CENTRE FOR APPLIED MACRO - AND PETROLEUM ECONOMICS (CAMP)

CAMP Working Paper Series
No 9/2015

Optimal Portfolio Choice under Decision-Based Model Combinations

Davide Pettenuzzo and Francesco Ravazzolo

NORWEGIAN BUSINESS SCHOOL

Optimal Portfolio Choice under Decision-Based Model Combinations*

Davide Pettenuzzo[†]
Brandeis University

Francesco Ravazzolo[‡]
Norges Bank, and Centre for Applied Macro and Petroleum economics at BI Norwegian Business School

August 15, 2015

Abstract

We extend the density combination approach of Billio et al. (2013) to feature combination weights that depend on the past forecasting performance of the individual models entering the combination through a utility-based objective function. We apply our model combination scheme to forecast stock returns, both at the aggregate level and by industry, and investigate its forecasting performance relative to a host of existing combination methods. Overall, we find that our combination scheme produces markedly more accurate predictions than the existing alternatives, both in terms of statistical and economic measures of out-of-sample predictability. We also investigate the performance of our model combination scheme in the presence of model instabilities, by considering individual predictive regressions that feature time-varying regression coefficients and stochastic volatility. We find that the gains from using our combination scheme 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, Bradley Paye, Allan Timmermann, and Ross Valkanov, as well as seminar participants at the Narodowy Bank Polski workshop on "Short Term Forecasting Workshop", Norges Bank, Brandeis University, the 2014 EC^2 conference, the 2014 CFE meeting, the 2015 NBER-SBIES meeting, the 2015 SoFiE meeting, and the 2015 IAAE conference for helpful comments. We also thank Jens Hilscher for assisting with the construction of the industry predictors.

[†]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.

[‡]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.

1 Introduction

Over the years, the question of whether stock returns are predictable has received considerable attention, both within academic and practitioner circles.[1] However, to this date, return predictability remains controversial, as emphasized by a number of recent studies on the subject.[2] For example, Welch and Goyal (2008) show that a long list of predictors from the literature is unable to consistently deliver superior out-of-sample forecasts of the equity premium relative to a simple forecast based on the historical average. In their view, the inconsistent out-of-sample performance of these predictors is due to structural instability. Figure 1 provides a graphical illustration of this point, showing the forecast accuracy of a representative subset of the predictors used in this literature, measured relative to the prevailing mean model (both the data and the evaluation criteria used are described in detail in sections 4 and 5). This figure paints a very uncertain and unstable environment for stock returns, where while at times some of the individual predictors appear to outperform the prevailing mean model, no single predictor seems able to consistently deliver superior forecasts.

Forecast combination methods offer a way to improve equity premium forecasts, reducing the uncertainty/instability risk associated with reliance on a single predictor or model.[3] Avramov (2002), Rapach et al. (2010), and Dangl and Halling (2012) confirm this point, and find that simple model combinations lead to improvements in the out-of-sample predictability of stock returns. Interestingly, the existing forecast combination methods weight the individual models entering the combination according to their statistical performance, without making any reference to the way the final forecasts will be put to use. For example, Rapach et al. (2010) propose combining different predictive models 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. We note, however, that in the case of stock returns the quality of the individual model predictions should depend on whether such models lead to profitable investment decisions, which in turns is directly related

[1]See Rapach and Zhou (2013) and references therein for a comprehensive review of the academic literature on aggregate U.S. stock return predictability.

[2]See Boudoukh et al. (2008), Campbell and Thompson (2008), Cochrane (2008), Lettau and Van Nieuwerburgh (2008) and Welch and Goyal (2008).

[3]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. See Timmermann (2006) for a comprehensive review on model combination methods.

to the investor's utility function. This creates a tension between the criterion used to combine the individual predictions and the final use to which the forecasts will be put.[4]

In this paper, we extend the literature on stock return predictability by proposing a model combination scheme where the predictive densities of the individual models are weighted based on how each model fares relative to the investor's utility function, as measured by its implied certainty equivalent return (CER) value. Accordingly, we label this model CER-based Density Combination, or CER-based DeCo in short. To implement this idea, we rely on the approach of Billio et al. (2013), who propose a Bayesian combination approach with time-varying weights, and use a non-linear state space model to estimate them. In addition, we introduce a mechanism that allows the combination weights to depend on the history of the individual models' past profitability, through the individual models' past CER values.

To test our combination scheme empirically, we evaluate how it fares relative to a host of alternative model combination methods, and consider as the individual models entering the combinations both linear and time-varying parameter with stochastic volatility (TVP-SV) models, each including as regressor one of the predictor variables used by Welch and Goyal (2008).[5] When implemented along the lines proposed in this paper, we find that the CER-based DeCo scheme leads to substantial improvements in the predictive accuracy of stock returns, both in statistical and economic terms. In the benchmark case of an investor endowed with power utility and a relative risk aversion of five, we find that the CER-based DeCo scheme yields an annualized CER that is almost 100 basis points higher than any of the competing model combinations. Switching from linear to TVP-SV models produces an increase in CER of more than 150 basis points, and an absolute CER level of 246 basis points. No other model combination scheme comes close to these gains.

Our paper contributes to a rapidly growing literature developing combination schemes with time-varying weights. In particular, our paper is related to the work of Elliott and Timmermann (2005) and Waggoner and Zha (2012), who develop model combination methods where the

[4]This point is closely related to the existing 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.

[5]Johannes et al. (2014) generalize the setting of Welch and Goyal (2008) by forecasting stock returns allowing both regression parameters and return volatility to change over time. However, their emphasis is not on model combination methods, and focus on a single predictor for stock returns, the dividend yield.

weights are driven by a regime switching process, Hoogerheide et al. (2010), Raftery et al. (2010), Koop and Korobilis (2011, 2012), Billio et al. (2013), and Del Negro et al. (2014), who propose model combinations whose weights change gradually over time, and Kapetanios et al. (2015), who develop a combination method where the weights depend on current and past values of the variable being forecasted, as determined by where in the forecast density the variable of interest is realized. To the best of our knowledge, ours is the first Bayesian combination scheme where the weights depend on a utility-based loss function. Our paper is also related to the literature on optimal portfolio choice, and to a number of recent papers, including Sentana (2005), Kan and Zhou (2007), Tu and Zhou (2011), and Paye (2012), exploring the benefits of combining individual portfolio strategies.

The plan of the paper is as follows. Section 2 reviews the standard Bayesian framework for predicting stock returns in the presence of model and parameter uncertainty. Section 3 introduces the CER-based DeCo combination scheme, and highlighs how it differs from the existing combination methods. Next, section 4 describes the data, while section 5 presents the main empirical results for a wide range of predictor variables and model combination strategies. Section 6 generalizes the previous results by introducing time-varying coefficients and stochastic volatility, while Section 7 reports the results of a number robustness checks and extensions, including an application of the methods described in the paper to forecast industry portfolio returns. Finally, section 8 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, x_τ:

$$r_{\tau+1} = \mu + \beta x_\tau + \varepsilon_{\tau+1}, \quad \tau = 1, ..., t-1, \tag{1}$$

where $\varepsilon_{\tau+1} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$. 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, μ and β, 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. See for example the comprehensive study of Welch and

Goyal (2008). They attribute the lack of out-of-sample predictability to a highly uncertain and constantly evolving environment, hard to appropriately characterize using the simple model in (1). In this context, model combination methods offer a valuable alternative. In particular, when implemented using Bayesian methods, model combinations allow to jointly incorporate parameter and model uncertainty into the estimation and inference steps and, compared to (1), promise to be more robust to model misspecifications. More 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 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 $(M_1, \ldots, M_N)$. After eliciting prior distributions on the parameters of each model, she can derive posterior estimates on all the parameters, and use them to obtain N distinct predictive distributions, one for each model entertained. Next, using Bayesian Model Averaging (BMA, henceforth) the individual predictive densities are combined into the predictive distribution $p(r_{t+1}|\mathcal{D}^t)$,

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

where $\mathcal{D}^t$ stands for the information set available at time t, i.e. $\mathcal{D}^t = \{r_{\tau+1}, x_\tau\}_{\tau=1}^{t-1} \cup x_t$, and $P\left(M_i|\, \mathcal{D}^t\right)$ is the posterior probability of model i, derived by Bayes' rule,

$$P\left(M_i|\, \mathcal{D}^t\right) = \frac{P\left(\mathcal{D}^t\middle|\, M_i\right) P\left(M_i\right)}{\sum_{j=1}^{N} P\left(\mathcal{D}^t|\, M_j\right) P\left(M_j\right)}, \qquad i = 1, ..., N \tag{3}$$

$P\left(M_i\right)$ is the prior probability of model M_i, and $P\left(\mathcal{D}^t\middle|\, M_i\right)$ denotes the corresponding marginal likelihood.[6] 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. First, BMA assumes that the true model is included in the model set. Indeed, under this assumption it can be shown that the posterior model probabilities 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 incom-*

[6] See Hoeting et al. (1999) for a review on BMA.

pleteness. Geweke and Amisano (2011) propose replacing the averaging as done in (2)-(3) with a linear prediction pool, where the individual model weights are computed by maximizing the log predictive likelihood, or log score (LS), of the combined model.[7] Geweke and Amisano (2011, 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. Second, BMA assumes 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 be changing over time.[8] Lastly, all existing Bayesian model combination methods, including BMA, are potentially subject to 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 the linear prediction pool), and how ultimately their forecasts are put to use. In particular, all the existing methods weight the 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 seems hardly the case. In fact, with stock returns the quality of the individual model predictions should not be assessed in terms of their statistical fit but rather on whether such predictions lead to profitable investment decisions.[9]

3 Our approach

In this section, we introduce an alternative model combination scheme that addresses the limitations discussed in section 2. We rely on the approach of Billio et al. (2013), who propose a Bayesian combination approach with time-varying weights, and use a non-linear state space model to estimate them. In addition, we introduce a mechanism that allows the combination weights to depend on the history of the individual models' past profitability. We now turn to explaining in more details how our model combination scheme works.

We continue to assume that at a generic point in time t, the investor has available N distinct

[7]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.

[8]The linear prediction pool of Geweke and Amisano (2011) also imposes time-invariant model combination weights. Del Negro et al. (2014) develop a dynamic version of the linear prediction pool approach which they show works well for combinations of two models. See also Waggoner and Zha (2012) and Billio et al. (2013).

[9]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.

models to predict excess returns, each model producing a predictive distribution $p\left(r_{t+1}\middle| M_i, \mathcal{D}^t\right)$, with $i = 1, ..., N$. To ease the notation in what follows, we first define with $\widetilde{\mathbf{r}}_{t+1} = (\widetilde{r}_{1,t+1}, ..., \widetilde{r}_{N,t+1})'$ the $N \times 1$ vector of predictions made at time t and with $p\left(\widetilde{\mathbf{r}}_{t+1}\middle| \mathcal{D}^t\right)$ its joint predictive density. Next, we write the composite predictive distribution $p(r_{t+1}|\mathcal{D}^t)$ as

$$p\left(r_{t+1}\middle| \mathcal{D}^t\right) = \int p(r_{t+1}|\widetilde{\mathbf{r}}_{t+1}, \mathbf{w}_{t+1}, \mathcal{D}^t)p(\mathbf{w}_{t+1}|\widetilde{\mathbf{r}}_{t+1}, \mathcal{D}^t)p\left(\widetilde{\mathbf{r}}_{t+1}\middle| \mathcal{D}^t\right) d\widetilde{\mathbf{r}}_{t+1}d\mathbf{w}_{t+1} \tag{4}$$

where $p(r_{t+1}|\widetilde{\mathbf{r}}_{t+1}, \mathbf{w}_{t+1}, \mathcal{D}^t)$ denotes the combination scheme based on the N predictions $\widetilde{\mathbf{r}}_{t+1}$ and the combination weights $\mathbf{w}_{t+1} \equiv (w_{1,t+1}, \ldots, w_{N,t+1})'$, and $p(\mathbf{w}_{t+1}|\widetilde{\mathbf{r}}_{t+1}, \mathcal{D}^t)$ denotes the posterior distribution of the combination weights $\mathbf{w}_{t+1}$. Equation (4) generalizes equation (2), taking into account the limitations discussed in the previous section. First, by specifying a stochastic process for the model combination scheme, $p(r_{t+1}|\widetilde{\mathbf{r}}_{t+1}, \mathbf{w}_{t+1}, \mathcal{D}^t)$, we allow for either model misspecification, or incompleteness, in the combination. Second, by introducing a proper distribution for $\mathbf{w}_{t+1}$, $p(\mathbf{w}_{t+1}|\widetilde{\mathbf{r}}_{t+1}, \mathcal{D}^t)$, we allow for the combination weights to change over time and, as we will show in subsection 3.2, to be driven by the individual models' past profitability.[10]

3.1 Individual models

We begin by describing how we specify the last term on the right-hand side of (4), $p\left(\widetilde{\mathbf{r}}_{t+1}\middle| \mathcal{D}^t\right)$, which we remind is short-hand for the N distinct predictive distributions entering the 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. As in (1), we project excess returns, $r_{\tau+1}$, on a lagged predictor, x_τ, where $\tau = 1, ..., t-1$.[11] Next, to estimate the model parameters, we follow Koop (2003, Section 4.2) and specify independent Normal-Inverse Gamma (NIG) priors on the parameter vector $\left(\mu, \beta, \sigma_\varepsilon^{-2}\right)$. Next, we rely on a Gibbs sampler to draw from the conditional posterior distributions of μ, β, and σ_ε^{-2}, given the information set available at time t, $\mathcal{D}^t$. Finally, once draws from the posterior distributions of μ, β, and σ_ε^{-2} are available, we use them to form a predictive density for r_{t+1} in the following way:

$$p\left(r_{t+1}\middle| M_i, \mathcal{D}^t\right) = \int p\left(r_{t+1}\middle| \mu, \beta, \sigma_\varepsilon^{-2}, M_i, \mathcal{D}^t\right) p\left(\mu, \beta, \sigma_\varepsilon^{-2}\middle| M_i, \mathcal{D}^t\right) d\mu d\beta d\sigma_\varepsilon^{-2}. \tag{5}$$

[10]Note also that the combination scheme in (4) 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.

