} = 0.01$ ,  $\underline{v}_0 = 100$ ), we witness an overall reduction in the economic performance of both the individual models and the various combination schemes. This should be expected, as we remind that our priors are centered on the “*no predictability*” view, and as a result more concentrated priors will tend to tilt more heavily the individual models in that direction. Interestingly, the DB-DeCo combination scheme still performs quite adequately, with an annualized CER of 48 basis points.

## 9 Conclusions

In this paper we develop a novel Bayesian model combination technique that features time-varying combination weights, model incompleteness, and allows the combination weights to depend on the individual models’ past profitability in a highly flexible way. We label our new method “Decision-Based Density Combination”, in the spirit of Pesaran and Skouras (2007). Our approach is related to the recent advances on Bayesian model combination, such as the work of Geweke and Amisano (2011), Waggoner and Zha (2012), Billio et al. (2013), and Del Negro et al. (2013), but generalize them in several ways.

We apply our model combination method to the problem of stock return predictability and optimal asset allocation, and show how to estimate model combination weights that depends on the past profitability of the individual models entering the combination. When evaluated

---

controlling the degree of model incompleteness, and find that the performance of DB-DeCo deteriorates when its value is too small, with combination weights shrinking to equal weights. On the other hand, we find that when the value of  $\underline{s}_{\sigma_\varepsilon}^{-2}$  is too large the estimation algorithm seems to converge very slowly.

empirically, we find that our Decision-Based Density Combination method improves both statistical and economic measures of out-of-sample predictability, relative to the best individual models entering the combination as well as a variety of existing model combination techniques, including equal-weighted combination, BMA, and the optimal prediction pool of Geweke and Amisano (2011).

We next employ our new methodology to look at the incremental role of model instability in the context of stock return predictability. We apply our Decision-Based Density Combination method to a set of models featuring time-varying coefficients and stochastic volatility. In this way, we are able to jointly assess the importance of model uncertainty, model instabilities, and parameter uncertainty on the statistical and economic predictability of stock returns. Overall we find that explicitly accounting for model instabilities in the model combination leads to sizable improvements in predictability. In terms of economic gains, the Decision-Based Density Combination delivers an improvement in certainty equivalent returns of 145 basis points, relative to the existing model combination methods. These gains appears to be robust to a large number of robustness checks.

## References

- Aastveit, K., F. Ravazzolo, and H. van Dijk (2014). Nowcasting the business cycle in an uncertain enviroment. Norges Bank.
- Aiolfi, M. and C. A. Favero (2005). Model uncertainty, thick modelling and the predictability of stock returns. *Journal of Forecasting* 24(4), 233–254.
- Amisano, G. and R. Giacomini (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics* 25(2), 177–190.
- Andersen, T. G., T. Bollerslev, P. F. Christoffersen, and F. X. Diebold (2006). Volatility and correlation forecasting. Volume 1 of *Handbook of Economic Forecasting*, pp. 777 – 878. Elsevier.
- Andrews, D. W. K. and J. C. Monahan (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 60(4), pp. 953–966.
- Ang, A. and A. Timmermann (2012). Regime changes and financial markets. *Annual Review of Financial Economics* 4(1), 313–337.
- Avramov, D. (2002). Stock return predictability and model uncertainty. *Journal of Financial Economics* 64(3), 423 – 458.
- Baker, M. and J. Wurgler (2000). The equity share in new issues and aggregate stock returns. *The Journal of Finance* 55(5), 2219 – 2257.

- Barberis, N. (2000). Investing for the long run when returns are predictable. *The Journal of Finance* 55(1), 225 – 264.
- Bates, J. M. and C. W. J. Granger (1969). Combination of Forecasts. *Operational Research Quarterly* 20, 451–468.
- Billio, M., R. Casarin, F. Ravazzolo, and H. K. van Dijk (2013). Time-varying combinations of predictive densities using nonlinear filtering. *Journal of Econometrics* 177(2), 213–232.
- Bossaerts, P. and P. Hillion (1999). Implementing statistical criteria to select return forecasting models: what do we learn? *Review of Financial Studies* 12(2), 405–428.
- Boudoukh, J., R. Michaely, M. Richardson, and M. R. Roberts (2007). On the importance of measuring payout yield: Implications for empirical asset pricing. *The Journal of Finance* 62(2), 877–915.
- Campbell, J. and R. Shiller (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies* 1(3), 195–228.
- Campbell, J. Y. (1987). Stock returns and the term structure. *Journal of Financial Economics* 18(2), 373 – 399.
- Campbell, J. Y. and S. B. Thompson (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *Review of Financial Studies* 21(4), 1509–1531.
- Carter, C. K. and R. Kohn (1994). On gibbs sampling for state space models. *Biometrika* 81(3), pp. 541–553.
- Cenesizoglu, T. and A. Timmermann (2012). Do return prediction models add economic value? *Journal of Banking & Finance* 36(11), 2974 – 2987.
- Clark, T. E. and F. Ravazzolo (2014). Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics*, in press.
- Cogley, T., S. Morozov, and T. J. Sargent (2005). Bayesian fan charts for u.k. inflation: Forecasting and sources of uncertainty in an evolving monetary system. *Journal of Economic Dynamics and Control* 29(11), 1893 – 1925.
- Dangl, T. and M. Halling (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics* 106(1), 157 – 181.
- Del Negro, M., R. Hasegawa, and F. Schorfheide (2013). Dynamic prediction pools: an investigation of financial frictions and forecasting performance. Technical report, Unpublished manuscript.

- Diebold, F. (1991). A note on bayesian forecast combination procedures. In A. Westlund and P. Hackl (Eds.), *Economic Structural Change: Analysis and Forecasting*, pp. 225–232. New York: Springer-Verlag.
- Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics* 13(3), 253–263.
- Durbin, J. and S. J. Koopman (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika* 89(3), pp. 603–615.
- Efron, B. (2010). *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*. Cambridge University Press.
- Fama, E. F. and K. R. French (1988). Dividend yields and expected stock returns. *Journal of Financial Economics* 22(1), 3 – 25.
- Fama, E. F. and K. R. French (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics* 25(1), 23 – 49.
- Fama, E. F. and G. Schwert (1977). Asset returns and inflation. *Journal of Financial Economics* 5(2), 115 – 146.
- Fawcett, N., G. Kapetanios, J. Mitchell, and S. Price (2014). Generalised density forecast combinations. Technical Report 492, Bank of England.
- Ferson, W. E. and C. R. Harvey (1991). The variation of economic risk premiums. *Journal of Political Economy* 99(2), pp. 385–415.
- Geweke, J. (2010). *Complete and Incomplete Econometric Models*. Princeton: Princeton University Press.
- Geweke, J. and G. Amisano (2010). Comparing and evaluating bayesian predictive distributions of asset returns. *International Journal of Forecasting* 26(2), 216 – 230.
- Geweke, J. and G. Amisano (2011). Optimal prediction pools. *Journal of Econometrics* 164(1), 130 – 141.
- Geweke, J. and G. Amisano (2012). Prediction using several macroeconomic models. Technical report, European Central Bank, Frankfurt.
- Gneiting, T. and A. E. Raftery (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* 102(477), 359–378.

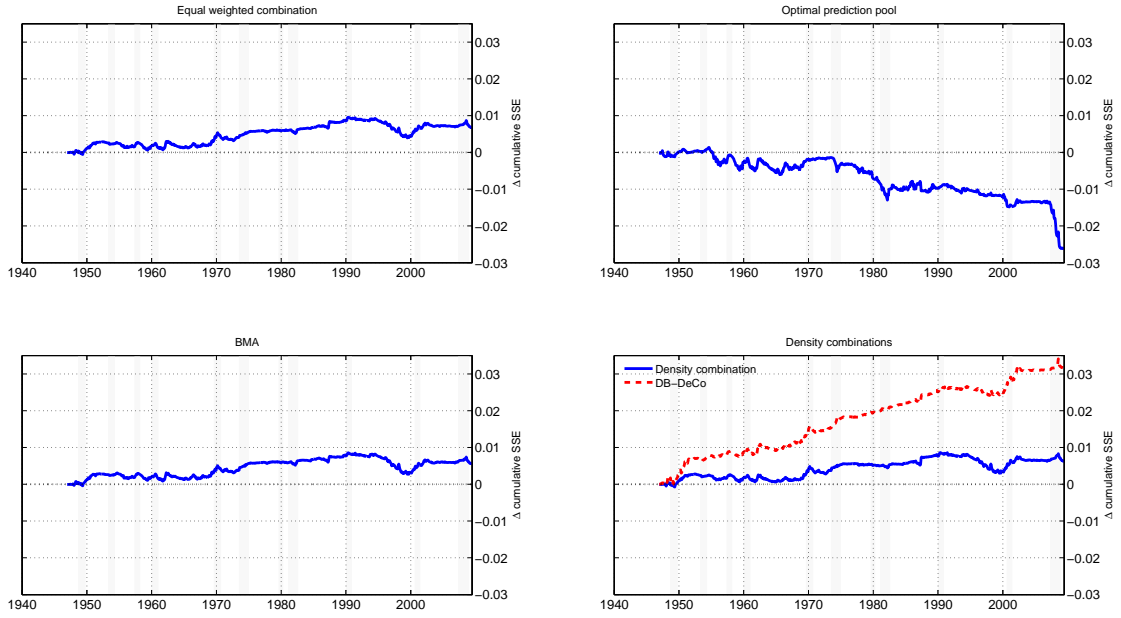


- Gneiting, T. and R. Ranjan (2011). Comparing density forecasts using threshold- and quantile-weighted scoring rules. *Journal of Business & Economic Statistics* 29(3), 411–422.
- Granger, C. W. and M. J. Machina (2006). Chapter 2 forecasting and decision theory. Volume 1 of *Handbook of Economic Forecasting*, pp. 81 – 98. Elsevier.
- Groen, J., R. Paap, and F. Ravazzolo (2013). Real-Time Inflation Forecasting in a Changing World. *Journal of Business and Economic Statistics* 31(1), 29–44.
- Hall, S. G. and J. Mitchell (2007). Combining density forecasts. *International Journal of Forecasting* 23(1), 1 – 13.
- Hendry, D. F. and M. P. Clements (2004). Pooling of forecasts. *The Econometrics Journal* 7(1), pp. 1–31.
- Henkel, S. J., J. S. Martin, and F. Nardari (2011). Time-varying short-horizon predictability. *Journal of Financial Economics* 99(3), 560 – 580.
- Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999). Bayesian model averaging: A tutorial. *Statistical Science* 14, 382–417.
- Johannes, M., A. Korteweg, and N. Polson (2014, April). Sequential learning, predictive regressions, and optimal portfolio returns. *Journal of Finance* 69(2), 611–644.
- Jore, A. S., J. Mitchell, and S. P. Vahey (2010). Combining forecast densities from VARs with uncertain instabilities. *Journal of Applied Econometrics* 25(4), 621–634.
- Kandel, S. and R. F. Stambaugh (1996). On the predictability of stock returns: An asset-allocation perspective. *The Journal of Finance* 51(2), pp. 385–424.
- Keim, D. B. and R. F. Stambaugh (1986). Predicting returns in the stock and bond markets. *Journal of Financial Economics* 17(2), 357 – 390.
- Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: Likelihood inference and comparison with arch models. *The Review of Economic Studies* 65(3), 361–393.
- Klein, R. W. and V. S. Bawa (1976). The effect of estimation risk on optimal portfolio choice. *Journal of Financial Economics* 3(3), 215 – 231.
- Koop, G. (2003). *Bayesian Econometrics*. John Wiley & Sons, Ltd.
- Lamont, O. (1998). Earnings and expected returns. *The Journal of Finance* 53(5), pp. 1563–1587.