[11]Note that x_τ can either be a scalar or a vector of regressors. 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 at once in (1), but we follow the bulk of the literature on stock return predictability and focus on a single predictor at a time.

Repeating this process for the N individual models entering the combination yields the joint predictive distribution $p\left(\widetilde{\mathbf{r}}_{t+1}\middle|\,\mathcal{D}^t\right)$. We refer the reader to an online appendix for more details on the specification of the priors, the implementation of the Gibbs sampler, and the evaluation of the integral in equation (5).

3.2 Combination weights

We now turn to describing how we specify the conditional density for the combination weights, $p(\mathbf{w}_{t+1}|\widetilde{\mathbf{r}}_{t+1}, \mathcal{D}^t)$. First, in order to have time-varying weights $\mathbf{w}_{t+1}$ that belong to the simplex $\Delta_{[0,1]^N}$, we introduce a vector of latent processes $\mathbf{z}_{t+1} = (z_{1,t+1}, \ldots, z_{N,t+1})'$, where N is the total number of models considered in the combination scheme, and[12]

$$w_{i,t+1} = \frac{\exp\{z_{i,t+1}\}}{\sum_{l=1}^{N} \exp\{z_{l,t+1}\}}, \quad i = 1, \ldots, N \tag{6}$$

Next, we need the combination weights to depend on the past profitability of the N individual models entering the combination. To accomplish this, we specify the following stochastic process for $\mathbf{z}_{t+1}$:

$$\begin{aligned} \mathbf{z}_{t+1} &\sim p(\mathbf{z}_{t+1}|\mathbf{z}_t, \Delta\boldsymbol{\zeta}_t, \boldsymbol{\Lambda}) \\ &\propto |\boldsymbol{\Lambda}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\mathbf{z}_{t+1} - \mathbf{z}_t - \Delta\boldsymbol{\zeta}_t\right)' \boldsymbol{\Lambda}^{-1} \left(\mathbf{z}_{t+1} - \mathbf{z}_t - \Delta\boldsymbol{\zeta}_t\right)\right\} \end{aligned} \tag{7}$$

where $\boldsymbol{\Lambda}$ is an $(N \times N)$ diagonal matrix, and $\Delta\boldsymbol{\zeta}_t = \boldsymbol{\zeta}_t - \boldsymbol{\zeta}_{t-1}$, with $\boldsymbol{\zeta}_t = (\zeta_{1,t}, \ldots, \zeta_{N,t})'$ denoting a distance vector, measuring the accuracy of the N prediction models up to time t.[13] As for the individual elements of $\boldsymbol{\zeta}_t$, we opt for an exponentially weighted average of the past performance of the N individual models entering the combination,

$$\zeta_{i,t} = (1-\lambda) \sum_{\tau=\underline{t}+1}^{t} \lambda^{t-\tau} f\left(r_\tau, \widetilde{r}_{i,\tau}\right), \quad i = 1, \ldots, N \tag{8}$$

where $\underline{t}+1$ denotes the beginning of the evaluation period, $\lambda \in (0, 1)$ is a smoothing parameter, $f\left(r_\tau, \widetilde{r}_{i,\tau}\right)$ is a measure of the accuracy of model i, and $\widetilde{r}_{i,\tau}$ denotes the one-step ahead density forecast of r_τ made by model i at time $\tau - 1$. $\widetilde{r}_{i,\tau}$ is thus short-hand for the i-th element of $p\left(\widetilde{\mathbf{r}}_\tau\middle|\,\mathcal{D}^{\tau-1}\right)$, $p(r_\tau|M_i, \mathcal{D}^{\tau-1})$. We set $\lambda = 0.95$ in our main analysis, and report in section 7 the

[12] Under this convexity constraint, the weights can be interpreted as discrete probabilities over the set of models entering the combination.

[13] We assume that the variance-covariance matrix $\boldsymbol{\Lambda}$ of the process $\mathbf{z}_{t+1}$ governing the combination weights is diagonal. We leave to further research the possibility of allowing for cross-correlation between model weights.

effect of altering this value.[14] As for the specific choice of $f\left(r_{\tau}, \widetilde{r}_{i,\tau}\right)$, we focus on a utility-based measure of predictability, the certainty equivalent return (CER).[15] 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, her CER is given by

$$f\left(r_{\tau}, \widetilde{r}_{i,\tau}\right) = \left[(1-A)\, U\left(W_{i,\tau}^{*}\right)\right]^{1/(1-A)} \tag{9}$$

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

$$U\left(W_{i,\tau}^{*}\right) = \frac{\left[\left(1-\omega_{i,\tau-1}^{*}\right)\exp\left(r_{\tau-1}^{f}\right) + \omega_{i,\tau-1}^{*}\exp\left(r_{\tau-1}^{f} + r_{\tau}\right)\right]^{1-A}}{1-A} \tag{10}$$

$r_{\tau-1}^{f}$ denotes the continuously compounded Treasury bill rate known at time $\tau-1$, A stands for the investor's relative risk aversion, r_{τ} is the realized excess return at time τ, and $\omega_{i,\tau-1}^{*}$ denotes the optimal allocation to stocks according to the prediction made for r_{τ} by model M_i, and given by the solution to[16]

$$\omega_{i,\tau-1}^{*} = \arg\max_{\omega_{\tau-1}} \int U\left(\omega_{\tau-1}, r_{\tau}\right) p(r_{\tau}|M_i, \mathcal{D}^{\tau-1}) dr_{\tau} \tag{11}$$

Combined, equations (6)–(9) imply that the combination weight of model i at time $t+1$, $w_{i,t+1}$, depends in a non-linear fashion on the time t combination weight $w_{i,t}$ and on an exponentially weighted sum of model i's past CER values. Accordingly, we label the model combination in (4) "CER-based Density Combination", or "CER-based DeCo" in short.

3.3 Combination scheme

We now turn to the first term on the right hand side of (4), $p\left(r_{t+1}|\widetilde{\mathbf{r}}_{t+1}, \mathbf{w}_{t+1}, \mathcal{D}^{t}\right)$, denoting the combination scheme. We note that since both the N original densities $p\left(\widetilde{\mathbf{r}}_{t+1}|\, \mathcal{D}^{t}\right)$ and the combination weights $\mathbf{w}_{t+1}$ are in the form of densities, the combination scheme for $p(r_{t+1}|\widetilde{\mathbf{r}}_{t+1}, \mathbf{w}_{t+1}, \mathcal{D}^{t})$ is based on a convolution mechanism, which guarantees that the product of N predictive densities with the combination weights results in a proper density.[17] Following

[14] We note that in principle the parameter λ could be estimated from the data, and one possibility would be to rely on a grid search to estimate it. Billio et al. (2013, section 6.2) discuss this option.

[15] Utility-based loss functions have been adopted before by Brown (1976), Frost and Savarino (1986), Stambaugh (1997), and Ter Horst et al. (2006) to evaluate portfolio rules.

[16] Throughout the paper, we restrict the allocation to the interval $0 \leq \omega_{\tau-1} < 1$, thus precluding short selling and buying on margin. See for example Barberis (2000).

[17] We refer the reader to Aastveit et al. (2014) for further discussion on convolution and its properties.

Billio et al. (2013), we apply a Gaussian combination scheme,

$$p(r_{t+1}|\widetilde{\mathbf{r}}_{t+1}, \mathbf{w}_{t+1}, \sigma_\kappa^{-2}) \propto \exp\left\{-\frac{1}{2}\left(r_{t+1} - \widetilde{\mathbf{r}}_{t+1}\mathbf{w}_{t+1}\right)' \sigma_\kappa^{-2} \left(r_{t+1} - \widetilde{\mathbf{r}}_{t+1}\mathbf{w}_{t+1}\right)\right\} \tag{12}$$

The combination relationship in (12) is linear and explicitly allows for model misspecification, possibly because all models in the combination may be false (incomplete model set or open model space). Furthermore, the combination residuals are estimated and their distribution follows a Gaussian process with mean zero and standard deviation σ_κ, providing a probabilistic measure of the incompleteness of the model set.[18,19]

We conclude this section by briefly describing how we estimate the posterior distributions $p(r_{t+1}|\widetilde{\mathbf{r}}_{t+1}, \mathbf{w}_{t+1}, \mathcal{D}^t)$ and $p(\mathbf{w}_{t+1}|\widetilde{\mathbf{r}}_{t+1}, \mathcal{D}^t)$.[20] Equations (4), (6), (7), and (12), as well as the individual model predictive densities $p\left(\widetilde{\mathbf{r}}_{t+1}\middle| \mathcal{D}^t\right)$ are first grouped into a non-linear state space model.[21] 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 $p(r_{t+1}|\widetilde{\mathbf{r}}_{t+1}, \mathbf{w}_{t+1}, \mathcal{D}^t)$ and $p(\mathbf{w}_{t+1}|\widetilde{\mathbf{r}}_{t+1}, \mathcal{D}^t)$. We refer the reader to an online appendix for more details on the prior specifications for $\mathbf{\Lambda}$ and σ_κ^{-2}, and the implementation of the sequential Monte Carlo.

4 Data

Our empirical analysis uses data on stock returns along with a set of fifteen economic variables which are popular stock return predictors and are directly linked to economic fundamentals and risk aversion. 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. As for the predictors, we use updated data from Welch and Goyal (2008), extending from January 1927

[18] 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.

[19] It is 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 (4) reduces to standard BMA. Hence, one can think of BMA as a special case of the combination scheme we propose here.

[20] As for all individual model parameters and their predictive densities $p\left(\widetilde{\mathbf{r}}_{t+1}\middle| \mathcal{D}^t\right)$, these are computed in a separate step before the model combination weights are estimated, as described in subsection 3.1. Hence, our approach differs from Waggoner and Zha (2012), as they implement a formal mixture between the individual candidate models, and is instead more along the lines of BMA and the optimal prediction pool of Geweke and Amisano (2011).

[21] The non-linearity is due to the logistic transformation mapping the latent process $\mathbf{z}_{t+1}$ into the model combination weights $\mathbf{w}_{t+1}$.

to December 2010. Most of the predictors 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.[22,23]

We use the first 20 years of data as a training sample for both the priors and the forecasts. Specifically, all priors hyperparameters are calibrated over this initial period, and held constant throughout the forecast evaluation period. As for the forecasts, we begin by estimating all 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 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.

5 Out-of-Sample Performance

In this section we answer the question of whether the CER-based DeCo model introduced in section 3 produces equity premium forecasts that are more accurate than those obtained from

[22] We follow Welch and Goyal (2008) and lag inflation an extra month to account for the delay in CPI releases.

[23] Johannes et al. (2014) find that accounting for net equity repurchases in addition to cash payouts produces a stronger predictor for equity returns.

the existing approaches, both in terms of statistical and economic criteria.

5.1 Statistical Performance

We compare the performance of CER-based DeCo to both the fifteen univariate models entering the combination as well as a number of alternative model combination methods, namely BMA, the optimal prediction pool of Geweke and Amisano (2011), and the equal weighted combination, and consider several evaluation statistics for both point and density forecasts. 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.[24]

As for assessing the accuracy of the point forecasts, we consider the Cumulative Sum of Squared prediction Error Difference (CSSED), introduced by Welch and Goyal (2008),

$$CSSED_{m,t} = \sum_{\tau=\underline{t}+1}^{t} \left(e^2_{PM,\tau} - e^2_{m,\tau}\right) \tag{13}$$

where m denotes the model under consideration (either univariate or model combination), and $e_{m,\tau}$ ($e_{PM,\tau}$) denotes model m's (PM's) prediction error from time τ forecast, obtained by synthesizing the corresponding predictive density 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. Next, 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,

$$R^2_{OoS,m} = 1 - \frac{\sum_{\tau=\underline{t}+1}^{\bar{t}} e^2_{m,\tau}}{\sum_{\tau=\underline{t}+1}^{\bar{t}} e^2_{PM,\tau}}. \tag{14}$$

whereby a positive $R^2_{OOS,m}$ 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 three 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,

$$LSD_m = \frac{\sum_{\tau=\underline{t}+1}^{\bar{t}} \left(LS_{m,\tau} - LS_{PM,\tau}\right)}{\sum_{\tau=\underline{t}+1}^{\bar{t}} LS_{PM,\tau}} \tag{15}$$

[24] For consistency, the prevailing mean model is estimated using priors that are analog to those we used for the model in (1). In particular, we slightly alter the prior on (μ, β) to impose a dogmatic "no predictability" prior on $\beta = 0$, while using the same prior for σ_ε^{-2}.

where $LS_{m,\tau}$ ($LS_{PM,\tau}$) denotes model m's (PM's) log predictive score computed at time τ. If LSD_m is positive, this indicates that on average the alternative model m produces more accurate density forecasts than the benchmark PM model. 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 mth model, $CLSD_{m,t} = \sum_{\tau=\underline{t}+1}^{t} (LS_{m,\tau} - LS_{PM,\tau})$. Again, 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. Lastly, 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),

$$CRPSD_m = \frac{\sum_{\tau=\underline{t}+1}^{\bar{t}} (CRPS_{PM,\tau} - CRPS_{m,\tau})}{\sum_{\tau=\underline{t}+1}^{\bar{t}} CRPS_{PM,\tau}} \tag{16}$$

where $CRPS_{m,\tau}$ ($CRPS_{PM,\tau}$) measures the average distance between the empirical cumulative distribution function (CDF) of r_τ (which is simply a step function in r_τ), and the empirical CDF that is associated with model m's (PM's) predictive density.[25]

Table 1 presents results on the accuracy of both point and density forecasts for all fifteen univariate models and a variety of model combination methods, including the CER-based DeCo scheme introduced in section 3. For all statistical metrics considered, positive values indicate that the alternative models perform better than the PM model. We also report stars to summarize the statistical significance of the results, where the underlying p-values are based on the Diebold and Mariano (1995) test of equal equal predictive accuracy 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 columns under the header "Linear". We will return later to the remaining columns of this table. Starting with the top part of panel A, the results for the point forecast accuracy of the individual models are reminiscent of the findings of Welch and Goyal (2008), where the R^2_{OoS}-values are negative for 13 out of the 15 predictor variables. Moving on to bottom part of panel A, we find that with the exception of the optimal prediction pool method of Geweke and Amisano (2011), model combinations lead to positive R^2_{OoS}-values. We note in particular that the CER-based DeCo model delivers the largest improvement in forecast performance among all model combinations, with an R^2_{OoS} of

[25] 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.

2.32%, statistically significant at the 1% level. The top two panels of Figure 2 plot the $CSSED$s for all the model combination methods considered. In particular, the second panel shows that with the exception of the first part of the 1990's, the CER-based DeCo scheme consistently outperforms the PM benchmark as well as all the alternative combination methods.

Turning next to the density forecast results in panels B and C of Table 1, we find that the CER-based DeCo scheme is the only model that yields positive and statistically significant results. This is true for both measures of density forecast accuracy, the average log score differential and the average CRPS differential. To shed light on the reasons for such improvements in both point and density predictability, we also compute a version of CER-based DeCo where we suppress the learning mechanism in the weight dynamics (that is, we remove the term $\Delta\boldsymbol{\zeta}_t$ from (7)). We label this combination scheme "DeCo". A quick look at the comparison between the CER-based DeCo and the DeCo results in Table 1 reveals that the learning mechanism introduced via equations (7)–(9) explains the lion's share of the increase in performance we see for the CER-based DeCo scheme. As for the point forecast improvement, this can also be seen by inspecting the gap between the CER-based DeCo (red dashed line) and DeCo (blue solid line) $CSSED$s displayed in the second panel of Figure 2.

5.2 Economic Performance

We now turn to evaluating the economic significance of the return forecasts by considering the portfolio choice of an investor who uses the forecasts to guide her investment decisions.[26] Having computed the optimal asset allocation weights for all the individual models and the various model combinations, we assess the economic predictability of all models by computing their implied (annualized) CER values. Under power utility, the investor's annualized CER is given by

$$CER_m = 12 \times \left[(1-A) \frac{1}{\bar{t} - \underline{t}} \sum_{\tau=\underline{t}+1}^{\bar{t}} U\left(W^{*}_{m,\tau}\right) \right]^{1/(1-A)} - 1 \tag{17}$$

where m denotes the model under consideration (either univariate or model combination). We next define the differential certainty equivalent return of model m, relative to the benchmark PM model, $CERD_m = CER_m - CER_{PM}$. We interpret a positive $CERD_m$ as evidence that

[26] One advantage of adopting a Bayesian approach is that it yields predictive densities that account for parameter estimation error. 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.

model m generates a higher (certainty equivalent) return than the benchmark model.

Panel A of Table 2 shows annualized $CERD$s for the same models listed in Table 1, assuming a coefficient of relative risk aversion of $A = 5$. Once again, we focus on the columns under the header "Linear", and for the time being restrict our focus to Panel A (we will return to the results in Panel B of this table in the next section). An inspection of the bottom half of panel A reveals that the statistical gains we saw for the CER-based DeCo scheme in Table 1 translate into CER gains of almost 100 basis points. No other combination scheme provides gains of a magnitude comparable to the CER-based DeCo scheme. Turning to the top part of panel A, it appears that some of the individual models generates positive CERD values, but in general these gains are at least 50 basis points smaller than the CER-based DeCo. Finally, the top two panels of Figure 3 plot the cumulative CER values of the various model combination schemes, relative to the PM benchmark. These plots parallel the cumulated differential plots of Figure 2. The figure shows how the economic performance of the CER-based 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 second panel of Figure 3. The only exception is during the second part of the 1990s, where the PM benchmark appears to outperform the CER-based DeCo model. Also, a comparison of the CER-based DECo with the DeCo scheme reveals once again that it is the learning mechanism introduced via equations (7)–(9) that is mainly responsible for these gains.

Along these lines, it would be informative to see whether the CERD of any of the alternative model combinations thus far considered could be improved by adding a similar CER-based learning feature into the calculation of its combination weights. To test this conjecture, we add to the set of model combinations a linear pool whose combination weights depend on the individual models' past profitability in the following way:

$$\widetilde{w}_i = \frac{\Delta\zeta_{i,t}}{\mathbf{1}'\Delta\boldsymbol{\zeta}_t}, \quad i = 1, \ldots, N \tag{18}$$

and where $\mathbf{1}$ is an $(N \times 1)$ unit vector.[27] We label this new combination scheme "CER-based linear pool", and report its annualized CERD at the bottom of Table 2. Comparing the results of the CER-based linear pool with the prediction pool of Geweke and Amisano (2011), we notice a significant improvement in CERD, which increases from -0.82% to -0.03%. We notice,

[27]The approach we propose in Equation 18 is similar to the recursive logarithmic score weight (RW) approach discussed in Jore et al. (2010). See also Amisano and Giacomini (2007) and Hall and Mitchell (2007).

however, that the CER-based linear pool does not improve over the equal weighted combination and BMA methods and, most importantly, falls significantly below the CER-based DeCo. It appears, therefore, that while having a utility-based learning mechanism in the formula for the combination weights can be quite beneficial, the gains we saw for the CER-based DeCo scheme are the result of an ensemble of features, including time-variation in the combination weights, modeling incompleteness, and the addition of a learning mechanism based on the individual models' past profitability.

6 Modeling Parameter Instability

Recent contributions to the literature on stock return predictability have found that it is important to account for two features. First, return volatility varies over time and time varying volatility models fit returns data far better than constant volatility models; see, e.g., Johannes et al. (2014) and Pettenuzzo et al. (2014). Stochastic volatility models can also account for fat tails—a feature that is clearly present in the monthly returns data. Second, the parameters of return predictability models are not stable over time but appear to undergo change; see Paye and Timmermann (2006), Pettenuzzo and Timmermann (2011), Dangl and Halling (2012), and Johannes et al. (2014). 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 combination is an aspect that has not yet been thoroughly investigated.

In this section, we extend the model in (1) 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 over time:

$$r_{\tau+1} = (\mu + \mu_{\tau+1}) + (\beta+\beta_{\tau+1})\, x_\tau + \exp\left(h_{\tau+1}\right) u_{\tau+1}, \quad \tau = 1, ..., t-1, \tag{19}$$

where $h_{\tau+1}$ denotes the (log of) stock return volatility at time $\tau+1$, and $u_{\tau+1} \sim \mathcal{N}(0,1)$. We assume that the time-varying parameters $\boldsymbol{\theta}_{\tau+1} = (\mu_{\tau+1}, \beta_{\tau+1})'$ follow a zero-mean, stationary process

$$\boldsymbol{\theta}_{\tau+1} = \boldsymbol{\gamma}_{\boldsymbol{\theta}}' \boldsymbol{\theta}_\tau + \boldsymbol{\eta}_{\tau+1}, \quad \boldsymbol{\eta}_{\tau+1} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{Q}\right), \tag{20}$$

where $\boldsymbol{\theta}_1 = \mathbf{0}$ and the elements in $\boldsymbol{\gamma}_{\boldsymbol{\theta}}$ are restricted to lie between -1 and 1.[28] The log-volatility $h_{\tau+1}$ is also assumed to follow a stationary and mean reverting process:

$$h_{\tau+1} = \lambda_0 + \lambda_1 h_\tau + \xi_{\tau+1}, \quad \xi_{\tau+1} \sim \mathcal{N}\left(0, \sigma_\xi^2\right), \tag{21}$$

where $|\lambda_1| < 1$ and u_τ, $\boldsymbol{\eta}_t$ and ξ_s are mutually independent for all τ, t, and s.

To estimate the model in (19)-(21), we first specify priors for all the parameters, μ, $\boldsymbol{\beta}$, $\boldsymbol{\theta}^t$, h^t, $\mathbf{Q}$, σ_ξ^{-2}, $\boldsymbol{\gamma}_{\boldsymbol{\theta}}$, λ_0, and λ_1. Next, we use a Gibbs sampler to draw from the conditional posterior distributions of all the parameters.[29] These draws are used to compute density forecasts for r_{t+1} as follows:

$$\begin{aligned} p\left(r_{t+1} \middle| M_i', \mathcal{D}^t\right) &= \int p\left(r_{t+1} \middle| \boldsymbol{\theta}_{t+1}, h_{t+1}, \boldsymbol{\Theta}, \boldsymbol{\theta}^t, h^t, M_i', \mathcal{D}^t\right) \\ &\quad \times p\left(\boldsymbol{\theta}_{t+1}, h_{t+1} \middle| \boldsymbol{\Theta}, \boldsymbol{\theta}^t, h^t, M_i', \mathcal{D}^t\right) \\ &\quad \times p\left(\boldsymbol{\Theta}, \boldsymbol{\theta}^t, h^t \middle| M_i', \mathcal{D}^t\right) d\boldsymbol{\Theta} d\boldsymbol{\theta}^{t+1} dh^{t+1}. \end{aligned} \tag{22}$$

where $\boldsymbol{\Theta} = \left(\mu, \beta, \mathbf{Q}, \sigma_\xi^{-2}, \boldsymbol{\gamma}_{\boldsymbol{\theta}}, \lambda_0, \lambda_1\right)$ contains the time-invariant parameters. We refer the reader to an online appendix for more details on the specification of the priors, the implementation of the Gibbs sampler, and the evaluation of the integral in (22).

Having produced the full set of predictive densities for the N distinct TVP-SV models, we use them to recompute all model combinations, including the CER-based DeCo scheme introduced in Section 3. Point and density forecast results for both the individual TVP-SV models as well as all the newly computed model combinations are reported in Table 1, under the column header "TVP-SV". Starting from the top half of Table 1 and focusing on panel A, we find that allowing for time-varying coefficients and volatilities leads to improvements in forecasting ability for almost all predictors. We note however that the R^2_{OoS} are still mostly negative, implying that at least in terms of point-forecast accuracy it remains very hard to beat the benchmark PM model. Moving on to the bottom of panel A, we find positive R^2_{OoS} for all model combinations methods, with the exception of the optimal prediction pool of Geweke and Amisano (2011). In particular, the R^2_{OoS} of the CER-based DeCo method remains large and significant, though we note a marginal decrease from the results based on the linear models. The bottom two panels

[28] Note that this is equivalent to writing $r_{\tau+1} = \widetilde{\mu}_{\tau+1} + \widetilde{\beta}_{\tau+1} x_\tau + \exp\left(h_{\tau+1}\right) u_{\tau+1}$, where $\left(\widetilde{\mu}_1, \widetilde{\beta}_1\right)$ is left unrestricted.

[29] In particular, we follow Primiceri (2005) after adjusting for the correction to the ordering of steps detailed in Del Negro and Primiceri (2014), and employ the algorithm of Carter and Kohn (1994) along with the approximation of Kim et al. (1998) to draw the history of stochastic volatilities.

of Figure 2 plot the $CSSED_t$ for all the TVP-SV based model combinations, and in particular the fourth panel of the figure shows that the CER-based DeCo outperforms the benchmark PM model throughout the whole forecast evaluation period.