- Leitch, G. and J. E. Tanner (1991). Economic forecast evaluation: Profits versus the conventional error measures. *The American Economic Review* 81(3), pp. 580–590.
- Lettau, M. and S. Ludvigson (2001). Consumption, aggregate wealth, and expected stock returns. *The Journal of Finance* 56(3), 815–849.
- Mitchell, J. and S. G. Hall (2005). Evaluating, comparing and combining density forecasts using the KLIC with an application to the Bank of England and NIESER “fan” charts of inflation. *Oxford Bulletin of Economics and Statistics* 67, 995–1033.
- Paye, B. S. and A. Timmermann (2006). Instability of return prediction models. *Journal of Empirical Finance* 13(3), 274 – 315.
- Pesaran, M. H. and S. Skouras (2007). *Decision-Based Methods for Forecast Evaluation*, pp. 241–267. Blackwell Publishing Ltd.
- Pettenuzzo, D. and A. Timmermann (2011). Predictability of stock returns and asset allocation under structural breaks. *Journal of Econometrics* 164(1), 60 – 78.
- Pettenuzzo, D., A. Timmermann, and R. Valkanov (2013). Forecasting stock returns under economic constraints. *Journal of Financial Economics* forthcoming.
- Poirier, D. J. (1995). *Intermediate Statistics and Econometrics: A Comparative Approach*. MIT Press.
- Polk, C., S. Thompson, and T. Vuolteenaho (2006). Cross-sectional forecasts of the equity premium. *Journal of Financial Economics* 81(1), 101 – 141.
- Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies* 72(3), 821–852.
- Rapach, D. E., J. K. Strauss, and G. Zhou (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies* 23(2), 821–862.
- Rapach, D. E. and G. Zhou (2013). Forecasting stock returns. In G. Elliot and A. Timmermann (Eds.), *Handbook of Economic Forecasting, volume 2*. North Holland.
- Stock, J. H. and M. W. Watson (1996). Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics* 14(1), 11–30.
- Stock, J. H. and M. W. Watson (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting* 23(6), 405–430.

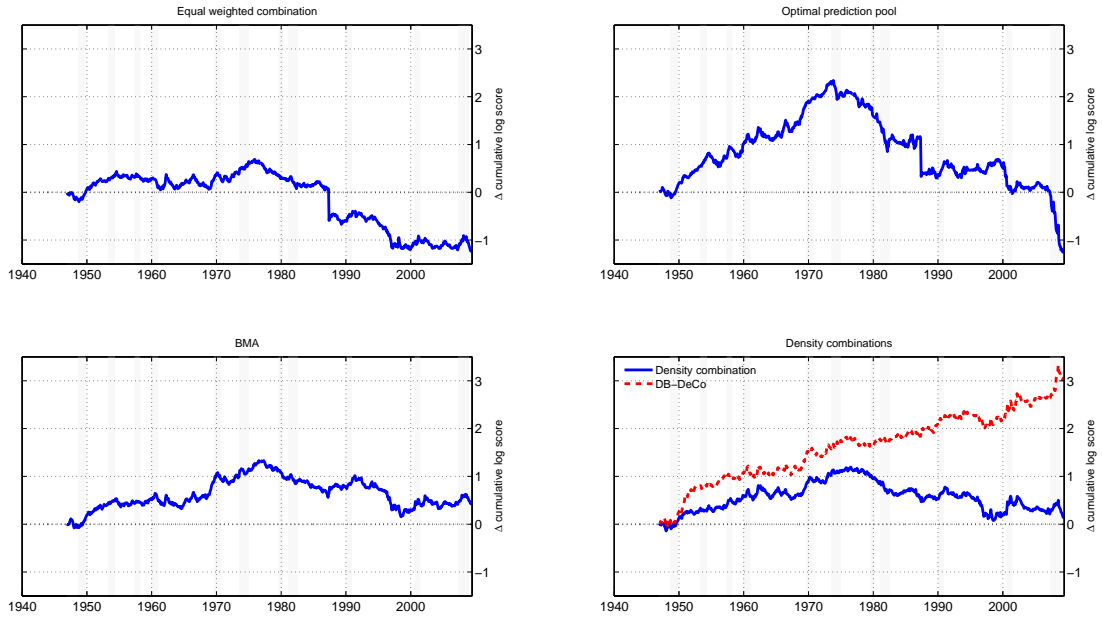
- Stock, J. H. and M. W. Watson (2006). Forecasting with many predictors. Volume 1 of *Handbook of Economic Forecasting*, pp. 515 – 554. Elsevier.
- Timmermann, A. (2006). Forecast combinations. In G. Elliot, C. Granger, and A. Timmermann (Eds.), *North-Holland*, Volume 1 of *Handbook of Economic Forecasting*, Chapter 4, pp. 135–196. Elsevier.
- Waggoner, D. F. and T. Zha (2012). Confronting model misspecification in macroeconomics. *Journal of Econometrics* 171(2), 167 – 184. Bayesian Models, Methods and Applications.
- Welch, I. and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies* 21(4), 1455–1508.
- Zhu, X. and J. Zhu (2013). Predicting stock returns: A regime-switching combination approach and economic links. *Journal of Banking & Finance* 37(11), 4120 – 4133.

Figure 1. **Cumulative sum of squared forecast error differentials**



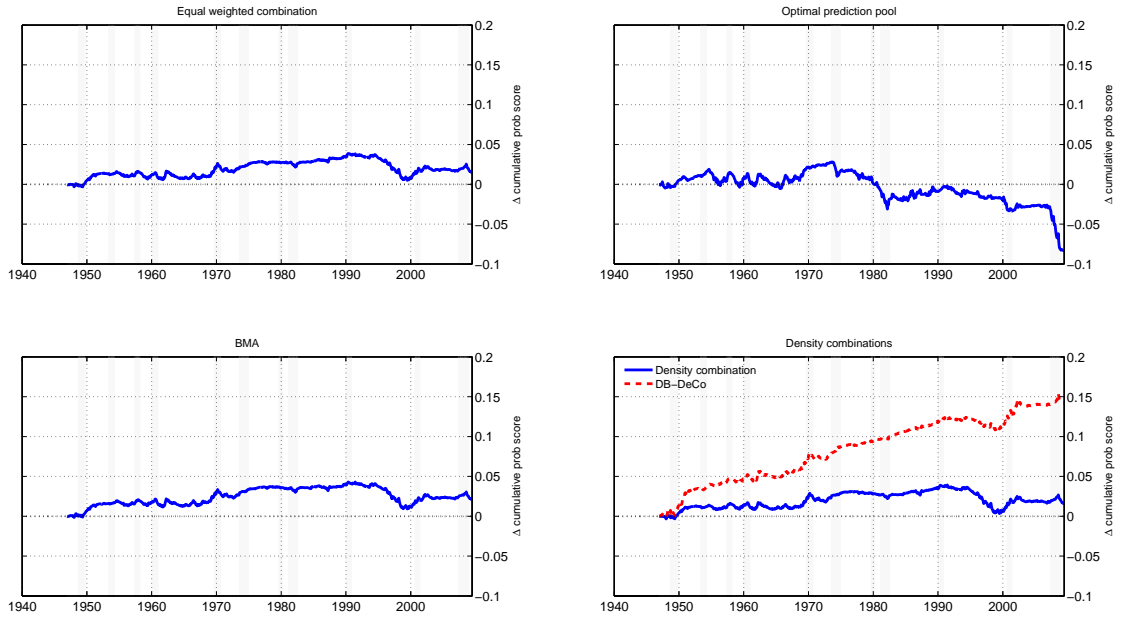
Notes: This figure shows the sum of squared forecast errors of the prevailing mean model (PM) model minus the sum of squared forecast errors of five alternative model combinations based on linear univariate models. We plot the cumulative sum of squared forecast errors of the PM forecasts relative to the model combination forecasts,  $CSSED_{m,t} = \sum_{\tau=t}^t (e_{PM,\tau}^2 - e_{m,\tau}^2)$ . Values above zero indicate that a model combination generates better performance than the PM benchmark, while negative values suggest the opposite. Each panel displays a different model combination method, with the equal weighted combination in the 1st panel, the optimal prediction pool in the 2nd panel, the BMA in the 3rd panel, and the density combination methods (with and without learning) in the 4th panel. Shaded areas indicate NBER-dated recessions.

Figure 2. Cumulative sum of log-score differentials



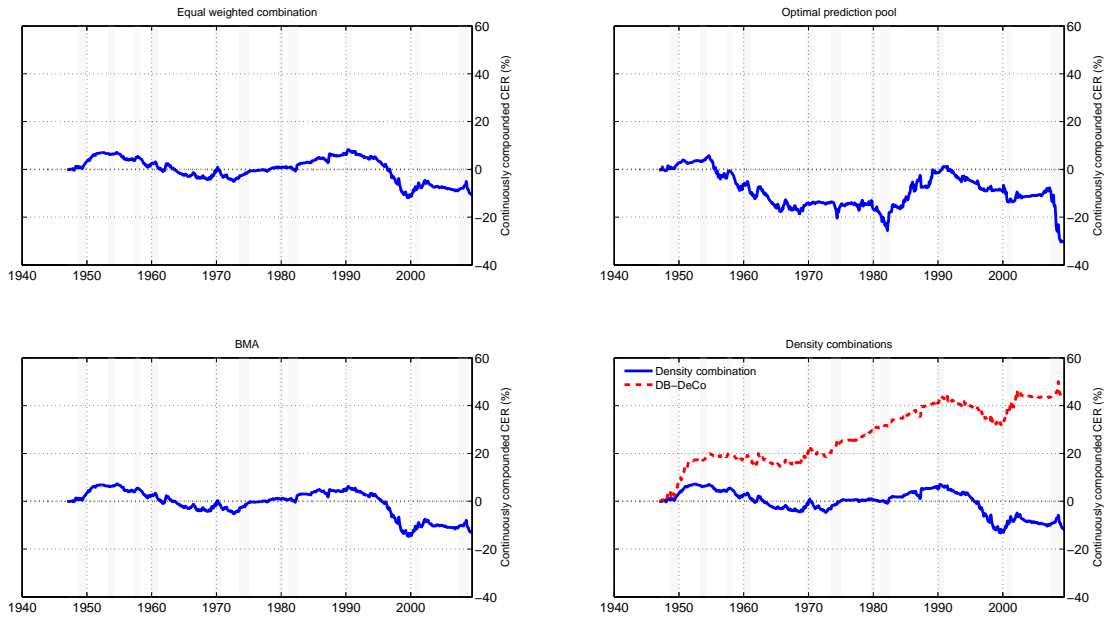
Notes: This figure shows the sum of log predictive scores of five alternative model combination methods minus the sum of log predictive scores of the PM model. We plot the cumulative sum of log-predictive scores for the model combinations based on linear models relative to the cumulative sum of log-predictive scores of the PM model,  $CLSD_{m,t} = \sum_{\tau=t}^t (LS_{m,\tau} - LS_{PM,\tau})$ . Values above zero indicate that a model combination method generates better performance than the PM benchmark, while negative values suggest the opposite. Each panel displays a different model combination method, with the equal weighted combination in the 1st panel, the optimal prediction pool in the 2nd panel, the BMA in the 3rd panel, and the density combination methods (with and without learning) in the 4th panel. Shaded areas indicate NBER-dated recessions.