Turning next to the density forecast results in panels B and C of Table 1, we find that allowing for instabilities in the individual models' coefficients and volatilities leads in all cases to improved density forecasts, with all comparison with the PM benchmark being significant at the 1% critical level. Moving on to the bottom halves of panels B and C, we find that for the CRPS measure the CER-based DeCo model generates the largest gains among all model combination methods, while for the log score measure the CER-based DeCo model ranks above the equal weighted combination, BMA, and DeCo but falls slightly below the Optimal prediction pool.[30] The stark contrast between the point and density forecast results in Table 1 is suggestive of the importance of also looking at metrics summarizing the accuracy 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 (2012) in a similar setting.

Moving on to the TVP-SV results in panel A of Table 2, we find that in all cases switching from linear to TVP-SV models produces large improvements in CERDs. This is true for the individual models, whose CERD values relative to the linear case increase on average by 96 basis points, and for the model combinations, whose CERD values increase on average by 140 basis points. As for the individual models, this result is in line with the findings of Johannes et al. (2014), but generalized to a richer set of predictors than those considered in their study. As for the model combinations, we note that the CER-based DeCo model produces the largest CERD, with a value of 246 basis points. This CERD value is more than twice the average CERD generated by the individual TVP-SV models entering the combination. The bottom two panels of Figure 3 offers a graphical illustration of the CERD results summarized in Table 2 for the TVP-SV based model combinations, showing over time the economic performance of the TVP-SV combination methods, relative to the PM benchmark. In particular, the fourth panel of Figure 3 shows that the cumulated CERD value at the end of the sample for the CER-based DeCo is approximately equal to 200%. This exceeds all other model combinations by approximately 40%.

[30]Interestingly, we also find that for the LSD metric, the individual model based on the Stock variance predictor yields a log score differential value of 11.81%, higher than the CER-based DeCo. In a non-reported set of results, we find that if the learning mechanism in equations (7)-(9) is modified to use the individual model past log score histories (i.e. $f(r_\tau, \widetilde{r}_{i,\tau}) = LS_{i,\tau}$), the resulting model combination LSD increases from 11.72% to 12.26% (and from 0.26% to 0.38% in the linear case).

One possible explanation for the improved CERD results we find for the TVP-SV models may have to do with the 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 modify our benchmark model to include stochastic volatility. We label this new benchmark Prevailing Mean with Stochastic Volatility, or PM-SV, and in panel B of Table 3 report the adjusted differential CER, $CERD'_m = CER_m - CER_{PM-SV}$. A quick comparison between panels A and B of Table 3 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 panels. Most notably, the CER-based DeCo results remain quite strong even after replacing the benchmark model, especially for the case of TVP-SV models, with a CERD of 168 basis points.

7 Robustness and Extensions

In this section we summarize the results of a number of robustness and extensions we have performed to validate the empirical results presented in sections 5 to 6. Additional details on these analysis can be found in an online appendix that accompanies the paper. In there, we also summarized the results of an extensive prior sensitivity to ascertain the role of our baseline prior choices on the overall results.

7.1 Robustness analysis

First, we investigated the effect on the profitability analysis presented in sections 5.2 and 6 of altering the investor's relative risk aversion coefficient A. We find that 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 it for the linear models. On the other hand, increasing the risk aversion coefficient to $A = 10$ leads to an overall decrease in CERD values, both for the individual models and the model combinations. In both cases, the CER-based DeCo scheme continues to dominate all the other methods. Second, we performed a subsample analysis to shed light on the robustness of our results to the choice of the forecasting evaluation period. In particular, we looked separately at recessions and expansions, as defined using NBER dating conventions, and we also used the 1973-1975 oil shock period to break the evaluation sample into two separate subsamples. We find that the CER-based DeCo scheme yields positive and large economic gains in all sub-periods, for both linear and TVP-SV models. Third, we modified our choice of the parameter λ controlling the degree of learning in the model combination weights and find that setting it to a lower value, $\lambda = 0.9$, has only minor consequences on the results. We find that this holds true for both the linear and TVP-SV models, and across all sub-periods.[31] Finally, we explored the sensitivity of our baseline 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. We find that the economic gains for power utility and mean variance utility are very similar in magnitude, and that under mean variance utility the CER-based DeCo scheme still produces sizable improvements in CERD relative to all the alternative models, especially in the case of the TVP-SV models.

7.2 Forecasting industry portfolios

We conclude our empirical analysis by investigating the performance of the CER-based DeCo scheme with a number of industry portfolios. While there is a vast literature examining the out-of-sample predictability of U.S. aggregate returns, analysis of out-of-sample return predictability for industry portfolios is relatively rare. Two notable exceptions are Rapach et al. (2015) and Huang et al. (2015). Relative to these studies, our focus is specifically on the predictive ability of the model combinations. We thus focus on a smaller set of industry portfolios, while at the same time significantly expanding the number of predictors used. The latter endeavor is necessary to fully take advantage of the model combination methods. In particular, we still rely on market-wide measures of bond yields and inflation, but in addition we construct industry-specific dividend yields, earning price ratios, book-to-market ratios, dividend payout ratios, net equity expansions, and stock variances. To the best of our knowledge, this is the first study that investigates industry portfolio predictability using such a detailed list of predictors.[32] Our focus is on model combinations based on the linear models we introduced in subsection 3.1,

$$r^j_{\tau+1} \quad = \quad \mu^j + \beta^j x^j_\tau + \varepsilon^j_{\tau+1}, \quad \tau = 1, ..., t-1, \tag{23}$$

[31] As for the case of a larger discount factor, note that when $\lambda = 1$ equation (8) implies that the CER-based 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.

[32] We use quarterly COMPUSTAT and monthly CRSP data, along with industry portfolio returns and industry classifications from Kenneth French's Data Library, available at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/datalibrary.html.

where $r^j_{\tau+1}$ is time $\tau+1$ monthly excess return for the j-th industry (the industries we consider are Consumers, Manufacturing, High-Tech, Health, and Other), x^j_τ is one of the industry predictors, and $\varepsilon^j_{\tau+1} \sim N(0, \sigma^{2,j}_\varepsilon)$.[33] Table 3 reports the results of this experiment, with the CERD of the individual models (relative to industry-specific PM benchmarks) presented on the top panel, and the model combinations in the bottom panel. There, we also include the results for the CER-based linear pool we introduced in subsection 5.2. Overall, the results we find for the various industry portfolios are largely consistent with those reported in Table 2 for the case of the S&P500. In particular, we find that compared to the average CERD from the individual models, model combinations generates larger economic gains. However, only in a few instances the model combinations manage to improve over the best individual models entering the combinations. On the other hand, the CER-based DeCo scheme appears to consistently improve over the best individual models entering the combination, with CERD values that are on average 172 basis points higher than the alternative model combination methods. The only exception is the Health industry, where the CER-based DeCo fails to improve over the CERD of the individual model based on the log dividend-yield. However, even in this case the CER-based DeCo scheme manages to deliver a CERD value that is 60 basis points higher than the next best model combination.

8 Conclusions

In this paper we extend the density combination approach of Billio et al. (2013) to feature combination weights that depend on the individual models' past profitability. We apply our model combination scheme to forecast stock returns, both at the aggregate level and by industry, and find improvements in 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. We also apply our combination scheme 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 even larger improvements in predictability. These gains appears to be robust to a large number of robustness checks.

[33] We provide further details on the industry definitions and classifications in the notes to Table 3.

References

Aastveit, K., F. Ravazzolo, and H. van Dijk (2014). Nowcasting the business cycle in an uncertain enviroment. Norges Bank.

Amisano, G. and R. Giacomini (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics 25*(2), 177–190.

Andrews, D. W. K. and J. C. Monahan (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica 60*(4), pp. 953–966.

Avramov, D. (2002). Stock return predictability and model uncertainty. *Journal of Financial Economics 64*(3), 423 – 458.

Banbura, M., D. Giannone, and L. Reichlin (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics 25*(1), 71–92.

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.

Boudoukh, J., M. Richardson, and R. F. Whitelaw (2008). The myth of long-horizon predictability. *Review of Financial Studies 21*(4), 1577–1605.

Brown, S. J. (1976). Optimal portfolio choice under uncertainty: a Bayesian approach. Ph.D. Dissertation, University of Chicago.

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. (2011). Real-time density forecasts from bayesian vector autoregressions with stochastic volatility. *Journal of Business & Economic Statistics 29*(3), 327–341.

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

Cochrane, J. H. (2008). The dog that did not bark: A defense of return predictability. *Review of Financial Studies 21*(4), 1533–1575.

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 (2014). Dynamic prediction pools: an investigation of financial frictions and forecasting performance. Penn Institute for Economic Research Working Paper 14-034.

Del Negro, M. and G. E. Primiceri (2014). Time-varying structural vector autoregressions and monetary policy: A corrigendum. Working Paper, Northwestern University.

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.

Elliott, G. and A. Timmermann (2005). Optimal forecast combination under regime switching. *International Economic Review 46*(4), 1081–1102.

Frost, P. A. and J. E. Savarino (1986). An empirical Bayes approach to efficient portfolio selection. *The Journal of Financial and Quantitative Analysis 21*(3), pp. 293–305.

Geweke, J. (2010). *Complete and Incomplete Econometric Models*. Princeton: Priceton 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). Forecasting and decision theory. In C. G. G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 1, 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 Stastistics 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.

Hoogerheide, L., R. Kleijn, R. Ravazzolo, H. K. van Dijk, and M. Verbeek (2010). Forecast Accuracy and Economic Gains from Bayesian Model Averaging using Time Varying Weights. *Journal of Forecasting 29*(1-2), 251–269.

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

Johannes, M., A. Korteweg, and N. Polson (2014). 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.

Kan, R. and G. Zhou (2007). Optimal portfolio choice with parameter uncertainty. *The Journal of Financial and Quantitative Analysis 42*(3), pp. 621–656.

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.

Kapetanios, G., J. Mitchell, S. Price, and N. Fawcett (2015). Generalised density forecast combinations. *Journal of Econometrics 188*(1), 150 – 165.

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.

Koop, G. and D. Korobilis (2011). UK macroeconomic forecasting with many predictors: Which models forecast best and when do they do so? *Economic Modelling 28*(5), 2307 – 2318.

Koop, G. and D. Korobilis (2012). Forecasting inflation using dynamic model averaging. *International Economic Review 53*(3), 867–886.

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. Van Nieuwerburgh (2008). Reconciling the return predictability evidence. *Review of Financial Studies 21*(4), 1607–1652.

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. (2012). The economic value of estimated porfolio rules under general utility specifications. Terry College of Business, University of Georgia Working Paper.

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 (2014). Forecasting stock returns under economic constraints. *Journal of Financial Economics 114*(3), 517–553.

Poirier, D. J. (1995). *Intermediate Statistics and Econometrics: A Comparative Approach.* MIT Press.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies 72*(3), 821–852.

Raftery, A., M. Karny, and P. Ettler (2010). Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. *Technometrics 52*, 52–66.

Rapach, D. E., J. K. Strauss, J. Tu, and G. Zhou (2015). Industry interdependencies and cross-industry return predictability. Working paper, Olin School of Business at Washington University.

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.

Sentana, E. (2005). Least squares predictions and mean-variance analysis. *Journal of Financial Econometrics 3*(1), 56–78.

Stambaugh, R. F. (1997). Analyzing investments whose histories differ in length. *Journal of Financial Economics 45*(3), 285 – 331.

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.

Ter Horst, J. R., F. De Roon, and B. J. Werker (2006). Incorporating estimation risk in portfolio choice. In L. Reneboog (Ed.), *Advances in Corporate Finance and Asset Prices*. Elsevier, Amsterdam.

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.

Tu, J. and G. Zhou (2011). Markowitz meets talmud: A combination of sophisticated and naive diversification strategies. *Journal of Financial Economics 99*(1), 204 – 215.

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.

Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g prior distributions. In P. Goel and A. Zellner (Eds.), *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, pp. 233–243. North-Holland.