Figure 3. Cumulative sum of probability score differentials



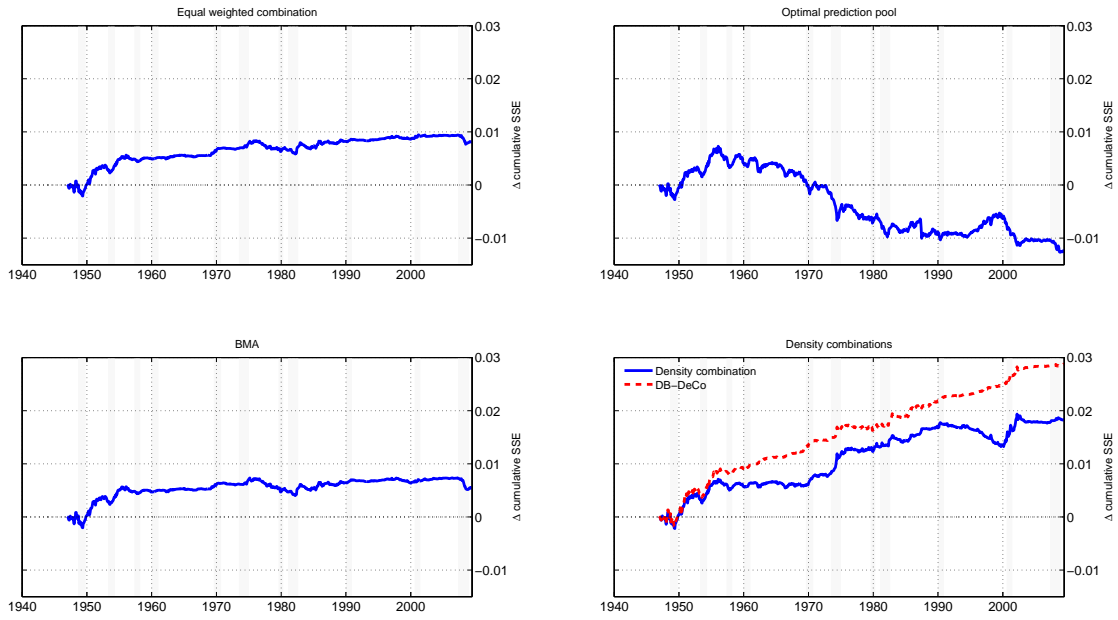
Notes: This figure shows the sum of probability scores of five alternative model combination methods minus the sum of probability scores of the PM model. We plot the cumulative sum of probability scores of the PM model relative to the cumulative sum of probability scores of the model combinations based on linear models,  $CCRPSD_{m,t} = \sum_{\tau=t}^t (CRPS_{PM,\tau} - CRPS_{m,\tau})$ . Values above zero indicate that a model combination method generates better performance than the PM benchmark, while negative values suggest the opposite. Each panel displays a different model combination method, with the equal weighted combination in the 1st panel, the optimal prediction pool in the 2nd panel, the BMA in the 3rd panel, and the density combination methods (with and without learning) in the 4th panel. Shaded areas indicate NBER-dated recessions.

Figure 4. **Economic value of out-of-sample forecasts**



Notes: This figure plots the cumulative certainty equivalent returns of five alternative model combination methods based on linear models, measured relative to the PM model. Each month we compute the optimal allocation to bonds and T-bills based on the predictive density of excess returns for both models. The investor is assumed to have power utility with a coefficient of relative risk aversion of five and the weight on stocks is constrained to lie in the interval  $[0, 0.99]$ . Each panel displays a different model combination method, with the equal weighted combination in the 1st panel, the optimal prediction pool in the 2nd panel, the BMA in the 3rd panel, and the density combination methods (with and without learning) in the 4th panel. Shaded areas indicate NBER-dated recessions.

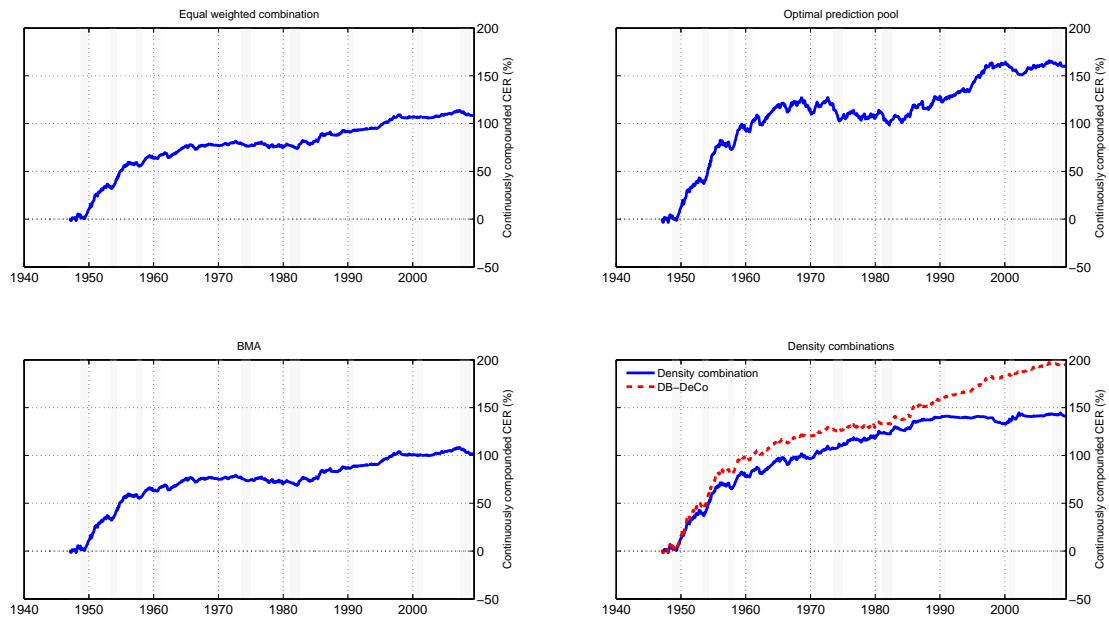
Figure 5. Cumulative sum of squared forecast error differentials, TVP-SV models



Notes: This figure shows the sum of squared forecast errors of the prevailing mean model (PM) model minus the sum of squared forecast errors of five alternative model combinations, based on TVP-SV models. We plot the cumulative sum of squared forecast errors of the PM forecasts relative to the model combination forecasts,  $CSSED_{m,t} = \sum_{\tau=t}^t (e_{PM,\tau}^2 - e_{m,\tau}^2)$ . Values above zero indicate that a model combination generates better performance than the PM benchmark, while negative values suggest the opposite. Each panel displays a different model combination method, with the equal weighted combination in the 1st panel, the optimal prediction pool in the 2nd panel, the BMA in the 3rd panel, and the density combination methods (with and without learning) in the 4th panel. Shaded areas indicate NBER-dated recessions.



Figure 6. **Economic value of out-of-sample forecasts, TVP-SV models**



Notes: This figure plots the cumulative certainty equivalent returns of five alternative model combination methods based on TVP-SV models, measured relative to the PM model. Each month we compute the optimal allocation to bonds and T-bills based on the predictive density of excess returns for both models. The investor is assumed to have power utility with a coefficient of relative risk aversion of five and the weight on stocks is constrained to lie in the interval  $[0, 0.99]$ . Each panel displays a different model combination method, with the equal weighted combination in the 1st panel, the optimal prediction pool in the 2nd panel, the BMA in the 3rd panel, and the density combination methods (with and without learning) in the 4th panel. Shaded areas indicate NBER-dated recessions.

Table 1. Summary Statistics

Variables	Mean	Std. dev.	Skewness	Kurthosis
Excess returns	0.005	0.056	-0.405	10.603
Log dividend yield	-3.324	0.450	-0.435	3.030
Log earning price ratio	-2.720	0.426	-0.708	5.659
Log smooth earning price ratio	-2.912	0.376	-0.002	3.559
Log dividend-payout ratio	-0.609	0.325	1.616	9.452
Book-to-market ratio	0.589	0.267	0.671	4.456
T-Bill rate	0.037	0.031	1.025	4.246
Long-term yield	0.053	0.028	0.991	3.407
Long-term return	0.005	0.024	0.618	8.259
Term spread	0.016	0.013	-0.218	3.128
Default yield spread	0.011	0.007	2.382	11.049
Default return spread	0.000	0.013	-0.302	11.490
Stock variance	0.003	0.005	5.875	48.302
Net equity expansion	0.019	0.024	1.468	10.638
Inflation	0.002	0.005	-0.069	6.535
Log total net payout yield	-2.137	0.224	-1.268	6.213

Notes: This table reports summary statistics for monthly excess returns, computed as returns on the S&P500 portfolio minus the T-bill rate, and for the predictor variables used in this study. The sample period is January 1927 - December 2010.

Table 2. Out-of-sample point forecast performance

	Linear	TVP-SV
Individual models		
Log dividend yield	-0.44%	<b>0.99%</b> *
Log earning price ratio	-2.27%	-0.07%
Log smooth earning price ratio	-1.51%	<b>0.68%</b>
Log dividend-payout ratio	-1.91%	-1.84%
Book-to-market ratio	-1.79%	-0.20%
T-Bill rate	-0.12%	<b>0.18%</b>
Long-term yield	-0.95%	-1.05%
Long-term return	-1.55%	-0.70%
Term spread	<b>0.09%</b>	<b>0.04%</b>
Default yield spread	-0.24%	-0.22%
Default return spread	-0.23%	-0.47%
Stock variance	<b>0.09%</b>	-0.99%
Net equity expansion	-0.93 %	-0.88%
Inflation	-0.19%	-0.20%
Log total net payout yield	-0.79%	<b>0.09%</b>
Combinations		
Equal weighted combination	<b>0.49%</b>	<b>0.62%</b> **
BMA	<b>0.39%</b>	<b>0.41%</b>
Optimal prediction pool	-1.93%	-0.86%
Density combination	<b>0.43%</b>	<b>1.33%</b> ***
Decision-based density combination	<b>2.32%</b> ***	<b>2.13%</b> ***

Notes: This table reports the out-of-sample  $R^2$  of the combination schemes and individual prediction models for monthly excess returns. The model “Decision-based density combination” refers to the case with  $A = 5$  and  $\lambda = 0.95$ . The out-of-sample  $R^2$  are measured relative to the prevailing mean model as:

$$R_{OoS,m}^2 = 1 - \frac{\sum_{\tau=\underline{t}}^{\bar{t}} e_{m,\tau}^2}{\sum_{\tau=\underline{t}}^{\bar{t}} e_{PM,\tau}^2},$$

where  $m$  denotes the model under consideration,  $\tau \in \{\underline{t}, \dots, \bar{t}\}$ , and  $e_{m,\tau}$  ( $e_{PM,\tau}$ ) denotes model  $m$ 's ( $PM$ 's) prediction error from the forecasts made at time  $\tau$ , obtained by synthesizing the predictive density into a point forecast. The column “Linear” refers to predictive return distributions based on a linear regression of monthly excess returns on an intercept and a lagged predictor variable,  $x_\tau$ :  $r_{\tau+1} = \mu + \beta x_\tau + \varepsilon_{\tau+1}$ , and combination of these  $N$  linear individual models; the column “TVP-SV” refers to predictive return distributions based on a time-varying parameter and stochastic volatility regression of monthly excess returns on an intercept and a lagged predictor variable,  $x_\tau$ :  $r_{\tau+1} = (\mu + \mu_{\tau+1}) + (\beta + \beta_{\tau+1}) x_\tau + \exp(h_{\tau+1}) u_{\tau+1}$ , and combination of these  $N$  time-varying parameter and stochastic volatility individual models. We measure statistical significance relative to the prevailing mean model using the Diebold and Mariano (1995)  $t$ -tests for equality of the average loss. The underlying  $p$ -values are based on  $t$ -statistics computed with a serial correlation-robust variance, using the pre-whitened quadratic spectral estimator of Andrews and Monahan (1992). One star \* indicates significance at 10% level; two stars \*\* significance at 5% level; and three stars \*\*\* significance at 1% level. Bold figures indicate all instances in which the out-of-sample  $R^2$  is greater than zero. All results are based on the whole forecast evaluation period, January 1947 - December 2010.

Table 3. Out-of-sample density forecast performance

	CRPSD		LSD	
	Linear	TVP-SV	Linear	TVP-SV
Individual models				
Log dividend yield	-0.37%	<b>8.21%</b> ***	-0.15%	<b>11.18%</b> ***
Log earning price ratio	-0.79%	<b>8.05%</b> ***	-0.17%	<b>11.24%</b> ***
Log smooth earning price ratio	-0.59%	<b>8.40%</b> ***	<b>0.02%</b>	<b>11.49%</b> ***
Log dividend-payout ratio	-0.45%	<b>6.64%</b> ***	-0.19%	<b>9.38%</b> ***
Book-to-market ratio	-0.61%	<b>8.25%</b> ***	-0.12%	<b>11.38%</b> ***
T-Bill rate	-0.07%	<b>7.17%</b> ***	-0.10%	<b>9.17%</b> ***
Long-term yield	-0.38%	<b>6.91%</b> ***	-0.22%	<b>9.48%</b> ***
Long-term return	-0.46%	<b>6.70%</b> ***	-0.14%	<b>9.06%</b> ***
Term spread	<b>0.08%</b>	<b>6.98%</b> ***	-0.03%	<b>8.87%</b> ***
Default yield spread	-0.07%	<b>7.17%</b> ***	-0.08%	<b>9.43%</b> ***
Default return spread	-0.11%	<b>7.03%</b> ***	-0.03%	<b>9.37%</b> ***
Stock variance	<b>0.02%</b>	<b>8.38%</b> ***	-0.02%	<b>11.86%</b> ***
Net equity expansion	0.00%	<b>7.22%</b> ***	<b>0.04%</b>	<b>9.36%</b> ***
Inflation	-0.05%	<b>7.56%</b> ***	-0.15%	<b>10.01%</b> ***
Log total net payout yield	-0.33%	<b>7.16%</b> ***	<b>0.06%</b>	<b>9.74%</b> ***
Combinations				
Equal weighted combination	<b>0.08%</b>	<b>7.88%</b> ***	-0.11%	<b>10.49%</b> ***
BMA	<b>0.10%</b>	<b>6.22%</b> ***	<b>0.03%</b>	<b>10.40%</b> ***
Optimal prediction pool	-0.43%	<b>8.36%</b> ***	-0.11%	<b>11.81%</b> ***
Density combination	<b>0.07%</b>	<b>8.53%</b> ***	<b>0.00%</b>	<b>11.17%</b> ***
Decision-based density combination	<b>0.73%</b> ***	<b>9.26%</b> ***	<b>0.26%</b> ***	<b>11.75%</b> ***

Notes: This table reports the average cumulative rank probability score differentials (columns under the heading “CRPSD”) and log predictive score differentials (columns under the heading “LSD”) of the combination schemes and individual prediction models for monthly excess returns. The model “Decision-based density combination” refers to the case with  $A = 5$  and  $\lambda = 0.95$ . The average CRPS differentials are expressed in percentage point differences relative to the prevailing mean model as:

$$CRPSD_m = \frac{\sum_{\tau=\underline{t}}^{\bar{t}} (CRPS_{PM,\tau} - CRPS_{m,\tau})}{\sum_{\tau=\underline{t}}^{\bar{t}} CRPS_{PM,\tau}},$$

where  $m$  denotes the model under consideration,  $t \in \{\underline{t}, \dots, \bar{t}\}$ , and  $CRPS_{m,\tau}$  ( $CRPS_{PM,\tau}$ ) denotes model  $m$ 's ( $PM$ 's) CRPS from the density forecasts made at time  $\tau$ . The average log predictive scores are expressed in percentage point differences relative to the prevailing mean model as:

$$LSD_m = \frac{\sum_{\tau=\underline{t}}^{\bar{t}-1} (LS_{m,\tau} - LS_{PM,\tau})}{\sum_{\tau=\underline{t}}^{\bar{t}} LS_{PM,\tau}},$$

where  $LS_{m,\tau}$  ( $LS_{PM,\tau}$ ) denotes model  $m$ 's ( $PM$ 's) log predictive score from the density forecasts made at time  $\tau$ . The columns “Linear” refer to predictive return distributions based on a linear regression of monthly excess returns on an intercept and a lagged predictor variable,  $x_\tau$ :  $r_{\tau+1} = \mu + \beta x_\tau + \varepsilon_{\tau+1}$ , and combination of these  $N$  linear individual models; the columns “TVP-SV” refer to predictive return distributions based on a time-varying parameter and stochastic volatility regression of monthly excess returns on an intercept and a lagged predictor variable,  $x_\tau$ :  $r_{\tau+1} = (\mu + \mu_{\tau+1}) + (\beta + \beta_{\tau+1}) x_\tau + \exp(h_{\tau+1}) u_{\tau+1}$ , and combination of these  $N$  time-varying parameter and stochastic volatility individual models. We measure statistical significance relative to the prevailing mean model using the Diebold and Mariano (1995)  $t$ -tests for equality of the average loss. The underlying  $p$ -values are based on  $t$ -statistics computed with a serial correlation-robust variance, using the pre-whitened quadratic spectral estimator of Andrews and Monahan (1992). One star \* indicates significance at 10% level; two stars \*\* significance at 5% level; three stars \*\*\* significance at 1% level. Bold figures indicate all instances in which  $CRPSD_m$  and  $LSD_m$  are greater than zero. All results are based on the whole forecast evaluation period, January 1947 - December 2010.

Table 4. Economic performance of portfolios based on out-of-sample return forecasts

	Linear	TVP-SV
Individual models		
Log dividend yield	-0.33%	<b>0.90%</b>
Log earning price ratio	<b>0.25%</b>	<b>0.91%</b>
Log smooth earning price ratio	-0.38%	<b>0.92%</b>
Log dividend-payout ratio	<b>0.41%</b>	<b>0.96%</b>
Book-to-market ratio	-0.58%	<b>0.71%</b>
T-Bill rate	-0.26%	<b>0.79%</b>
Long-term yield	-0.34%	<b>0.50%</b>
Long-term return	-0.42%	<b>0.77%</b>
Term spread	<b>0.15%</b>	<b>0.89%</b>
Default yield spread	-0.20%	<b>0.90%</b>
Default return spread	-0.14%	<b>0.64%</b>
Stock variance	<b>0.00%</b>	<b>0.98%</b>
Net equity expansion	-0.14%	<b>0.76%</b>
Inflation	-0.17%	<b>0.76%</b>
Log total net payout yield	-0.37%	<b>0.47%</b>
Combinations		
Equal weighted combination	<b>0.02%</b>	<b>1.17%</b>
BMA	-0.05%	<b>1.03%</b>
Optimal prediction pool	-0.82%	<b>0.96%</b>
Density combination	-0.01%	<b>1.74%</b>
Decision-based density combination	<b>0.94%</b>	<b>2.49%</b>

Notes: This table reports the certainty equivalent values for portfolio decisions based on recursive out-of-sample forecasts of monthly excess returns. Each period an investor with power utility and coefficient of relative risk aversion  $A = 5$  selects stocks and T-bills based on different predictive densities, precisely the combination schemes and individual prediction models for monthly excess returns. The model “Decision-based density combination” refers to the case with  $A = 5$  and  $\lambda = 0.95$ . The column “Linear” refers to predictive return distributions based on a linear regression of monthly excess returns on an intercept and a lagged predictor variable,  $x_\tau$ :  $r_{\tau+1} = \mu + \beta x_\tau + \varepsilon_{\tau+1}$ , and combination of these  $N$  linear individual models; the column “TVP-SV” refers to predictive return distributions based on a time-varying parameter and stochastic volatility regression of monthly excess returns on an intercept and a lagged predictor variable,  $x_\tau$ :  $r_{\tau+1} = (\mu + \mu_{\tau+1}) + (\beta + \beta_{\tau+1}) x_\tau + \exp(h_{\tau+1}) u_{\tau+1}$ , and combination of these  $N$  time-varying parameter and stochastic volatility individual models. Certainty equivalent values are annualized and are measured relative to the prevailing mean model which assumes a constant equity premium. Bold figures indicate all instances in which the CER is greater than zero. All results are based on the whole forecast evaluation period, January 1947 - December 2010.

Table 5. Effect of risk aversion on economic performance measures

	A=2		A=10	
	Linear	TVP-SV	Linear	TVP-SV
Individual models				
Log dividend yield	-0.98%	<b>1.12%</b>	-0.16%	<b>0.46%</b>
Log earning price ratio	<b>0.12%</b>	<b>1.52%</b>	<b>0.13%</b>	<b>0.51%</b>
Log smooth earning price ratio	-1.36%	<b>1.22%</b>	-0.19%	<b>0.48%</b>
Log dividend-payout ratio	<b>0.99%</b>	<b>1.14%</b>	<b>0.21%</b>	<b>0.43%</b>
Book-to-market ratio	-1.37%	<b>1.41%</b>	-0.28%	<b>0.34%</b>
T-Bill rate	-0.66%	<b>1.23%</b>	-0.13%	<b>0.41%</b>
Long-term yield	-0.86%	<b>0.83%</b>	-0.17%	<b>0.27%</b>
Long-term return	-0.81%	<b>0.94%</b>	-0.19%	<b>0.40%</b>
Term spread	<b>0.47%</b>	<b>1.64%</b>	<b>0.06%</b>	<b>0.45%</b>
Default yield spread	-0.49%	<b>1.09%</b>	-0.10%	<b>0.46%</b>
Default return spread	-0.06%	<b>1.07%</b>	-0.09%	<b>0.34%</b>
Stock variance	<b>0.02%</b>	<b>1.31%</b>	<b>0.02%</b>	<b>0.52%</b>
Net equity expansion	<b>0.54%</b>	<b>1.19%</b>	-0.08%	<b>0.39%</b>
Inflation	-0.41%	<b>0.90%</b>	-0.07%	<b>0.40%</b>
Log total net payout yield	-1.07%	<b>0.40%</b>	-0.18%	<b>0.24%</b>
Combinations				
Equal weighted combination	<b>0.06%</b>	<b>1.34%</b>	<b>0.02%</b>	<b>0.59%</b>
BMA	-0.09%	<b>1.25%</b>	-0.02%	<b>0.51%</b>
Optimal prediction pool	-1.02%	<b>1.30%</b>	-0.41%	<b>0.48%</b>
Density combination	0.00%	<b>1.73%</b>	<b>0.01%</b>	<b>0.88%</b>
Decision-based density combination	<b>2.63%</b>	<b>2.35%</b>	<b>0.50%</b>	<b>1.25%</b>

Notes: This table reports the certainty equivalent values for portfolio decisions based on recursive out-of-sample forecasts of monthly excess returns. Each period an investor with power utility and coefficient of relative risk aversion of two (columns two and three) or ten (columns four and five) selects stocks and T-bills based on different predictive densities, precisely the combination schemes and individual prediction models for monthly excess returns. The model “Decision-based density combination” refers to the case with  $A$  matching the values in the headings ( $A = 2, 10$ ) and  $\lambda = 0.95$ . The columns “Linear” refer to predictive return distributions based on a linear regression of monthly excess returns on an intercept and a lagged predictor variable,  $x_\tau$ :  $r_{\tau+1} = \mu + \beta x_\tau + \varepsilon_{\tau+1}$ , and combination of these  $N$  linear individual models; the columns “TVP-SV” refer to predictive return distributions based on a time-varying parameter and stochastic volatility regression of monthly excess returns on an intercept and a lagged predictor variable,  $x_\tau$ :  $r_{\tau+1} = (\mu + \mu_{\tau+1}) + (\beta + \beta_{\tau+1}) x_\tau + \exp(h_{\tau+1}) u_{\tau+1}$ , and combination of these  $N$  time-varying parameter and stochastic volatility individual models. Certainty equivalent values are annualized and are measured relative to the prevailing mean model which assumes a constant equity premium. Bold figures indicate all instances in which the CER is greater than zero. All results are based on the whole forecast evaluation period, January 1947 - December 2010.