Table 1. Out-of-sample forecast performance

Predictor	Panel A: OoS R^2		Panel B: CRPSD		Panel C: LSD	
	Linear	TVP-SV	Linear	TVP-SV	Linear	TVP-SV
Individual models						
Log dividend yield	-0.44 %	**1.03 %** *	-0.37 %	**8.06 %** ***	-0.15 %	**10.68 %** ***
Log earning price ratio	-2.27 %	**0.06 %**	-0.79 %	**8.01 %** ***	-0.17 %	**10.94 %** ***
Log smooth earning price ratio	-1.51 %	**0.60 %**	-0.59 %	**8.33 %** ***	**0.02 %**	**11.42 %** ***
Log dividend-payout ratio	-1.91 %	-1.77 %	-0.45 %	**6.46 %** ***	-0.19 %	**9.01 %** ***
Book-to-market ratio	-1.79 %	-0.29 %	-0.61 %	**8.14 %** ***	-0.12 %	**11.48 %** ***
T-Bill rate	-0.12 %	**0.38 %**	-0.07 %	**6.90 %** ***	-0.10 %	**8.91 %** ***
Long-term yield	-0.95 %	-0.92 %	-0.38 %	**6.24 %** ***	-0.22 %	**8.42 %** ***
Long-term return	-1.55 %	-0.68 %	-0.46 %	**6.45 %** ***	-0.14 %	**8.71 %** ***
Term spread	**0.09 %**	**0.12 %**	**0.08 %**	**6.77 %** ***	-0.03 %	**8.75 %** ***
Default yield spread	-0.24 %	-0.22 %	-0.07 %	**7.04 %** ***	-0.08 %	**8.03 %** ***
Default return spread	-0.23 %	-0.48 %	-0.11 %	**6.87 %** ***	-0.03 %	**9.07 %** ***
Stock variance	**0.09 %**	-0.89 %	**0.02 %**	**8.37 %** ***	-0.02 %	**11.81 %** ***
Net equity expansion	-0.93 %	-0.84 %	**0.00 %**	**7.10 %** ***	**0.04 %**	**9.59 %** ***
Inflation	-0.19 %	-0.15 %	-0.05 %	**7.50 %** ***	-0.15 %	**9.96 %** ***
Log total net payout yield	-0.79 %	**0.18 %**	-0.33 %	**7.05 %** ***	**0.06 %**	**9.54 %** ***
Model combinations						
Equal weighted combination	**0.49 %**	**0.52 %** *	**0.08 %**	**7.66 %** ***	-0.11 %	**10.05 %** ***
BMA	**0.39 %**	**0.50 %** *	**0.10 %**	**7.67 %** ***	**0.03 %**	**10.18 %** ***
Optimal prediction pool	-1.93 %	-0.77 %	-0.43 %	**8.37 %** ***	-0.11 %	**11.80 %** ***
DeCo	**0.43 %**	**1.37 %** ***	**0.07 %**	**8.41 %** ***	**0.00 %**	**10.91 %** ***
CER-based DeCo	**2.32 %** ***	**2.14 %** ***	**0.73 %** ***	**9.14 %** ***	**0.26 %** ***	**11.72 %** ***

This table reports the out-of-sample R^2 ("OoS R^2"), the average cumulative rank probability score differentials ("CRPSD"), and the average log predictive score differentials ("LSD") for the combination schemes and the individual prediction models of monthly excess returns. The out-of-sample R^2 are measured relative to the prevailing mean (PM) model as: $R^2_{m,OoS} = 1 - \left[\sum_{\tau=\underline{t}+1}^{\overline{t}} e^2_{m,\tau} / \sum_{\tau=\underline{t}+1}^{\overline{t}} e^2_{PM,\tau}\right]$ where m denotes either an individual model or a model combination, $\tau \in \{\underline{t}+1, ..., \overline{t}\}$, and $e_{m,\tau}$ ($e_{PM,\tau}$) stands for model m's (PM's) prediction error from the forecasts made at time τ,obtained by synthesizing the predictive density into a point forecast. The average CRPS differentials are expressed as percentage point differences relative to the PM model as $CRPSD_m = \sum_{\tau=\underline{t}+1}^{\overline{t}} \left(CRPS_{PM,\tau} - CRPS_{m,\tau}\right) / \sum_{\tau=\underline{t}+1}^{\overline{t}} CRPS_{PM,\tau}$, where $CRPS_{m,\tau}$ ($CRPS_{PM,\tau}$) denotes model m's (PM's) CRPS from the density forecasts made at time τ. The average log predictive score differentials are expressed as percentage point differences relative to the PM model as: $LSD_m = \sum_{\tau=\underline{t}+1}^{\overline{t}} \left(LS_{m,\tau} - LS_{PM,\tau}\right) / \sum_{\tau=\underline{t}+1}^{\overline{t}} LS_{PM,\tau}$, where $LS_{m,\tau}$ ($LS_{PM,\tau}$) denotes model m's (PM's) log predictive score from the density forecasts made at time τ. 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_τ: $r_{\tau+1} = \mu + \beta x_\tau + \varepsilon_{\tau+1}$, 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_τ: $r_{\tau+1} = (\mu + \mu_{\tau+1}) + (\beta + \beta_{\tau+1}) x_\tau + \exp(h_{\tau+1}) u_{\tau+1}$, and combination of these N TVP-SV individual models. The model "CER-based DeCo" refers to the case with $A = 5$ and $\lambda = 0.95$. We measure statistical significance relative to the prevailing mean model using the Diebold and Mariano (1995) t-tests for equality of the average loss. 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 the forecast accuracy measures are greater than zero. All results are based on the whole forecast evaluation period, January 1947 - December 2010.

Table 2. Economic performance of portfolios based on out-of-sample return forecasts

Individual models				
Predictor	Panel A: vs. PM		Panel B: vs. PM-SV	
	Linear	TVP-SV	Linear	TVP-SV
Log dividend yield	-0.33 %	**0.90 %**	-1.12 %	**0.11 %**
Log earning price ratio	**0.25 %**	**1.11 %**	-0.54 %	**0.32 %**
Log smooth earning price ratio	-0.38 %	**0.91 %**	-1.17 %	**0.12 %**
Log dividend-payout ratio	**0.41 %**	**0.94 %**	-0.38 %	**0.15 %**
Book-to-market ratio	-0.58 %	**0.61 %**	-1.36 %	-0.18 %
T-Bill rate	-0.26 %	**0.87 %**	-1.05 %	**0.08 %**
Long-term yield	-0.34 %	**0.52 %**	-1.13 %	-0.27 %
Long-term return	-0.42 %	**0.74 %**	-1.21 %	-0.05 %
Term spread	**0.15 %**	**0.81 %**	-0.64 %	**0.02 %**
Default yield spread	-0.20 %	**0.86 %**	-0.99 %	**0.07 %**
Default return spread	-0.14 %	**0.62 %**	-0.93 %	-0.17 %
Stock variance	**0.00 %**	**0.97 %**	-0.79 %	**0.18 %**
Net equity expansion	-0.14 %	**0.79 %**	-0.92 %	**0.00 %**
Inflation	-0.17 %	**0.79 %**	-0.96 %	**0.00 %**
Log total net payout yield	-0.37 %	**0.46 %**	-1.16 %	-0.33 %
Model combinations				
Equal weighted combination	**0.02 %**	**1.06 %**	-0.77 %	**0.27 %**
BMA	-0.05 %	**1.03 %**	-0.84 %	**0.24 %**
Optimal prediction pool	-0.82 %	**0.96 %**	-1.61 %	**0.17 %**
CER-based linear pool	-0.03 %	**1.13 %**	-0.82 %	**0.34 %**
DeCo	-0.01 %	**1.78 %**	-0.80 %	**0.99 %**
CER-based DeCo	**0.94 %**	**2.46 %**	**0.15 %**	**1.68 %**

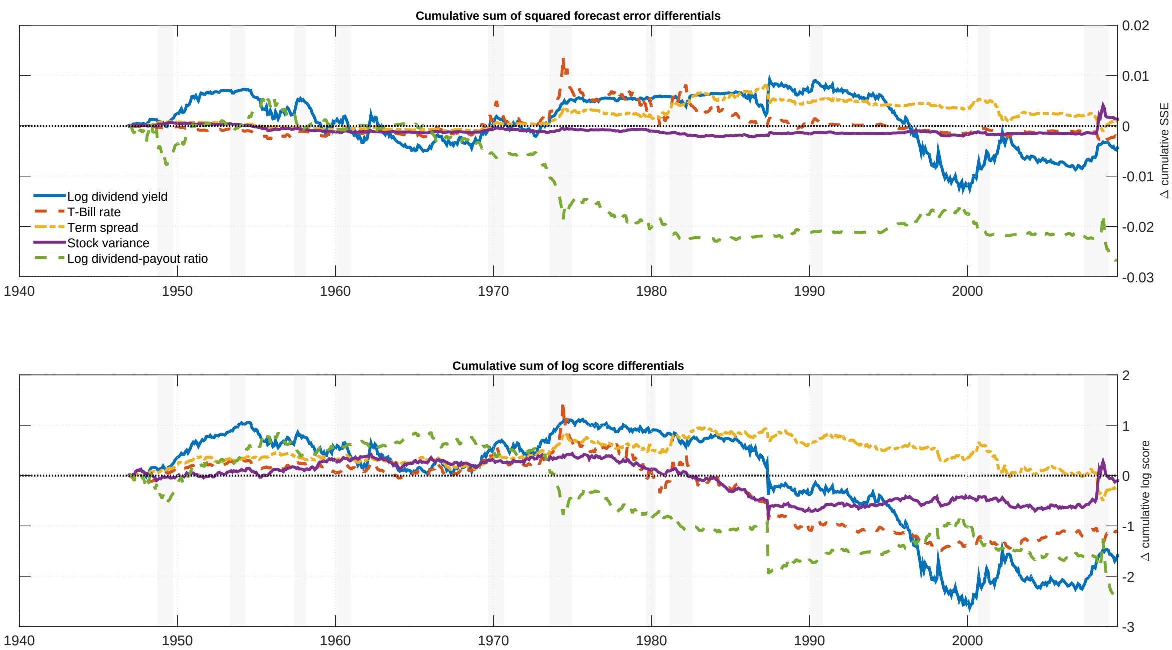
This table reports the annualized certainty equivalent return differentials (CERD) for portfolio decisions based on recursive out-of-sample forecasts of excess returns. Each period an investor with power utility and coefficient of relative risk aversion $A = 5$ selects stocks and T-bills based on a different predictive density, based either on a combination scheme or on an individual prediction model of the monthly excess returns. The columns "Linear" refers to predictive return distributions based on a linear regression of monthly excess returns on an intercept and a lagged predictor variable, x_τ: $r_{\tau+1} = \mu + \beta x_\tau + \varepsilon_{\tau+1}$, 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_τ: $r_{\tau+1} = (\mu + \mu_{\tau+1}) + (\beta{+}\beta_{\tau+1})\, x_\tau + \exp(h_{\tau+1})\, u_{\tau+1}$, and combination of these N TVP-SV individual models. The models "CER-based linear pool" and "CER-based DeCo" refer to the case with $A = 5$ and, in the case of "CER-based DeCo", $\lambda = 0.95$. Panel A reports CERD that are measured relative to the prevailing mean (PM) benchmark, while panel B presents CERD that are computed relative to the prevailing mean model with stochastic volatility (PM-SV) benchmark. Bold figures indicate all instances in which the CERD is greater than zero. All results are based on the whole forecast evaluation period, January 1947 - December 2010.

Table 3. Economic performance of industry-sorted portfolio returns

	Individual models				
Predictor	Cnsmr	Manuf	HiTec	Hlth	Other
Log dividend yield	-0.82 %	**0.05 %**	-0.53 %	**0.97 %**	-0.67 %
Log earning price ratio	-0.49 %	-0.81 %	**0.25 %**	**0.30 %**	**0.83 %**
Log dividend-payout ratio	-1.01 %	-1.08 %	-1.37 %	-0.42 %	-2.45 %
Book-to-market ratio	0.00 %	**0.60 %**	**0.52 %**	**0.26 %**	-0.69 %
T-Bill rate	-0.45 %	-0.41 %	-0.25 %	-1.05 %	-1.09 %
Long-term yield	-0.45 %	-0.54 %	**0.25 %**	-0.55 %	-0.43 %
Long-term return	**1.36 %**	**0.41 %**	**0.11 %**	**0.25 %**	**1.17 %**
Term spread	**0.57 %**	-0.07 %	**0.27 %**	-0.67 %	-0.30 %
Default yield spread	**0.12 %**	-2.22 %	-0.64 %	-0.44 %	-1.92 %
Default return spread	-0.40 %	-0.54 %	-0.40 %	**0.06 %**	-0.90 %
Stock variance	-0.73 %	**0.11 %**	**0.52 %**	-0.36 %	**0.12 %**
Inflation	-0.53 %	-0.53 %	-0.31 %	-0.46 %	-1.32 %
Net equity expansion	-1.26 %	-0.52 %	-0.34 %	-1.31 %	-0.14 %
	Model combinations				
Equal weighted combination	**0.37 %**	**0.50 %**	**1.18 %**	-0.34 %	**0.42 %**
BMA	**0.33 %**	**0.33 %**	**1.27 %**	-0.41 %	**0.39 %**
Optimal prediction pool	-1.43 %	-0.19 %	**1.11 %**	-1.69 %	-0.13 %
CER-based linear pool	**0.51 %**	**0.51 %**	**1.24 %**	-0.22 %	**0.42 %**
DeCo	**1.00 %**	**0.95 %**	**2.19 %**	**0.02 %**	**1.00 %**
CER-based DeCo	**2.30 %**	**1.70 %**	**3.41 %**	**0.62 %**	**2.44 %**