Table 6. Economic performance measures: subsamples

	NBER expansions		NBER recessions		1947-1978		1979-2010	
	Linear	TVP-SV	Linear	TVP-SV	Linear	TVP-SV	Linear	TVP-SV
Individual models								
Log dividend yield	-1.11%	<b>0.73%</b>	<b>3.31%</b>	<b>1.66%</b>	<b>0.14%</b>	<b>1.77%</b>	-0.81%	<b>0.02%</b>
Log earning price ratio	<b>0.14%</b>	<b>1.36%</b>	<b>0.73%</b>	-1.08%	<b>0.58%</b>	<b>1.17%</b>	-0.10%	<b>0.65%</b>
Log smooth earning price ratio	-1.07%	<b>1.20%</b>	<b>2.80%</b>	-0.32%	-0.40%	<b>1.56%</b>	-0.37%	<b>0.27%</b>
Log dividend-payout ratio	<b>0.96%</b>	<b>1.89%</b>	-2.05%	-3.14%	<b>0.46%</b>	<b>1.35%</b>	<b>0.35%</b>	<b>0.56%</b>
Book-to-market ratio	-0.94%	<b>1.37%</b>	<b>1.11%</b>	-2.22%	-0.18%	<b>1.25%</b>	-0.99%	<b>0.16%</b>
T-Bill rate	-0.52%	<b>0.82%</b>	<b>0.96%</b>	<b>0.67%</b>	<b>0.09%</b>	<b>1.79%</b>	-0.61%	-0.21%
Long-term yield	-0.64%	<b>0.20%</b>	<b>1.02%</b>	<b>1.91%</b>	-0.06%	<b>1.58%</b>	-0.63%	-0.59%
Long-term return	-0.32%	<b>1.07%</b>	-0.85%	-0.60%	-0.88%	<b>0.89%</b>	<b>0.06%</b>	<b>0.63%</b>
Term spread	-0.01%	<b>1.11%</b>	<b>0.87%</b>	-0.14%	<b>0.14%</b>	<b>1.53%</b>	<b>0.16%</b>	<b>0.23%</b>
Default yield spread	-0.27%	<b>1.44%</b>	<b>0.10%</b>	-1.51%	-0.24%	<b>1.47%</b>	-0.16%	<b>0.31%</b>
Default return spread	-0.12%	<b>1.19%</b>	-0.28%	-1.82%	-0.34%	<b>0.53%</b>	<b>0.06%</b>	<b>0.75%</b>
Stock variance	-0.13%	<b>1.94%</b>	<b>0.62%</b>	-3.24%	-0.14%	<b>1.49%</b>	<b>0.14%</b>	<b>0.46%</b>
Net equity expansion	0.76%	<b>1.94%</b>	-4.06%	-4.39%	<b>0.06%</b>	<b>1.43%</b>	-0.34%	<b>0.08%</b>
Inflation	-0.19%	<b>1.26%</b>	-0.11%	-1.45%	-0.16%	<b>1.24%</b>	-0.19%	<b>0.28%</b>
Log total net payout yield	-0.73%	<b>0.67%</b>	<b>1.28%</b>	-0.44%	<b>0.24%</b>	<b>1.59%</b>	-0.98%	-0.67%
Combinations								
Equal weighted combination	-0.18%	<b>1.45%</b>	<b>0.89%</b>	-0.12%	<b>0.13%</b>	<b>1.70%</b>	-0.10%	<b>0.62%</b>
BMA	-0.27%	<b>1.33%</b>	<b>0.96%</b>	-0.32%	<b>0.13%</b>	<b>1.56%</b>	-0.24%	<b>0.49%</b>
Optimal prediction pool	-0.39%	<b>1.86%</b>	-2.75%	-3.02%	-0.48%	<b>1.42%</b>	-1.17%	<b>0.49%</b>
Density combination	-0.25%	<b>1.66%</b>	<b>1.05%</b>	<b>2.10%</b>	<b>0.10%</b>	<b>2.73%</b>	-0.14%	<b>0.74%</b>
Decision-based density combination	<b>0.67%</b>	<b>2.67%</b>	<b>2.15%</b>	<b>1.66%</b>	<b>1.00%</b>	<b>3.20%</b>	<b>0.87%</b>	<b>1.77%</b>

Notes: This table reports the certainty equivalent values for portfolio decisions based on recursive out-of-sample forecasts of monthly excess returns, over four alternative subsamples (NBER-dated expansions, NBER-dated recessions, 1947-1978, and 1979-2010). Each period an investor with power utility and coefficient of relative risk aversion  $A = 5$  selects stocks and T-bills based on different predictive densities, precisely the combination schemes and individual prediction models for monthly excess returns. The model “Decision-based density combination” refers to the case with  $A = 5$  and  $\lambda = 0.95$ . The columns “Linear” refer to predictive return distributions based on a linear regression of monthly excess returns on an intercept and a lagged predictor variable,  $x_{\tau}$ :  $r_{\tau+1} = \mu + \beta x_{\tau} + \varepsilon_{\tau+1}$ , and combination of these  $N$  linear individual models; the columns “TVP-SV” refer to predictive return distributions based on a time-varying parameter and stochastic volatility regression of monthly excess returns on an intercept and a lagged predictor variable,  $x_{\tau}$ :  $r_{\tau+1} = (\mu + \mu_{\tau+1}) + (\beta + \beta_{\tau+1}) x_{\tau} + \exp(h_{\tau+1}) u_{\tau+1}$ , and combination of these  $N$  time-varying parameter and stochastic volatility individual models. Certainty equivalent values are annualized and are measured relative to the prevailing mean model which assumes a constant equity premium. Bold figures indicate all instances in which the CER is greater than zero.

Table 7. Economic performance measures: alternative benchmark

	A=2		A=5		A=10	
	Linear	TVP-SV	Linear	TVP-SV	Linear	TVP-SV
Individual models						
Log dividend yield	-2.14%	-0.04%	-1.23%	<b>0.00%</b>	-0.62%	-0.00%
Log earning price ratio	-1.04%	<b>0.36%</b>	-0.65%	<b>0.02%</b>	-0.33%	<b>0.05%</b>
Log smooth earning price ratio	-2.52%	<b>0.06%</b>	-1.28%	<b>0.02%</b>	-0.65%	<b>0.02%</b>
Log dividend-payout ratio	-0.16%	-0.02%	-0.49%	<b>0.06%</b>	-0.25%	-0.03%
Book-to-market ratio	-2.53%	<b>0.25%</b>	-1.47%	-0.19%	-0.74%	-0.12%
T-Bill rate	-1.82%	<b>0.08%</b>	-1.15%	-0.10%	-0.59%	-0.05%
Long-term yield	-2.02%	-0.33%	-1.24%	-0.39%	-0.63%	-0.19%
Long-term return	-1.97%	-0.21%	-1.31%	-0.13%	-0.65%	-0.06%
Term spread	-0.68%	<b>0.48%</b>	-0.75%	-0.01%	-0.40%	-0.01%
Default yield spread	-1.64%	-0.07%	-1.10%	-0.00%	-0.56%	-0.00%
Default return spread	-1.22%	-0.09%	-1.04%	-0.26%	-0.55%	-0.12%
Stock variance	-1.14%	<b>0.16%</b>	-0.90%	<b>0.09%</b>	-0.44%	<b>0.06%</b>
Net equity expansion	-0.62%	<b>0.04%</b>	-1.03%	-0.14%	-0.54%	-0.07%
Inflation	-1.56%	-0.25%	-1.07%	-0.13%	-0.53%	-0.06%
Log total net payout yield	-2.23%	-0.76%	-1.26%	-0.43%	-0.64%	-0.22%
Combinations						
Equal weighted combination	-1.10%	<b>0.18%</b>	-0.88%	<b>0.27%</b>	-0.44%	<b>0.13%</b>
BMA	-1.13%	<b>0.05%</b>	-0.59%	<b>0.13%</b>	-0.28%	<b>0.05%</b>
Optimal prediction pool	-2.18%	<b>0.15%</b>	-1.72%	<b>0.06%</b>	-0.87%	<b>0.02%</b>
Density combination	-1.16%	<b>0.57%</b>	-0.91%	<b>0.84%</b>	-0.45%	<b>0.42%</b>
Decision-based density combination	<b>1.47%</b>	<b>1.20%</b>	<b>0.04%</b>	<b>1.59%</b>	<b>0.04%</b>	<b>0.79%</b>

Notes: This table reports the certainty equivalent values for portfolio decisions based on recursive out-of-sample forecasts of monthly excess returns, relative to the alternative benchmark model with constant mean and stochastic volatility. Each period an investor with power utility and coefficient of relative risk aversion  $A$  selects stocks and T-bills based on different predictive densities, precisely the combination schemes and individual prediction models for monthly excess returns. The model “Decision-based density combination” refers to the case with  $A$  matching the values in the column headings ( $A = 2, 5, 10$ ) and  $\lambda = 0.95$ . The columns “Linear” refer to predictive return distributions based on a linear regression of monthly excess returns on an intercept and a lagged predictor variable,  $x_\tau: r_{\tau+1} = \mu + \beta x_\tau + \varepsilon_{\tau+1}$ , and combination of these  $N$  linear individual models; the columns “TVP-SV” refer to predictive return distributions based on a time-varying parameter and stochastic volatility regression of monthly excess returns on an intercept and a lagged predictor variable,  $x_\tau: r_{\tau+1} = (\mu + \mu_{\tau+1}) + (\beta + \beta_{\tau+1}) x_\tau + \exp(h_{\tau+1}) u_{\tau+1}$ , and combination of these  $N$  time-varying parameter and stochastic volatility individual models. Certainty equivalent values are annualized and are measured relative to the alternative benchmark model which assumes a constant equity premium and stochastic volatility. Bold figures indicate all instances in which the CER is greater than zero. All results are based on the whole forecast evaluation period, January 1947 - December 2010.



Table 8. Economic performance measures: alternative learning dynamics

$\lambda$	Full sample		NBER expansions		NBER recessions		1947-1978		1979-2010	
	Linear	TVP-SV	Linear	TVP-SV	Linear	TVP-SV	Linear	TVP-SV	Linear	TVP-SV
A=2										
lambda=0.9	<b>2.31%</b>	<b>2.20%</b>	<b>0.36%</b>	<b>1.27%</b>	<b>5.09%</b>	<b>2.48%</b>	<b>2.31%</b>	<b>2.38%</b>	<b>2.32%</b>	<b>2.02%</b>
lambda=0.95	<b>2.63%</b>	<b>2.35%</b>	<b>0.37%</b>	<b>1.32%</b>	<b>5.86%</b>	<b>2.89%</b>	<b>2.66%</b>	<b>2.50%</b>	<b>2.60%</b>	<b>2.20%</b>
A=5										
lambda=0.9	<b>0.93%</b>	<b>2.49%</b>	<b>0.67%</b>	<b>2.66%</b>	<b>2.09%</b>	<b>1.70%</b>	<b>1.02%</b>	<b>3.22%</b>	<b>0.83%</b>	<b>1.74%</b>
lambda=0.95	<b>0.94%</b>	<b>2.49%</b>	<b>0.67%</b>	<b>2.67%</b>	<b>2.15%</b>	<b>1.66%</b>	<b>1.00%</b>	<b>3.20%</b>	<b>0.87%</b>	<b>1.77%</b>
A=10										
lambda=0.9	<b>0.50%</b>	<b>1.21%</b>	<b>1.67%</b>	<b>2.13%</b>	<b>1.14%</b>	<b>0.93%</b>	<b>0.58%</b>	<b>1.56%</b>	<b>0.42%</b>	<b>0.85%</b>
lambda=0.95	<b>0.50%</b>	<b>1.25%</b>	<b>1.89%</b>	<b>2.23%</b>	<b>1.13%</b>	<b>0.90%</b>	<b>0.57%</b>	<b>1.62%</b>	<b>0.43%</b>	<b>0.86%</b>

Notes: This table reports the certainty equivalent values for portfolio decisions based on recursive out-of-sample forecasts of monthly excess returns, under two alternative parametrizations for the learning dynamics,  $\lambda=0.95$  (our benchmark case) and  $\lambda = 0.9$ . Each period an investor with power utility and coefficient of relative risk aversion  $A = 5$  selects stocks and T-bills based on two utility-based density combinations with  $A$  matching the values in three panels ( $A = 2, 5, 10$ ). The columns “Linear” refer to predictive return distributions based on a linear regression of monthly excess returns on an intercept and a lagged predictor variable,  $x_\tau: r_{\tau+1} = \mu + \beta x_\tau + \varepsilon_{\tau+1}$ , and combination of these  $N$  linear individual models; the columns “TVP-SV” refer to predictive return distributions based on a time-varying parameter and stochastic volatility regression of monthly excess returns on an intercept and a lagged predictor variable,  $x_\tau: r_{\tau+1} = (\mu + \mu_{\tau+1}) + (\beta + \beta_{\tau+1}) x_\tau + \exp(h_{\tau+1}) u_{\tau+1}$ , and combination of these  $N$  time-varying parameter and stochastic volatility individual models. Certainty equivalent values are annualized and are measured relative to the prevailing mean model which assumes a constant equity premium. Bold figures indicate all instances in which the CER is greater than zero. The results are based on five different samples: full sample (1947-2010), NBER expansions, NBER recessions, 1947-1978, and 1979-2010.

Table 9. Economic performance measures: Mean Variance preferences

	A=2			A=5			A=10		
	Linear	TVP-SV	Linear	TVP-SV	Linear	TVP-SV	Linear	TVP-SV	Linear
Individual models									
Log dividend yield	-0.80%	<b>1.63%</b>	-0.33%	<b>0.82%</b>	-0.16%	<b>0.41%</b>	-0.16%	<b>0.41%</b>	-0.16%
Log earning price ratio	<b>0.31%</b>	<b>2.03%</b>	<b>0.20%</b>	<b>0.92%</b>	<b>0.10%</b>	<b>0.47%</b>	<b>0.10%</b>	<b>0.47%</b>	<b>0.10%</b>
Log smooth earning price ratio	-0.99%	<b>1.96%</b>	-0.39%	<b>0.83%</b>	-0.20%	<b>0.42%</b>	-0.20%	<b>0.42%</b>	-0.20%
Log dividend-payout ratio	<b>1.06%</b>	<b>1.96%</b>	<b>0.39%</b>	<b>0.83%</b>	<b>0.20%</b>	<b>0.44%</b>	<b>0.20%</b>	<b>0.44%</b>	<b>0.20%</b>
Book-to-market ratio	-1.48%	<b>1.99%</b>	-0.57%	<b>0.65%</b>	-0.29%	<b>0.34%</b>	-0.29%	<b>0.34%</b>	-0.29%
T-Bill rate	-0.59%	<b>1.87%</b>	-0.24%	<b>0.76%</b>	-0.13%	<b>0.38%</b>	-0.13%	<b>0.38%</b>	-0.13%
Long-term yield	-0.81%	<b>1.65%</b>	-0.33%	<b>0.54%</b>	-0.16%	<b>0.27%</b>	-0.16%	<b>0.27%</b>	-0.16%
Long-term return	-0.96%	<b>1.55%</b>	-0.38%	<b>0.75%</b>	-0.20%	<b>0.37%</b>	-0.20%	<b>0.37%</b>	-0.20%
Term spread	<b>0.37%</b>	<b>2.73%</b>	<b>0.13%</b>	<b>0.91%</b>	<b>0.06%</b>	<b>0.47%</b>	<b>0.06%</b>	<b>0.47%</b>	<b>0.06%</b>
Default yield spread	-0.46%	<b>1.90%</b>	-0.19%	<b>0.88%</b>	-0.09%	<b>0.45%</b>	-0.09%	<b>0.45%</b>	-0.09%
Default return spread	-0.22%	<b>1.64%</b>	-0.15%	<b>0.63%</b>	-0.08%	<b>0.31%</b>	-0.08%	<b>0.31%</b>	-0.08%
Stock variance	<b>0.01%</b>	<b>2.02%</b>	<b>0.01%</b>	<b>0.98%</b>	-0.01%	<b>0.50%</b>	-0.01%	<b>0.50%</b>	-0.01%
Net equity expansion	<b>0.46%</b>	<b>1.99%</b>	-0.07%	<b>0.80%</b>	-0.04%	<b>0.40%</b>	-0.04%	<b>0.40%</b>	-0.04%
Inflation	-0.37%	<b>1.72%</b>	-0.14%	<b>0.74%</b>	-0.09%	<b>0.37%</b>	-0.09%	<b>0.37%</b>	-0.09%
Log total net payout yield	-0.91%	<b>0.85%</b>	-0.36%	<b>0.41%</b>	-0.18%	<b>0.21%</b>	-0.18%	<b>0.21%</b>	-0.18%
Combinations									
Equal weighted combination	<b>0.03%</b>	<b>2.52%</b>	<b>0.01%</b>	<b>1.09%</b>	<b>0.00%</b>	<b>0.55%</b>	<b>0.00%</b>	<b>0.55%</b>	<b>0.00%</b>
BMA	-0.11%	<b>2.26%</b>	-0.05%	<b>0.95%</b>	-0.03%	<b>0.49%</b>	-0.03%	<b>0.49%</b>	-0.03%
Optimal prediction pool	-1.66%	<b>2.05%</b>	-0.74%	<b>0.95%</b>	-0.38%	<b>0.49%</b>	-0.38%	<b>0.49%</b>	-0.38%
Density combination	-0.04%	<b>3.39%</b>	-0.02%	<b>1.62%</b>	-0.02%	<b>0.80%</b>	-0.02%	<b>0.80%</b>	-0.02%
Decision-based density combination	<b>2.55%</b>	<b>4.46%</b>	<b>0.85%</b>	<b>2.27%</b>	<b>0.45%</b>	<b>1.12%</b>	<b>0.45%</b>	<b>1.12%</b>	<b>0.45%</b>

Notes: This table reports the certainty equivalent values for portfolio decisions based on recursive out-of-sample forecasts of monthly excess returns. Each period an investor with mean variance utility and coefficient of relative risk aversion  $A$  selects stocks and T-bills based on different predictive densities, precisely the combination schemes and individual prediction models for monthly excess returns. The model “Decision-based density combination” refers to the case with  $A$  matching the values in the column headings ( $A = 2, 5, 10$ ) and  $\lambda = 0.95$ . The columns “Linear” refer to predictive return distributions based on a linear regression of monthly excess returns on an intercept and a lagged predictor variable,  $x_\tau: r_{\tau+1} = \mu + \beta x_\tau + \varepsilon_{\tau+1}$ , and combination of these  $N$  linear individual models; the columns “TVP-SV” refer to predictive return distributions based on a time-varying parameter and stochastic volatility regression of monthly excess returns on an intercept and a lagged predictor variable,  $x_\tau: r_{\tau+1} = (\mu + \mu_{\tau+1}) + (\beta + \beta_{\tau+1}) x_\tau + \exp(h_{\tau+1}) u_{\tau+1}$ , and combination of these  $N$  time-varying parameter and stochastic volatility individual models. Certainty equivalent values are annualized and are measured relative to the alternative benchmark model which assumes a constant equity premium and stochastic volatility. Bold figures indicate all instances in which the CER is greater than zero. All results are based on the whole forecast evaluation period, January 1947 - December 2010.

Table 10. Prior sensitivity analysis: economic performance

	$\psi = 10, v_0 = 0.1$	$\psi = 0.01, v_0 = 100$
Individual models		
Log dividend yield	-0.25%	-0.19%
Log earning price ratio	<b>0.27%</b>	<b>0.08%</b>
Log smooth earning price ratio	-0.29%	-0.16%
Log dividend-payout ratio	<b>0.30%</b>	<b>0.06%</b>
Book-to-market ratio	-0.70%	-0.26%
T-Bill rate	-0.16%	-0.19%
Long-term yield	-0.24%	-0.15%
Long-term return	-0.06%	-0.31%
Term spread	<b>0.33%</b>	-0.23%
Default yield spread	<b>0.00%</b>	-0.11%
Default return spread	<b>0.01%</b>	-0.03%
Stock variance	<b>0.27%</b>	<b>0.00%</b>
Net equity expansion	-0.02%	-0.03%
Inflation	-0.01%	-0.12%
Log total net payout yield	-0.23%	-0.23%
Combinations		
Equal weighted combination	<b>0.16%</b>	-0.08%
BMA	<b>0.17%</b>	-0.06%
Optimal prediction pool	-0.56%	-0.07%
Density combination	-0.06%	<b>0.00%</b>
Decision-based density combination	<b>0.94%</b>	<b>0.48%</b>

Notes: This table reports the certainty equivalent values for portfolio decisions based on recursive out-of-sample forecasts of monthly excess returns. Each period an investor with power utility and coefficient of relative risk aversion  $A = 5$  selects stocks and T-bills based on different predictive densities, precisely the combination schemes and individual prediction models for monthly excess returns. The model “Decision-based density combination” refers to the case with  $A = 5$  and  $\lambda = 0.95$ . Predictive return distributions are based on a linear regression of monthly excess returns on an intercept and a lagged predictor variable,  $x_\tau$ :  $r_{\tau+1} = \mu + \beta x_\tau + \varepsilon_{\tau+1}$ , and combination of these  $N$  linear individual models for two different set of priors. The prior set with  $\psi = 10$  and  $v_0 = 0.1$  refers to a diffuse prior assumption and  $\psi = 10$  and  $v_0 = 0.1$  to an informative prior assumption. Certainty equivalent values are annualized and are measured relative to the prevailing mean model which assumes a constant equity premium. Bold figures indicate all instances in which the CER is greater than zero. All results are based on the whole forecast evaluation period, January 1947 - December 2010.

## Appendices (not for publication)

### Appendix A Convolution

As noted in Section 3.3, the pdf  $p(r_{t+1}|\tilde{\mathbf{r}}_{t+1}, \mathbf{w}_{t+1}, \sigma_\kappa^{-2})$  is obtained using a convolution mechanism that combines the  $N$  original densities  $\{p(r_{t+1}|M_i, \mathcal{D}^t)\}_{i=1}^N$  using as weights  $\mathbf{w}_{t+1}$ . That is,

$$p(r_{t+1}|\tilde{\mathbf{r}}_{t+1}, \mathbf{w}_{t+1}, \sigma_\kappa^{-2}) \propto \exp \left\{ -\frac{1}{2} (r_{t+1} - \tilde{\mathbf{r}}_{t+1}' \mathbf{w}_{t+1})' \sigma_\kappa^{-2} (r_{t+1} - \tilde{\mathbf{r}}_{t+1}' \mathbf{w}_{t+1}) \right\} \quad (\text{A-1})$$

The convolution applies in equation (20) and it precisely creates the following combined density:

$$(\tilde{\mathbf{r}}_{t+1} * \mathbf{w}_{t+1}) = \int_{-\inf}^{+\inf} (\tilde{r}_{t+1,1}(\xi) * \dots * \tilde{r}_{t+1,N}(\xi)) * (w_{t+1,1}(d - \xi) * \dots * w_{t+1,N}(d - \xi)) d\xi \quad (\text{A-2})$$

where  $\tilde{\mathbf{r}}_{t+1} = (\tilde{r}_{t+1,1} * \dots * \tilde{r}_{t+1,N})$ ,  $\mathbf{w}_{t+1} = (w_{t+1,1} * \dots * w_{t+1,N})'$ . For each value of  $\xi$ , the convolution formula can be described as a weighted average of the function  $\tilde{r}_{t+1}(\xi)$  with weight  $\mathbf{w}_{t+1}(d - \xi)$ . As  $d$  changes, the weighting function emphasizes different parts of the input function. We follow Billio et al. (2013) and use different draws from the  $N$  individual predictive densities as values for  $\xi$ .

Convolution has several important mathematical properties that we exploit to derive equation (A-2):

- Property 1:  $(\tilde{y} * w)_{t_q+h} = (w * \tilde{y})_{t_q+1}$ .
- Property 2:  $\tilde{y} * (w * \gamma)_{t_q+1} = ((w * \tilde{y}) * \gamma)_{t_q+1}$ .
- Property 3:  $\tilde{y} * (w + \gamma)_{t_q+1} = (\tilde{y} * w)_{t_q+1} + (\tilde{y} * \gamma)_{t_q+1}$ .
- Property 4:  $\alpha(\tilde{y} * w)_{t_q+1} = \alpha(\tilde{y})_{t_q+1} * w_{t_q+1}$ , for any real or complex  $\alpha$ .

Property 1 implies that the operation is invariant to the order of the operands (commutative property). Property 2 implies that the order in which the operations are performed does not matter as long as the sequence of the operands is not changed (associative property). Property 3 implies that multiplying a density by a group of added densities yields the same outcome as multiplying each density separately and then adding them together (distributive property). Property 4 implies that associativity holds for any scalar multiplication.