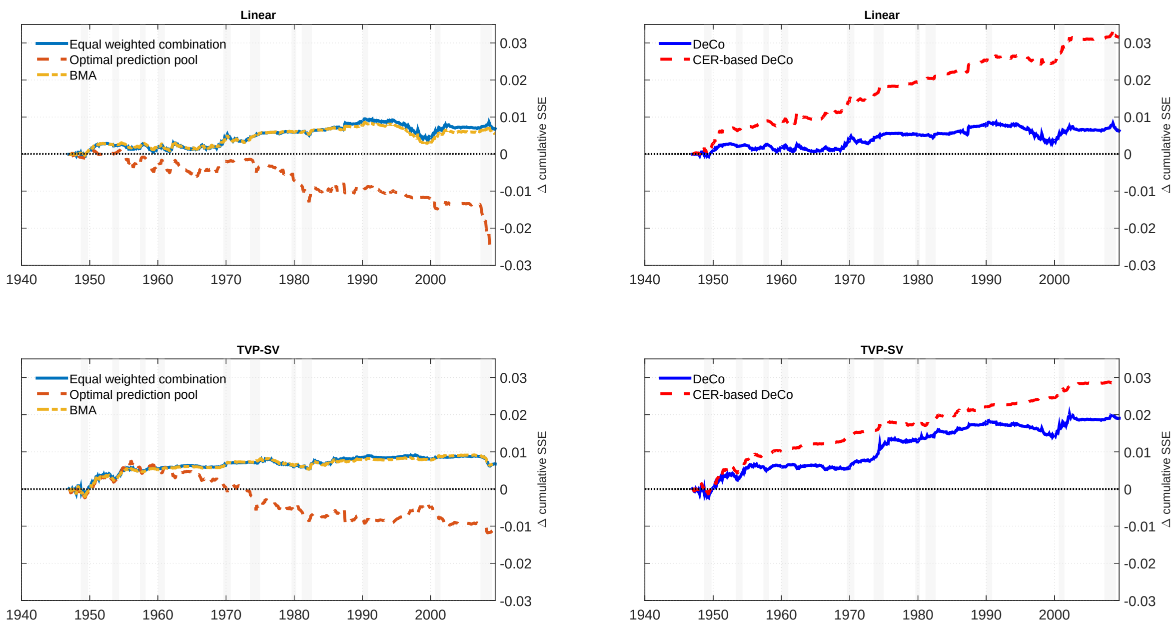
This table reports the annualized certainty equivalent return differentials (CERD) for portfolio decisions based on recursive out-of-sample forecasts of industry excess returns. Each period an investor with power utility and coefficient of relative risk aversion $A = 5$ selects stocks from a given industry and T-bills based on a different predictive density, based either on a combination scheme or on an individual prediction model of the monthly excess returns from that industry. The classification of stocks into industries is based on the industry definitions provided by Kenneth French's and available at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/datalibrary.html. Specifically: (1) *Cnsmr* includes Consumer Durables, NonDurables, Wholesale, Retail, and Some Services (Laundries, Repair Shops); *Manuf* includes Manufacturing, Energy, and Utilities; *HiTec* includes Business Equipment, Telephone and Television Transmission, as well as a number of other high-tech services; *Hlth* includes Healthcare, Medical Equipment, and Drugs; *Other* includes Mines, Constr, BldMt, Trans, Hotels, Bus Serv, Entertainment, and Finance. For each industry, predictive densities are obtained from a linear regression of monthly excess returns for that industry on an intercept and a lagged predictor variable, x^j_τ (j goes from 1 to 5): $r^j_{\tau+1} = \mu^j + \beta^j x^j_\tau + \varepsilon^j_{\tau+1}$, and combination of these N linear individual models. The models "CER-based linear pool" and "CER-based DeCo" refer to the case with $A = 5$ and, in the case of "CER-based DeCo", $\lambda = 0.95$. CERD are measured relative to the prevailing mean (PM) benchmark, and bold figures indicate all instances in which the CERD is greater than zero. All results are based on the whole forecast evaluation period, January 1980 - December 2010.

Figure 1. **Out-of-sample forecast performance of univariate models**



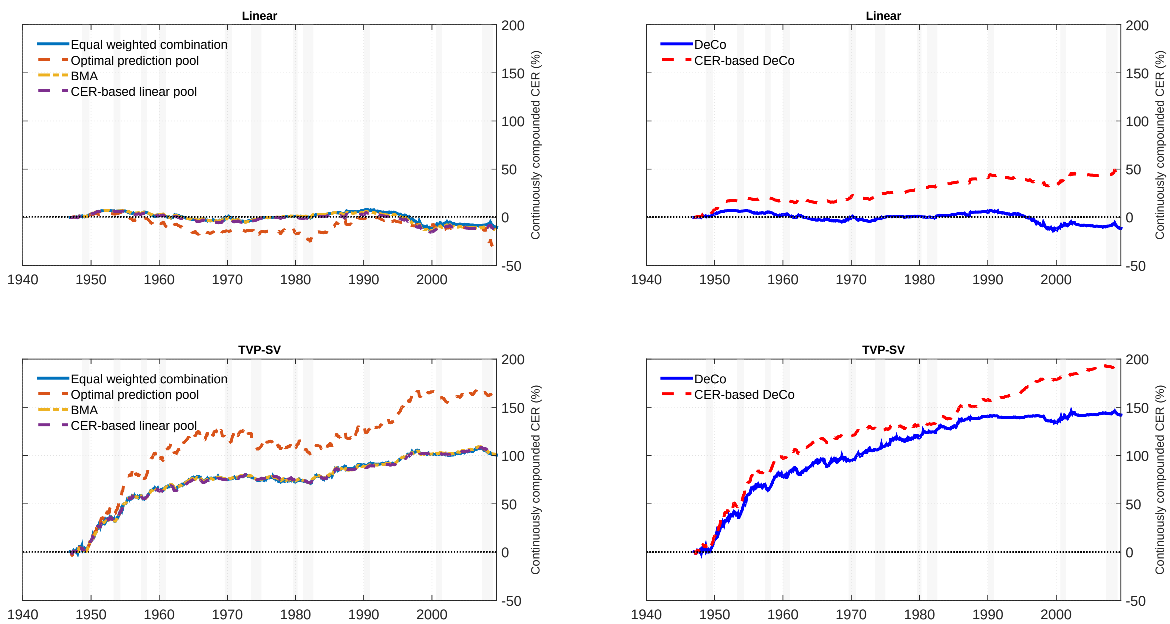
The top panel shows the sum of squared forecast errors of the prevailing mean model (PM) model minus the sum of squared forecast errors for each of five different univariate models. Each model is estimated from a linear regression of monthly excess returns on an intercept and a lagged predictor variable, x_τ: $r_{\tau+1} = \mu + \beta x_\tau + \varepsilon_{\tau+1}$. We plot the cumulative sum of squared forecast errors of the PM forecasts relative to the individual model m, $CSSED_{m,t} = \sum_{\tau=\underline{t}+1}^{t}\left(e^2_{PM,\tau} - e^2_{m,\tau}\right)$. Values above zero indicate that a given predictor generates better performance than the PM benchmark, while negative values suggest the opposite. The bottom panel 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 same five univariate models relative to the cumulative sum of log-predictive scores of the PM model, $CLSD_{m,t} = \sum_{\tau=\underline{t}+1}^{t}\left(LS_{m,\tau} - LS_{PM,\tau}\right)$. Values above zero indicate that a model combination method generates better performance than the PM benchmark, while negative values suggest the opposite. We show results based on the forecasts generated using the log dividend-yield (blue solid line), the T-bill rate (red dashed line), the Term Spread (yellow dotted line), the Stock variance (purple solid line), and the log dividend-payout ratio (green dashed line). Shaded areas indicate NBER-dated recessions.

Figure 2. **Cumulative sum of squared forecast error differentials for model combinations**



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 (top panels) and univariate time-varying parameter stochastic volatility (TVP-SV) models (bottom panels). For each model combination, we plot the cumulative sum of squared forecast errors of the PM forecasts relative to the model combination forecasts, $CSSED_{m,t} = \sum_{\tau=\underline{t}+1}^{t}\left(e_{PM,\tau}^{2} - e_{m,\tau}^{2}\right)$. Values above zero indicate that a model combination generates better performance than the PM benchmark, while negative values suggest the opposite. The left panels present results for the equal weighted combination (blue solid line), the optimal prediction pool (red dashed line), and Bayesian model averaging (yellow dashed line). The right-hand side panels plots the the cumulative sum of squared forecast error differentials for the Density Combination schemes, with and without learning. The model "CER-based DeCo" refers to the case with $A = 5$ and $\lambda = 0.95$. Shaded areas indicate NBER-dated recessions.

Figure 3. **Economic value of out-of-sample forecasts**



This figure plots the cumulative certainty equivalent returns of six alternative model combination methods based on linear models (top panels) and univariate time-varying parameter stochastic volatility (TVP-SV) models (bottom panels), 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, while the weight on stocks is constrained to lie in the interval $[0, 0.99]$. The left panels present results for the equal weighted combination (blue solid line), the optimal prediction pool (red dashed line), Bayesian model averaging (yellow dashed line), and a CER-based linear pool where the combination weights depends on the past profitability of the univariate models (purple dashed line). The right-hand side panels plots the the cumulative sum of squared forecast error differentials for the Density Combination methods, with and without learning. The models "CER-based linear pool" and "CER-based DeCo" refer to the case with $A = 5$ and, in the case of "CER-based DeCo", $\lambda = 0.95$. Shaded areas indicate NBER-dated recessions.

Online Appendix

This Appendix is organized as follows. A description of the prior specifications and the posterior simulation algorithms employed to estimate both the linear and the time-varying parameter with stochastic volatility models in the paper is provided in section A. Next, section B sketches the Sequential Monte Carlo algorithm used to obtain the predictive density for the CER-based density combination scheme, along with a description of the priors employed. Section C reports the results of several robustness checks to the main results presented in sections 5 and 6 of the paper. Finally, section D provides a number of supplementary tables and charts, including results for a shorter evaluation sample ending in 2007 before the onset of the latest recession, and a graphical summary of the time dynamics of the CER-based DeCo combination weights.

A Prior and posterior simulations

A.1 Linear models

The individual linear models regress stock returns, measured in excess of a risk-free rate, $r_{\tau+1}$, on a constant and a lagged predictor variable, x_τ:

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

A.1.1 Priors

Following standard practice, the priors for the parameters μ and β in (1) are assumed to be normal and independent of σ_ε^2,[34]

$$\begin{bmatrix} \mu \\ \beta \end{bmatrix} \sim \mathcal{N}\left(\underline{\mathbf{b}}, \underline{\mathbf{V}}\right), \tag{A-2}$$

with the hyperparameters $\underline{\mathbf{b}}$ and $\underline{\mathbf{V}}$ calibrated over the initial twenty years of data, January 1927 to December 1946.[35] In particular, we set all the elements of $\underline{\mathbf{b}}$ to zero, except for the term corresponding to μ, which is set to $\bar{r}_{\underline{t}}$, the average excess return calculated over the initial training sample. As for the elements of $\underline{\mathbf{V}}$, we use a g-prior (see Zellner (1986))

$$\underline{\mathbf{V}} = \underline{\psi}^2 \left[s_{r,\underline{t}}^2 \left(\sum_{\tau=1}^{\underline{t}-1} x_\tau x_\tau' \right)^{-1} \right], \tag{A-3}$$

where $s_{r,\underline{t}}^2$ denotes the standard deviation of excess returns, calculated over the initial training sample, and $\underline{t} = 240$. Note that our choice of the prior mean vector $\underline{\mathbf{b}}$ 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. In (A-3), $\underline{\psi}$ is a constant that controls the tightness of the prior, with $\underline{\psi} \to \infty$ corresponding to a diffuse prior on μ and β. Our benchmark analysis sets $\underline{\psi} = 1$.

We assume a standard gamma prior for the error precision of the return innovation, σ_ε^{-2}:

$$\sigma_\varepsilon^{-2} \sim \mathcal{G}\left(s_{r,\underline{t}}^{-2}, \underline{v}_0 \left(\underline{t} - 1\right) \right), \tag{A-4}$$

[34] See for example Koop (2003), Section 4.2.