## Appendix B Posterior simulations

### Appendix B.1 Linear models

For the linear models the goal is to obtain draws from the joint posterior distribution  $p(\mu, \beta, \sigma_\varepsilon^{-2} | M_i, \mathcal{D}^t)$ , where  $\mathcal{D}^t$  denotes all information available up to time  $t$ , and  $M_i$  denotes model  $i$ , with  $i = 1, \dots, N$ . Combining the priors in (21)-(23) with the likelihood function yields the following conditional posteriors:

$$\begin{bmatrix} \mu \\ \beta \end{bmatrix} \Big| \sigma_\varepsilon^{-2}, M_i, \mathcal{D}^t \sim \mathcal{N}(\bar{b}, \bar{V}), \quad (\text{B-1})$$

and

$$\sigma_\varepsilon^{-2} | \mu, \beta, M_i, \mathcal{D}^t \sim \mathcal{G}(\bar{s}^{-2}, \bar{v}), \quad (\text{B-2})$$

where

$$\begin{aligned} \bar{V} &= \left[ \underline{V}^{-1} + \sigma_\varepsilon^{-2} \sum_{\tau=1}^{t-1} x_\tau x_\tau' \right]^{-1}, \\ \bar{b} &= \bar{V} \left[ \underline{V}^{-1} \underline{b} + \sigma_\varepsilon^{-2} \sum_{\tau=1}^{t-1} x_\tau r_{\tau+1} \right], \\ \bar{v} &= \underline{v}_0 + (t-1). \end{aligned} \quad (\text{B-3})$$

and

$$\bar{s}^2 = \frac{\sum_{\tau=1}^{t-1} (r_{\tau+1} - \mu - \beta x_\tau)^2 + (s_{r,t}^2 \times \underline{v}_0 (t-1))}{\bar{v}}. \quad (\text{B-4})$$

A Gibbs sampler algorithm can be used to iterate back and forth between (B-1) and (B-2), yielding a series of draws for the parameter vector  $(\mu, \beta, \sigma_\varepsilon^{-2})$ . Draws from the predictive density  $p(r_{t+1} | \mathcal{D}^t)$  can then be obtained by noting that

$$p(r_{t+1} | M_i, \mathcal{D}^t) = \int_{\mu, \beta, \sigma_\varepsilon^{-2}} p(r_{t+1} | \mu, \beta, \sigma_\varepsilon^{-2}, M_i, \mathcal{D}^t) p(\mu, \beta, \sigma_\varepsilon^{-2} | M_i, \mathcal{D}^t) d\mu d\beta d\sigma_\varepsilon^{-2}. \quad (\text{B-5})$$

### Appendix B.2 Time-varying Parameter, Stochastic Volatility Models

Let denote with  $\theta_t$  the time varying parameters,  $\theta_t = (\mu_t, \beta_t)$  and with  $\theta^t$  the sequence of time varying parameters  $\mu_t, \beta_t$  up to time  $t$ ,  $\theta^t = \{\theta_1, \dots, \theta_t\}$ . Similarly, define with  $h^t$  the sequence of excess return log volatilities up to time  $t$ ,  $h^t = \{h_1, \dots, h_t\}$ . Finally, to simplify the notation, let's group all the time invariant parameters of the TVP-SV model into the matrix  $\Theta$ , where  $\Theta = (\mu, \beta, \mathbf{Q}, \sigma_\xi^{-2}, \gamma_\mu, \gamma_\beta, \lambda_0, \lambda_1)$ .

To obtain draws from the joint posterior distribution  $p(\Theta, \theta^t, h^t | M_i', \mathcal{D}^t)$  under the TVP-SV model, we use the Gibbs sampler to draw recursively from the following seven conditional distributions:<sup>23</sup>

---

<sup>23</sup>Using standard set notation, we define  $A_{-b}$  as the complementary set of  $b$  in  $A$ , i.e.  $A_{-b} = \{x \in A : x \neq b\}$ .

1.  $p(\theta^t | \Theta, h^t, M'_i, \mathcal{D}^t)$ .
2.  $p(\mu, \beta | \Theta_{-\mu, \beta}, \theta^t, h^t, M'_i, \mathcal{D}^t)$ .
3.  $p(\mathbf{Q} | \Theta_{-\mathbf{Q}}, \theta^t, h^t, M'_i, \mathcal{D}^t)$
4.  $p(h^t | \Theta, \theta^t, M'_i, \mathcal{D}^t)$ .
5.  $p(\sigma_\xi^{-2} | \Theta_{-\sigma_\xi^{-2}}, \theta^t, h^t, M'_i, \mathcal{D}^t)$
6.  $p(\gamma_\mu, \gamma_\beta | \Theta_{-\gamma_\mu, \gamma_\beta}, \theta^t, h^t, M'_i, \mathcal{D}^t)$
7.  $p(\lambda_0, \lambda_1 | \Theta_{-\lambda_0, \lambda_1}, \theta^t, h^t, M'_i, \mathcal{D}^t)$

We simulate from each of these blocks as follows. Starting with  $\theta^t$ , we focus on  $p(\theta^t | \Theta, h^t, M'_i, \mathcal{D}^t)$ . Define  $\tilde{r}_{\tau+1} = r_{\tau+1} - \mu - \beta x_\tau$  and rewrite (32) as follows:

$$\tilde{r}_{\tau+1} = \mu_\tau - \beta_\tau x_\tau + \exp(h_{\tau+1}) u_{\tau+1} \quad (\text{B-6})$$

Note that knowledge of  $\mu$  and  $\beta$  makes  $\tilde{r}_{\tau+1}$  observable, and reduces (32) to the measurement equation of a standard linear Gaussian state space model with heteroskedastic errors. Thus the sequence of time varying parameters  $\theta^t$  can be drawn from (B-6) using, for example, the algorithm of Carter and Kohn (1994).

Moving on to  $p(\mu, \beta | \Theta_{-\mu, \beta}, \theta^t, h^t, M'_i, \mathcal{D}^t)$ , conditional on  $\theta^t$  it is straightforward to draw  $\mu, \beta$ , by applying standard results. Specifically,

$$\begin{bmatrix} \mu \\ \beta \end{bmatrix} \Big| \Theta_{-\mu, \beta}, \theta^t, h^t, M'_i, \mathcal{D}^t \sim N(\bar{b}, \bar{V}), \quad (\text{B-7})$$

where

$$\begin{aligned} \bar{V} &= \left[ \underline{V}^{-1} + \sum_{\tau=1}^{t-1} \frac{1}{\exp(h_{\tau+1})^2} x_\tau x_\tau' \right]^{-1}, \\ \bar{b} &= \bar{V} \left[ \underline{V}^{-1} \underline{b} + \sum_{\tau=1}^{t-1} \frac{1}{\exp(h_{\tau+1})^2} x_\tau (r_{\tau+1} - \mu_\tau - \beta_\tau x_\tau) \right], \end{aligned} \quad (\text{B-8})$$

As for  $p(\mathbf{Q} | \Theta_{-\mathbf{Q}}, \theta^t, h^t, M'_i, \mathcal{D}^t)$ , we have that

$$\mathbf{Q} | \Theta_{-\mathbf{Q}}, \theta^t, h^t, M'_i, \mathcal{D}^t \sim \mathcal{IW}(\bar{\mathbf{Q}}, 2t - 3), \quad (\text{B-9})$$

where

$$\bar{\mathbf{Q}} = \underline{\mathbf{Q}} + \sum_{\tau=1}^{t-1} (\theta_{\tau+1} - \text{diag}\{\gamma_\mu, \gamma_\beta\} \times \theta_\tau) (\theta_{\tau+1} - \text{diag}\{\gamma_\mu, \gamma_\beta\} \times \theta_\tau)'. \quad (\text{B-10})$$

Moving on to  $p(h^t | \Theta, \theta^t, M'_i, \mathcal{D}^t)$ , we employ the algorithm of [Kim et al. \(1998\)](#). Define  $r_{\tau+1}^* = r_{\tau+1} - (\mu + \mu_{\tau+1}) - (\beta + \beta_{\tau+1})x_\tau$  and note that  $r_{\tau+1}^*$  is observable conditional on  $\mu, \beta$ , and  $\theta^t$ . Next, rewrite (32) as

$$r_{\tau+1}^* = \exp(h_{\tau+1})u_{\tau+1}. \quad (\text{B-11})$$

Squaring and taking logs on both sides of (B-11) yields a new state space system that replaces (32)-(34) with

$$r_{\tau+1}^{**} = 2h_{\tau+1} + u_{\tau+1}^{**}, \quad (\text{B-12})$$

$$h_{\tau+1} = \lambda_0 + \lambda_1 h_\tau + \xi_{\tau+1}, \quad (\text{B-13})$$

where  $r_{\tau+1}^{**} = \ln \left[ (r_{\tau+1}^*)^2 \right]$ , and  $u_{\tau+1}^{**} = \ln(u_{\tau+1}^2)$ , with  $u_\tau^{**}$  independent of  $\xi_s$  for all  $\tau$  and  $s$ . Since  $u_{t+1}^{**} \sim \ln(\chi_1^2)$ , we cannot resort to standard Kalman recursions and simulation algorithms such as those in [Carter and Kohn \(1994\)](#) or [Durbin and Koopman \(2002\)](#). To obviate this problem, [Kim et al. \(1998\)](#) employ a data augmentation approach and introduce a new state variable  $s_{\tau+1}$ ,  $\tau = 1, \dots, t-1$ , turning their focus on drawing from  $p(h^t | \Theta, \theta^t, s^t, M'_i, \mathcal{D}^t)$  instead of  $p(h^t | \Theta, \theta^t, M'_i, \mathcal{D}^t)$ .<sup>24</sup> The introduction of the state variable  $s_{\tau+1}$  allows us to rewrite the linear non-Gaussian state space representation in (B-12)-(B-13) as a linear Gaussian state space model, making use of the following approximation,

$$u_{\tau+1}^{**} \approx \sum_{j=1}^7 q_j \mathcal{N}(m_j - 1.2704, v_j^2), \quad (\text{B-14})$$

where  $m_j, v_j^2$ , and  $q_j$ ,  $j = 1, 2, \dots, 7$ , are constants specified in [Kim et al. \(1998\)](#) and thus need not be estimated. In turn, (B-14) implies

$$u_{\tau+1}^{**} | s_{\tau+1} = j \sim \mathcal{N}(m_j - 1.2704, v_j^2), \quad (\text{B-15})$$

where each state has probability

$$\Pr(s_{\tau+1} = j) = q_j. \quad (\text{B-16})$$

Conditional on  $s^t$ , we can rewrite the nonlinear state space system as follows:

$$\begin{aligned} r_{\tau+1}^{**} &= 2h_{\tau+1} + e_{\tau+1}, \\ h_{\tau+1} &= \lambda_0 + \lambda_1 h_\tau + \xi_{\tau+1}, \end{aligned} \quad (\text{B-17})$$

where  $e_{\tau+1} \sim N(m_j - 1.2704, v_j^2)$  with probability  $q_j$ . For this linear Gaussian state space system, we can use the algorithm of [Carter and Kohn \(1994\)](#) to draw the whole sequence of stochastic volatilities,  $h^t$ .

---

<sup>24</sup>Here  $s^t = \{s_2, s_3, \dots, s_t\}$  denotes the history up to time  $t$  of the new state variable  $s$ .

Finally, conditional on the whole sequence  $h^t$ , draws for the sequence of states  $s^t$  can easily be obtained, noting that

$$\Pr(s_{\tau+1} = j | r_{\tau+1}^{**}, h_{\tau+1}) = \frac{f_h(r_{\tau+1} | 2h_{\tau+1} - m_j + 1.2704, v_j^2)}{\sum_{l=1}^7 f_h(r_{\tau+1} | 2h_{\tau+1} - m_l + 1.2704, v_l^2)}. \quad (\text{B-18})$$

Next, the posterior distribution for  $p(\sigma_\xi^{-2} | \mu, \beta, \theta^t, \mathbf{Q}, h^t, \lambda_0, \lambda_1, \gamma_\mu, \gamma_\beta, M'_i, \mathcal{D}^t)$  is readily available as,

$$\sigma_\xi^{-2} | \Theta_{-\sigma_\xi^{-2}}, \theta^t, h^t, M'_i, \mathcal{D}^t \sim \mathcal{G} \left( \left[ \frac{k_\xi + \sum_{\tau=2}^{t-1} (h_{\tau+1} - \lambda_0 - \lambda_1 h_\tau)^2}{t-1} \right]^{-1}, t-1 \right). \quad (\text{B-19})$$

Finally, obtaining draws from  $p(\gamma_\mu, \gamma_\beta | \Theta_{-\gamma_\mu, \gamma_\beta}, \theta^t, h^t, M'_i, \mathcal{D}^t)$  and  $p(\lambda_0, \lambda_1 | \Theta_{-\lambda_0, \lambda_1}, \theta^t, h^t, M'_i, \mathcal{D}^t)$  is straightforward. As for  $p(\gamma_\mu, \gamma_\beta | \Theta_{-\gamma_\mu, \gamma_\beta}, \theta^t, h^t, M'_i, \mathcal{D}^t)$ ,

$$\gamma_\mu | \Theta_{-\gamma_\mu}, \theta^t, h^t, M'_i, \mathcal{D}^t \sim \mathcal{N}(\bar{m}_{\gamma_\mu}, \bar{V}_{\gamma_\mu}) \times \gamma_\mu \in (-1, 1) \quad (\text{B-20})$$

and

$$\gamma_\beta | \Theta_{-\gamma_\mu}, \theta^t, h^t, M'_i, \mathcal{D}^t \sim \mathcal{N}(\bar{m}_{\gamma_\beta}, \bar{V}_{\gamma_\beta}) \times \gamma_\beta \in (-1, 1) \quad (\text{B-21})$$

where

$$\begin{aligned} \bar{V}_{\gamma_\mu} &= \left[ \underline{V}_\gamma^{-1} + \mathbf{Q}^{11} \sum_{\tau=1}^{t-1} \mu_\tau^2 \right]^{-1}, \\ \bar{m}_{\gamma_\mu} &= \bar{V}_{\gamma_\mu} \left[ \underline{V}_\gamma^{-1} \underline{m}_\gamma + \mathbf{Q}^{11} \sum_{\tau=1}^{t-1} \mu_\tau \mu_{\tau+1} \right], \end{aligned} \quad (\text{B-22})$$

and

$$\begin{aligned} \bar{V}_{\gamma_\beta} &= \left[ \underline{V}_\gamma^{-1} + \mathbf{Q}^{22} \sum_{\tau=1}^{t-1} \beta_\tau^2 \right]^{-1}, \\ \bar{m}_{\gamma_\beta} &= \bar{V}_{\gamma_\beta} \left[ \underline{V}_\gamma^{-1} \underline{m}_\gamma + \mathbf{Q}^{22} \sum_{\tau=1}^{t-1} \beta_\tau \beta_{\tau+1} \right], \end{aligned} \quad (\text{B-23})$$

and where  $\mathbf{Q}^{11}$  and  $\mathbf{Q}^{22}$  and the diagonal elements of  $\mathbf{Q}^{-1}$ . As for  $p(\lambda_0, \lambda_1 | \Theta_{-\lambda_0, \lambda_1}, \theta^t, h^t, M'_i, \mathcal{D}^t)$ , we have that

$$\lambda_0, \lambda_1 | \Theta_{-\lambda_0, \lambda_1}, \theta^t, h^t, M'_i, \mathcal{D}^t \sim \mathcal{N} \left( \left[ \frac{\bar{m}_{\lambda_0}}{\bar{m}_{\lambda_1}} \right], \bar{V}_\lambda \right) \times \lambda_1 \in (-1, 1)$$

where

$$\bar{V}_\lambda = \left\{ \left[ \begin{array}{cc} \underline{V}_{\lambda_0}^{-1} & 0 \\ 0 & \underline{V}_{\lambda_1}^{-1} \end{array} \right] + \sigma_\xi^{-2} \sum_{\tau=1}^{t-1} \left[ \begin{array}{c} 1 \\ h_\tau \end{array} \right] [1, h_\tau] \right\}^{-1} \quad (\text{B-24})$$



and

$$\begin{bmatrix} \bar{m}_{\lambda_0} \\ \bar{m}_{\lambda_1} \end{bmatrix} = \bar{V}_\lambda \left\{ \begin{bmatrix} V_{\lambda_0}^{-1} & 0 \\ 0 & V_{\lambda_1}^{-1} \end{bmatrix} \begin{bmatrix} \underline{m}_{\lambda_0} \\ \underline{m}_{\lambda_1} \end{bmatrix} + \sigma_\xi^{-2} \sum_{\tau=1}^{t-1} \begin{bmatrix} 1 \\ h_\tau \end{bmatrix} h_{\tau+1} \right\}. \quad (\text{B-25})$$

Finally, draws from the predictive density  $p(r_{t+1}|M'_i, \mathcal{D}^t)$  can be obtained by noting than

$$\begin{aligned} p(r_{t+1}|M'_i, \mathcal{D}^t) &= \int_{\Theta, \theta^{t+1}, h^{t+1}} p(r_{t+1}|\theta_{t+1}, h_{t+1}, \Theta, \theta^t, h^t, M'_i, \mathcal{D}^t) \\ &\quad \times p(\theta_{t+1}, h_{t+1}|\Theta, \theta^t, h^t, M'_i, \mathcal{D}^t) \\ &\quad \times p(\Theta, \theta^t, h^t|M'_i, \mathcal{D}^t) d\Theta d\theta^{t+1} dh^{t+1}. \end{aligned} \quad (\text{B-26})$$

To obtain draws for  $p(r_{t+1}|M'_i, \mathcal{D}^t)$ , we proceed in three steps:

1. Draws from  $p(\Theta, \theta^t, h^t|M'_i, \mathcal{D}^t)$  are obtained from the Gibbs sampling algorithm described above;
2. Draws from  $p(\theta_{t+1}, h_{t+1}|\Theta, \theta^t, h^t, M'_i, \mathcal{D}^t)$ : having processed data up to time  $t$ , the next step is to simulate the future volatility,  $h_{t+1}$ , and the future parameters,  $\theta_{t+1}$ . We have that

$$h_{t+1}|\Theta, \theta^t, h^t, M'_i, \mathcal{D}^t \sim N(h_t, \sigma_\xi^2). \quad (\text{B-27})$$

and

$$\theta_{t+1}|\Theta, \theta^t, h^t, M'_i, \mathcal{D}^t \sim N(\theta_t, \mathbf{Q}). \quad (\text{B-28})$$

3. Draws from  $p(r_{t+1}|\theta_{t+1}, h_{t+1}, \Theta, \theta^t, h^t, M'_i, \mathcal{D}^t)$ : we have that

$$r_{t+1}|\theta_{t+1}, h_{t+1}, \Theta, \theta^t, h^t, M'_i, \mathcal{D}^t \sim N((\mu + \mu_{t+1}) + (\beta + \beta_{t+1})x_t, \exp(h_{t+1})). \quad (\text{B-29})$$

## Appendix C Sequential combination

We conclude by briefly summarizing the estimation density combination algorithm proposed in [Billio et al. \(2013\)](#), which we extend with a learning mechanism based on the past economic performance of the individual models entering the combination to obtain posterior distributions for both  $p(r_{t+1}|\tilde{\mathbf{r}}_{t+1}, \mathbf{w}_{t+1}, \mathcal{D}^t)$  and  $p(\mathbf{w}_{t+1}|\tilde{\mathbf{r}}_{t+1}, \mathcal{D}^t)$ .

Let  $\boldsymbol{\theta}$  be the parameter vector of the combination model, that is  $\boldsymbol{\theta} = (\sigma_\kappa^2, \Lambda)$ . Assume that  $\tilde{\mathbf{r}}_s$ ,  $s = 1, \dots, t+1$  is computed using formulas in the previous section and given; define the vector of observable  $r_{1:t} = (r_1, \dots, r_t)' \in \mathcal{D}^t$ , the augmented state vector  $\mathbf{Z}_{t+1} = (\mathbf{w}_{t+1}, \mathbf{z}_{t+1}, \boldsymbol{\theta}_{t+1})$ , where  $\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}$ ,  $\forall t$ . Notice that  $\mathbf{x}_{t+1}$  can be modelled as a stochastic latent process depending on the past forecasting performance of the  $N$  individual prediction models, see equation (12),

or just on its lagged value, see equation (10). Our combination model writes in the state space form as

$$r_t \sim p(r_t|\tilde{\mathbf{r}}_t, \mathbf{Z}_t) \quad (\text{measurement density}) \quad (\text{B-30})$$

$$\mathbf{Z}_t \sim p(\mathbf{Z}_t|\mathbf{Z}_{t-1}, r_{1:t}, \tilde{\mathbf{r}}_t) \quad (\text{transition density}) \quad (\text{B-31})$$

$$\mathbf{Z}_0 \sim p(\mathbf{Z}_0) \quad (\text{initial density}) \quad (\text{B-32})$$

The state predictive and filtering densities, which provide the posterior densities of the combination weights, see equation (B-35) are

$$p(\mathbf{Z}_{t+1}|r_{1:t}, \tilde{\mathbf{r}}_{1:t}) = \int_{\mathbf{Z}_t} p(\mathbf{Z}_{t+1}|\mathbf{Z}_t, r_{1:t}, \tilde{\mathbf{r}}_{1:t})p(\mathbf{Z}_t|r_{1:t}, \tilde{\mathbf{r}}_{1:t})d\mathbf{Z}_t \quad (\text{B-33})$$

$$p(\mathbf{Z}_{t+1}|r_{1:t+1}, \tilde{\mathbf{r}}_{1:t+1}) = \frac{p(r_{t+1}|\mathbf{Z}_{t+1}, \tilde{\mathbf{r}}_{t+1})p(\mathbf{Z}_{t+1}|r_{1:t}, \tilde{\mathbf{r}}_{1:t})}{p(r_{t+1}|r_{1:t}, \tilde{\mathbf{r}}_{1:t})} \quad (\text{B-34})$$

and the marginal predictive density of the observable variables is then

$$p(r_{t+1}|r_{1:t}) = \int_{\tilde{\mathbf{r}}_{t+1}} p(r_{t+1}|r_{1:t}, \tilde{\mathbf{r}}_{t+1})p(\tilde{\mathbf{r}}_{t+1}|r_{1:t})d\tilde{\mathbf{r}}_{t+1}$$

where  $p(r_{t+1}|r_{1:t}, \tilde{\mathbf{r}}_{t+1})$  is defined as

$$\int_{\mathbf{Z}_{t+1}} p(r_{t+1}|\mathbf{Z}_{t+1}, \tilde{\mathbf{r}}_{t+1})p(\mathbf{Z}_{t+1}|r_{1:t}, \tilde{\mathbf{r}}_{1:t})d\mathbf{Z}_{t+1}$$

and represents the conditional predictive density of the observable given the predictors and the past values of the observable.

The analytical solution of the optimal combination problem is generally not known. We use  $M$  parallel conditional SMC filters, where each filter, is conditioned on the predictor vector sequence  $\tilde{\mathbf{r}}_s$ ,  $s = 1, \dots, t$ .

We initialize independently the  $M$  particle sets:  $\Xi_0^j = \{\mathbf{Z}_0^{i,j}, \omega_0^{i,j}\}_{i=1}^N$ ,  $j = 1, \dots, M$ . Each particle set  $\Xi_0^j$  contains  $N$  iid random variables  $\mathbf{Z}_0^{i,j}$  with random weights  $\omega_0^{i,j}$ . We initialize the set of predictors, by generating iid samples  $\tilde{\mathbf{r}}_1^j$ ,  $j = 1, \dots, M$ , from  $p(\tilde{\mathbf{r}}_1|r_0)$  where  $r_0$  is an initial set of observations for the variable of interest. Then, at the iteration  $t + 1$  of the combination algorithm, we approximate the predictive density  $p(\tilde{\mathbf{r}}_{t+1}|r_{1:t})$  with  $M$  iid samples from the predictive densities, and  $\delta_x(y)$  denotes the Dirac mass at  $x$ .

Precisely, we assume an independent sequence of particle sets  $\Xi_t^j = \{\mathbf{Z}_{1:t}^{i,j}, \omega_t^{i,j}\}_{i=1}^N$ ,  $j = 1, \dots, M$ , is available at time  $t$  and that each particle set provides the approximation

$$p_{N,j}(\mathbf{z}_t|r_{1:t}, \tilde{\mathbf{r}}_{1:t}^j) = \sum_{i=1}^N \omega_t^{i,j} \delta_{\mathbf{z}_t^{i,j}}(\mathbf{z}_t) \quad (\text{B-35})$$

of the filtering density,  $p(\mathbf{Z}_t|\mathbf{y}_{1:t}, \tilde{\mathbf{r}}_{1:t}^j)$ , conditional on the  $j$ -th predictor realization,  $\tilde{\mathbf{r}}_{1:t}^j$ . The prediction (including the weights  $\mathbf{w}_{t+1}$ ) are computed using the state predictive  $p(\mathbf{Z}_{t+1}|r_{1:t}, \tilde{\mathbf{r}}_{1:t})$ .

After collecting the results from the different particle sets, it is possible to obtain the following empirical predictive density for the stock returns

$$p_{M,N}(r_{t+1}|r_{1:t}) = \frac{1}{MN} \sum_{j=1}^M \sum_{i=1}^N \omega_t^{i,j} \delta_{r_{t+1}^{i,j}}(r_{t+1}) \quad (\text{B-36})$$

At the next observation,  $M$  independent conditional SMC algorithms are used to find a new sequence of  $M$  particle sets, which include the information available from the new observation and the new predictors.