[35] The approach of calibrating some of the prior hyperparameters using statistics computed over an initial training sample is quite standard in the Bayesian literature; see, e.g., Primiceri (2005), Clark (2011), Clark and Ravazzolo (2015), and Banbura et al. (2010).

where $\underline{v}_0$ is a prior hyperparameter that controls the degree of informativeness of this prior, with $\underline{v}_0 \to 0$ corresponding to a diffuse prior on σ_ε^{-2}. Our baseline analysis sets $\underline{v}_0 = 1$.[36]

A.1.2 Posterior simulation

For the linear models the goal is to obtain draws from the joint posterior distribution $p\left(\mu, \beta, \sigma_\varepsilon^{-2} \middle| M_i, \mathcal{D}^t\right)$, where $\mathcal{D}^t$ denotes all information available up to time t, and M_i denotes model i, with $i = 1, .., N$. Combining the priors in (A-2)-(A-4) with the likelihood function yields the following conditional posteriors:

$$\left.\left[\begin{array}{c}\mu \\ \beta\end{array}\right]\right| \sigma_\varepsilon^{-2}, M_i, \mathcal{D}^t \sim \mathcal{N}\left(\overline{\mathbf{b}}, \overline{\mathbf{V}}\right), \tag{A-5}$$

and

$$\sigma_\varepsilon^{-2}\Big| \mu, \beta, M_i, \mathcal{D}^t \sim \mathcal{G}\left(\overline{s}^{-2}, \overline{v}\right), \tag{A-6}$$

where

$$\begin{aligned} \overline{\mathbf{V}} &= \left[\underline{\mathbf{V}}^{-1} + \sigma_\varepsilon^{-2} \sum_{\tau=1}^{t-1} x_\tau x_\tau^{'}\right]^{-1}, \\ \overline{\mathbf{b}} &= \overline{\mathbf{V}}\left[\underline{\mathbf{V}}^{-1}\underline{\mathbf{b}} + \sigma_\varepsilon^{-2} \sum_{\tau=1}^{t-1} x_\tau r_{\tau+1}\right], \\ \overline{v} &= \underline{v}_0\left(\underline{t} - 1\right) + (t-1). \end{aligned} \tag{A-7}$$

and

$$\overline{s}^2 = \frac{\sum_{\tau=1}^{t-1}\left(r_{\tau+1} - \mu - \beta x_\tau\right)^2 + \left(s_{r,\underline{t}}^2 \times \underline{v}_0\left(\underline{t} - 1\right)\right)}{\overline{v}}. \tag{A-8}$$

A Gibbs sampler algorithm can be used to iterate back and forth between (A-5) and (A-6), yielding a series of draws for the parameter vector $\left(\mu, \boldsymbol{\beta}, \sigma_\varepsilon^{-2}\right)$. Draws from the predictive density $p\left(r_{t+1}\middle| M_i, \mathcal{D}^t\right)$ can then be obtained by noting that

$$p\left(r_{t+1}\middle| M_i, \mathcal{D}^t\right) = \int p\left(r_{t+1}\middle| \mu, \beta, \sigma_\varepsilon^{-2}, M_i, \mathcal{D}^t\right) p\left(\mu, \beta, \sigma_\varepsilon^{-2}\middle| M_i, \mathcal{D}^t\right) d\mu d\beta d\sigma_\varepsilon^{-2}. \tag{A-9}$$

A.2 Time-varying Parameter, Stochastic Volatility Models

The time-varying parameter, stochastic volatility (TVP-SV) model allows both the regression coefficients and the return volatility to change over time:

$$r_{\tau+1} = \left(\mu + \mu_{\tau+1}\right) + \left(\beta + \beta_{\tau+1}\right) x_\tau + \exp\left(h_{\tau+1}\right) u_{\tau+1}, \quad \tau = 1, ..., t-1, \tag{A-10}$$

[36]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 $Var(Y) = 2\mu^2/v$.

where $h_{\tau+1}$ denotes the (log of) stock return volatility at time $\tau+1$, and $u_{\tau+1} \sim \mathcal{N}(0,1)$. We assume that the time-varying parameters $\boldsymbol{\theta}_{\tau+1} = (\mu_{\tau+1}, \beta_{\tau+1})'$ follow a zero-mean, stationary process

$$\boldsymbol{\theta}_{\tau+1} = \boldsymbol{\gamma}_{\boldsymbol{\theta}}' \boldsymbol{\theta}_{\tau} + \boldsymbol{\eta}_{\tau+1}, \quad \boldsymbol{\eta}_{\tau+1} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{Q}\right), \tag{A-11}$$

where $\boldsymbol{\theta}_1 = \mathbf{0}$ and the elements in $\boldsymbol{\gamma}_{\boldsymbol{\theta}}$ are restricted to lie between -1 and 1.[37] The log-volatility $h_{\tau+1}$ is also assumed to follow a stationary and mean reverting process:

$$h_{\tau+1} = \lambda_0 + \lambda_1 h_{\tau} + \xi_{\tau+1}, \quad \xi_{\tau+1} \sim \mathcal{N}\left(0, \sigma_{\xi}^2\right), \tag{A-12}$$

where $|\lambda_1| < 1$ and u_{τ}, $\boldsymbol{\eta}_t$ and ξ_s are mutually independent for all τ, t, and s.

A.2.1 Priors

Our choice of priors for (μ, β) are the same as those in (A-2). The TVP-SV model in (19)-(21) also requires eliciting priors for the sequence of time-varying parameters, $\boldsymbol{\theta}^t = \{\boldsymbol{\theta}_2, ..., \boldsymbol{\theta}_t\}$ the variance covariance matrix $\boldsymbol{Q}$, the sequence of log return volatilities, $h^t = \{h_1, ..., h_t\}$, the error precision σ_{ξ}^{-2}, and the parameters $\boldsymbol{\gamma}_{\boldsymbol{\theta}}$, λ_0, and λ_1. Using the decomposition $p\left(\boldsymbol{\theta}^t, \boldsymbol{\gamma}_{\boldsymbol{\theta}}, \mathbf{Q}\right) = p\left(\boldsymbol{\theta}^t \middle| \boldsymbol{\gamma}_{\boldsymbol{\theta}}, \mathbf{Q}\right) p\left(\boldsymbol{\gamma}_{\boldsymbol{\theta}}\right) p\left(\mathbf{Q}\right)$, we note that (20) along with the assumption that $\boldsymbol{\theta}_1 = \mathbf{0}$ implies

$$p\left(\boldsymbol{\theta}^t \middle| \boldsymbol{\gamma}_{\boldsymbol{\theta}}, \mathbf{Q}\right) = \prod_{\tau=1}^{t-1} p\left(\boldsymbol{\theta}_{\tau+1} \middle| \boldsymbol{\gamma}_{\boldsymbol{\theta}}, \boldsymbol{\theta}_t, \mathbf{Q}\right), \tag{A-13}$$

with $\boldsymbol{\theta}_{\tau+1} | \boldsymbol{\gamma}_{\boldsymbol{\theta}}, \boldsymbol{\theta}_{\tau}, \mathbf{Q} \sim \mathcal{N}\left(\boldsymbol{\gamma}_{\boldsymbol{\theta}}' \boldsymbol{\theta}_{\tau}, \mathbf{Q}\right)$, for $\tau = 1, ..., t-1$. To complete the prior elicitation for $p\left(\boldsymbol{\theta}^t, \boldsymbol{\gamma}_{\boldsymbol{\theta}}, \mathbf{Q}\right)$, we specify priors for $\mathbf{Q}$ and $\boldsymbol{\gamma}_{\boldsymbol{\theta}}$ as follows. As for $\boldsymbol{Q}$, we choose an Inverted Wishart distribution

$$\boldsymbol{Q} \sim \mathcal{IW}\left(\underline{\boldsymbol{Q}}, \underline{t} - 2\right), \tag{A-14}$$

with

$$\underline{\boldsymbol{Q}} = \underline{k}_Q \left(\underline{t} - 2\right) \left[s_{r,\underline{t}}^2 \left(\sum_{\tau=1}^{\underline{t}-1} x_{\tau} x_{\tau}' \right)^{-1} \right]. \tag{A-15}$$

The constant $\underline{k}_Q$ controls the degree of variation in the time-varying regression coefficients $\boldsymbol{\theta}_{\tau}$, where larger values of $\underline{k}_Q$ imply greater variation in $\boldsymbol{\theta}_{\tau}$.[38] We set $\underline{k}_Q = 0.01$ to limit the extent

[37]Note that this is equivalent to writing $r_{\tau+1} = \widetilde{\mu}_{\tau+1} + \widetilde{\beta}_{\tau+1} x_{\tau} + \exp\left(h_{\tau+1}\right) u_{\tau+1}$, where $\left(\widetilde{\mu}_1, \widetilde{\beta}_1\right)$ is left unrestricted.

[38]In this way, the scale of the Wishart distribution for $\boldsymbol{Q}$ is specified to be a fraction of the OLS estimates of the variance covariance matrix $s_{r,\underline{t}}^2 \left(\sum_{\tau=1}^{\underline{t}-1} x_{\tau} x_{\tau}'\right)^{-1}$, multiplied by the degrees of freedom, $\underline{t} - 2$, 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., Primiceri (2005).

to which the parameters can change over time. We specify the elements of $\boldsymbol{\gamma_\theta}$ to be a priori independent of each other with generic element $\gamma_{\boldsymbol{\theta}}^i$

$$\gamma_{\boldsymbol{\theta}}^i \sim \mathcal{N}\left(\underline{m}_{\gamma_{\boldsymbol{\theta}}}, \underline{V}_{\gamma_{\boldsymbol{\theta}}}\right), \quad \gamma_{\boldsymbol{\theta}}^i \in (-1,1), \quad i = 1,2 \tag{A-16}$$

where $\underline{m}_{\gamma_{\boldsymbol{\theta}}} = 0.95$, and $\underline{V}_{\gamma_{\boldsymbol{\theta}}} = 1.0e^{-6}$, implying high autocorrelations.

Next, consider the sequence of log-volatilities, h^t, the error precision, σ_ξ^{-2}, and the parameters λ_0 and λ_1. Decomposing the joint probability of these parameters $p\left(h^t, \lambda_0, \lambda_1, \sigma_\xi^{-2}\right) = p\left(h^t \middle| \lambda_0, \lambda_1, \sigma_\xi^{-2}\right) p\left(\lambda_0, \lambda_1\right) p\left(\sigma_\xi^{-2}\right)$ and using (21), we have

$$\begin{aligned} p\left(h^t \middle| \lambda_0, \lambda_1, \sigma_\xi^{-2}\right) &= \prod_{\tau=1}^{t-1} p\left(h_{\tau+1} \middle| \lambda_0, \lambda_1, h_\tau, \sigma_\xi^{-2}\right) p\left(h_1\right), \\ h_{\tau+1} \middle| \lambda_0, \lambda_1, h_\tau, \sigma_\xi^{-2} &\sim \mathcal{N}\left(\lambda_0 + \lambda_1 h_\tau, \sigma_\xi^2\right). \end{aligned} \tag{A-17}$$

To complete the prior elicitation for $p\left(h^t, \lambda_0, \lambda_1 \sigma_\xi^{-2}\right)$, we choose priors for λ_0, λ_1, the initial log volatility h_1, and σ_ξ^{-2} from the normal-gamma family:

$$h_1 \sim \mathcal{N}\left(\ln\left(s_{r,\underline{t}}\right), \underline{k}_h\right), \tag{A-18}$$

$$\begin{bmatrix} \lambda_0 \\ \lambda_1 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \underline{m}_{\lambda_0} \\ \underline{m}_{\lambda_1} \end{bmatrix}, \begin{bmatrix} \underline{V}_{\lambda_0} & 0 \\ 0 & \underline{V}_{\lambda_1} \end{bmatrix}\right), \quad \lambda_1 \in (-1,1), \tag{A-19}$$

and

$$\sigma_\xi^{-2} \sim \mathcal{G}\left(1/\underline{k}_\xi, 1\right). \tag{A-20}$$

We set $\underline{k}_\xi = 0.01$ and choose the remaining hyperparameters in (A-18) and (A-19) 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_τ, we set $\underline{m}_{\lambda_1} = 0.95$, and $\underline{V}_{\lambda_1} = 1.0e^{-06}$, which imply a high autocorrelation in $h_{\tau+1}$.

A.2.2 Posterior simulation

Let $s^t = \{s_1, s_2, ..., s_t\}$ be the history up to time t of the states for the mixture distribution used to approximate the χ^2 distribution under the Kim et al. (1998) algorithm. Also, to simplify the notation, let us group all the time invariant parameters of the TVP-SV model into the matrix $\boldsymbol{\Theta}$, where $\boldsymbol{\Theta} = \left(\mu, \beta, \boldsymbol{Q}, \boldsymbol{\gamma_\theta}, \sigma_\xi^{-2}, \lambda_0, \lambda_1\right)$.

To obtain draws from the joint posterior distribution $p\left(\boldsymbol{\Theta}, \boldsymbol{\theta}^t, h^t \middle| M_i', \mathcal{D}^t\right)$ under the TVP-SV model, we use the Gibbs sampler to draw recursively from the following eight conditional distributions:[39]

1. $p\left(\boldsymbol{\theta}^t \middle| \boldsymbol{\Theta}, h^t, M_i', \mathcal{D}^t\right)$.

2. $p\left(\mu, \beta \middle| \boldsymbol{\Theta}_{-\mu,\beta}, \boldsymbol{\theta}^t, h^t, M_i', \mathcal{D}^t\right)$.

3. $p\left(\boldsymbol{Q} \middle| \boldsymbol{\Theta}_{-\boldsymbol{Q}}, \boldsymbol{\theta}^t, h^t, M_i', \mathcal{D}^t\right)$

4. $p\left(s^t \middle| \boldsymbol{\Theta}, \boldsymbol{\theta}^t, h^t, M_i', \mathcal{D}^t\right)$.

5. $p\left(h^t \middle| \boldsymbol{\Theta}, \boldsymbol{\theta}^t, s^t, M_i', \mathcal{D}^t\right)$.

6. $p\left(\sigma_\xi^{-2} \middle| \boldsymbol{\Theta}_{-\sigma_\xi^{-2}}, \boldsymbol{\theta}^t, h^t, M_i', \mathcal{D}^t\right)$

7. $p\left(\boldsymbol{\gamma}_{\boldsymbol{\theta}} \middle| \boldsymbol{\Theta}_{-\boldsymbol{\gamma}_{\boldsymbol{\theta}}}, \boldsymbol{\theta}^t, h^t, M_i', \mathcal{D}^t\right)$

8. $p\left(\lambda_0, \lambda_1 \middle| \boldsymbol{\Theta}_{-\lambda_0,\lambda_1}, \boldsymbol{\theta}^t, h^t, M_i', \mathcal{D}^t\right)$

We simulate from each of these blocks as follows. Starting with $\boldsymbol{\theta}^t$, we focus on $p\left(\boldsymbol{\theta}^t \middle| \boldsymbol{\Theta}, h^t, M_i', \mathcal{D}^t\right)$. Define $\widetilde{r}_{\tau+1} = r_{\tau+1} - \mu - \beta x_\tau$ and rewrite (19) as follows:

$$\widetilde{r}_{\tau+1} = \mu_\tau - \beta_\tau x_\tau + \exp\left(h_{\tau+1}\right) u_{\tau+1} \tag{A-21}$$

Note that knowledge of μ and β makes $\widetilde{r}_{\tau+1}$ observable, and reduces (19) to the measurement equation of a standard linear Gaussian state space model with heteroskedastic errors. Thus the sequence of time varying parameters $\boldsymbol{\theta}^t$ can be drawn from (A-21) using, for example, the algorithm of Carter and Kohn (1994).

Moving on to $p\left(\mu, \beta \middle| \boldsymbol{\Theta}_{-\mu,\beta}, \boldsymbol{\theta}^t, h^t, M_i', \mathcal{D}^t\right)$, conditional on $\boldsymbol{\theta}^t$ it is straightforward to draw μ, β, by applying standard results. Specifically,

$$\left[\begin{array}{c} \mu \\ \beta \end{array}\right] \middle| \boldsymbol{\Theta}_{-\mu,\beta}, \boldsymbol{\theta}^t, h^t, M_i', \mathcal{D}^t \sim N\left(\overline{\mathbf{b}}, \overline{\mathbf{V}}\right), \tag{A-22}$$

where

$$\begin{aligned} \overline{\mathbf{V}} &= \left[\underline{\mathbf{V}}^{-1} + \sum_{\tau=1}^{t-1} \frac{1}{\exp\left(h_{\tau+1}\right)^2} x_\tau x_\tau'\right]^{-1}, \\ \overline{\mathbf{b}} &= \overline{\mathbf{V}}\left[\underline{\mathbf{V}}^{-1}\underline{\mathbf{b}} + \sum_{\tau=1}^{t-1} \frac{1}{\exp\left(h_{\tau+1}\right)^2} x_\tau \left(r_{\tau+1} - \mu_\tau - \beta_\tau x_\tau\right)\right], \end{aligned} \tag{A-23}$$

[39] Using standard set notation, we define A_{-b} as the complementary set of b in A, i.e. $A_{-b} = \{x \in A : x \neq b\}$.

As for $p\left(\boldsymbol{Q}|\,\boldsymbol{\Theta}_{-\boldsymbol{Q}},\boldsymbol{\theta}^t,h^t,M_i',\mathcal{D}^t\right)$, we have that

$$\boldsymbol{Q}|\,\boldsymbol{\Theta}_{-\boldsymbol{Q}},\boldsymbol{\theta}^t,h^t,M_i',\mathcal{D}^t \sim \mathcal{IW}\left(\overline{\boldsymbol{Q}},t+\underline{t}-3\right), \tag{A-24}$$

where

$$\overline{\boldsymbol{Q}} = \underline{\boldsymbol{Q}} + \sum_{\tau=1}^{t-1}\left(\boldsymbol{\theta}_{\tau+1}-\boldsymbol{\gamma}_{\boldsymbol{\theta}}'\boldsymbol{\theta}_\tau\right)\left(\boldsymbol{\theta}_{\tau+1}-\boldsymbol{\gamma}_{\boldsymbol{\theta}}'\boldsymbol{\theta}_\tau\right)'. \tag{A-25}$$

Moving on to the vector of states $p\left(s^t\middle|\,\boldsymbol{\Theta},\boldsymbol{\theta}^t,h^t,M_i',\mathcal{D}^t\right)$ and the time varying volatilities $p\left(h^t\middle|\,\boldsymbol{\Theta},\boldsymbol{\theta}^t,s^t,M_i',\mathcal{D}^t\right)$, we follow Primiceri (2005) and employ the algorithm of Kim et al. (1998).[40] 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 μ, β, and $\boldsymbol{\theta}^t$. Next, rewrite (19) as

$$r^*_{\tau+1} = \exp\left(h_{\tau+1}\right)u_{\tau+1}. \tag{A-26}$$

Squaring and taking logs on both sides of (A-26) yields a new state space system that replaces (19)-(21) with

$$r^{**}_{\tau+1} = 2h_{\tau+1} + u^{**}_{\tau+1}, \tag{A-27}$$

$$h_{\tau+1} = \lambda_0 + \lambda_1 h_\tau + \xi_{\tau+1}, \tag{A-28}$$

where $r^{**}_{\tau+1} = \ln\left[\left(r^*_{\tau+1}\right)^2\right]$, and $u^{**}_{\tau+1} = \ln\left(u^2_{\tau+1}\right)$, with u^{**}_τ independent of ξ_s for all τ and s. Since $u^{**}_{\tau+1} \sim \ln\left(\chi^2_1\right)$, 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 $s_{\tau+1}$, $\tau = 1,..,t-1$, turning their focus on drawing from $p\left(h^t\middle|\,\boldsymbol{\Theta},\boldsymbol{\theta}^t,s^t,M_i',\mathcal{D}^t\right)$ instead of $p\left(h^t\middle|\,\boldsymbol{\Theta},\boldsymbol{\theta}^t,M_i',\mathcal{D}^t\right)$. The introduction of the state variable $s_{\tau+1}$ allows us to rewrite the linear non-Gaussian state space representation in (A-27)-(A-28) as a linear Gaussian state space model, making use of the following approximation,

$$u^{**}_{\tau+1} \approx \sum_{j=1}^{7} q_j \mathcal{N}\left(m_j - 1.2704, v_j^2\right), \tag{A-29}$$

where m_j, v_j^2, and q_j, $j = 1,2,...,7$, are constants specified in Kim et al. (1998) and thus need not be estimated. In turn, (A-29) implies

$$u^{**}_{\tau+1}\middle|\, s_{\tau+1} = j \sim \mathcal{N}\left(m_j - 1.2704, v_j^2\right), \tag{A-30}$$

[40] However, we modify the algorithm of Primiceri (2005) to reflect the correction to the ordering of steps detailed in Del Negro and Primiceri (2014).

where each state has probability

$$\Pr\left(s_{\tau+1} = j\right) = q_j. \tag{A-31}$$

Draws for the sequence of states s^t can easily be obtained, noting that each of its elements can be independently drawn from the discrete density defined by

$$\Pr\left(s_{\tau+1} = j \middle| \boldsymbol{\Theta}, \boldsymbol{\theta}^t, h^t, M_i', \mathcal{D}^t\right) = \frac{q_j f_{\mathcal{N}}\left(r_{\tau+1}^{**} \middle| 2h_{\tau+1} + m_j - 1.2704, v_j^2\right)}{\sum_{l=1}^{7} q_l f_{\mathcal{N}}\left(r_{\tau+1}^{**} \middle| 2h_{\tau+1} + m_l - 1.2704, v_l^2\right)}. \tag{A-32}$$

for $\tau = 1, ..., t-1$ and $j = 1, ..., 7$, and where $f_{\mathcal{N}}$ denotes the kernel of a normal density. Next, conditional on s^t, we can rewrite the nonlinear state space system as follows:

$$\begin{aligned} r_{\tau+1}^{**} &= 2h_{\tau+1} + e_{\tau+1}, \\ h_{\tau+1} &= \lambda_0 + \lambda_1 h_\tau + \xi_{\tau+1}, \end{aligned} \tag{A-33}$$

where $e_{\tau+1} \sim N\left(m_j - 1.2704, v_j^2\right)$ with probability $\Pr\left(s_{\tau+1} = j \middle| \boldsymbol{\Theta}, \boldsymbol{\theta}^t, h^t, M_i', \mathcal{D}^t\right)$. 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.

Next, the posterior distribution for $p\left(\sigma_\xi^{-2} \middle| \mu, \beta, \boldsymbol{\theta}^t, \boldsymbol{Q}, h^t, \lambda_0, \lambda_1, \boldsymbol{\gamma}_{\boldsymbol{\theta}}, M_i', \mathcal{D}^t\right)$ is readily available as,

$$\sigma_\xi^{-2} \middle| \boldsymbol{\Theta}_{-\sigma_\xi^{-2}}, \boldsymbol{\theta}^t, h^t, M_i', \mathcal{D}^t \sim \mathcal{G}\left(\left[\frac{\underline{k}_\xi + \sum_{\tau=1}^{t-1} \left(h_{\tau+1} - \lambda_0 - \lambda_1 h_\tau\right)^2}{t}\right]^{-1}, t\right). \tag{A-34}$$

Finally, obtaining draws from $p\left(\boldsymbol{\gamma}_{\boldsymbol{\theta}} \middle| \boldsymbol{\Theta}_{-\boldsymbol{\gamma}_{\boldsymbol{\theta}}}, \boldsymbol{\theta}^t, h^t, M_i', \mathcal{D}^t\right)$ and $p\left(\lambda_0, \lambda_1 \middle| \boldsymbol{\Theta}_{-\lambda_0,\lambda_1}, \boldsymbol{\theta}^t, h^t, M_i', \mathcal{D}^t\right)$ is straightforward. As for $p\left(\boldsymbol{\gamma}_{\boldsymbol{\theta}} \middle| \boldsymbol{\Theta}_{-\boldsymbol{\gamma}_{\boldsymbol{\theta}}}, \boldsymbol{\theta}^t, h^t, M_i', \mathcal{D}^t\right)$, we separately draw each of its elements. The $i-$th element $\gamma_{\boldsymbol{\theta}}^i$ is drawn from the following distribution

$$\gamma_{\boldsymbol{\theta}}^i \middle| \boldsymbol{\Theta}_{-\boldsymbol{\gamma}_{\boldsymbol{\theta}}}, \boldsymbol{\theta}^t, h^t, \mathcal{D}^t \sim \mathcal{N}\left(\overline{m}_{\gamma_{\boldsymbol{\theta}}}^i, \overline{V}_{\gamma_{\boldsymbol{\theta}}}^i\right) \times \gamma_{\boldsymbol{\theta}}^i \in (-1, 1) \tag{A-35}$$

where $i = 1, 2$ and

$$\begin{aligned} \overline{V}_{\gamma_{\boldsymbol{\theta}}}^i &= \left[\underline{V}_{\gamma_{\boldsymbol{\theta}}}^{-1} + \mathbf{Q}^{ii} \sum_{\tau=1}^{t-1} \left(\theta_\tau^i\right)^2\right]^{-1}, \\ \overline{m}_{\gamma_{\boldsymbol{\theta}}}^i &= \overline{V}_{\gamma_{\boldsymbol{\theta}}}^i \left[\underline{V}_{\gamma_{\boldsymbol{\theta}}}^{-1} \underline{m}_{\gamma_{\boldsymbol{\theta}}} + \mathbf{Q}^{ii} \sum_{\tau=1}^{t-1} \theta_\tau^i \theta_{\tau+1}^i\right], \end{aligned} \tag{A-36}$$

and $\mathbf{Q}^{ii}$ is the $i-$th diagonal element of $\mathbf{Q}^{-1}$. As for $p\left(\lambda_0, \lambda_1 \middle| \boldsymbol{\Theta}_{-\lambda_0,\lambda_1}, \boldsymbol{\theta}^t, h^t, M_i', \mathcal{D}^t\right)$, we have that

$$\lambda_0, \lambda_1 | \boldsymbol{\Theta}_{-\lambda_0,\lambda_1}, \boldsymbol{\theta}^t, h^t, M_i', \mathcal{D}^t \sim \mathcal{N}\left(\left[\begin{array}{c} \overline{m}_{\lambda_0} \\ \overline{m}_{\lambda 1} \end{array}\right], \overline{V}_\lambda\right) \times \lambda_1 \in (-1,1)$$

where

$$\overline{V}_\lambda = \left\{\left[\begin{array}{cc} \underline{V}_{\lambda_0}^{-1} & 0 \\ 0 & \underline{V}_{\lambda_1}^{-1} \end{array}\right] + \sigma_\xi^{-2} \sum_{\tau=1}^{t-1} \left[\begin{array}{c} 1 \\ h_\tau \end{array}\right] [1, h_\tau]\right\}^{-1} \tag{A-37}$$

and

$$\left[\begin{array}{c} \overline{m}_{\lambda_0} \\ \overline{m}_{\lambda 1} \end{array}\right] = \overline{V}_\lambda \left\{\left[\begin{array}{cc} \underline{V}_{\lambda_0}^{-1} & 0 \\ 0 & \underline{V}_{\lambda_1}^{-1} \end{array}\right]\left[\begin{array}{c} \underline{m}_{\lambda_0} \\ \underline{m}_{\lambda_1} \end{array}\right] + \sigma_\xi^{-2} \sum_{\tau=1}^{t-1} \left[\begin{array}{c} 1 \\ h_\tau \end{array}\right] h_{\tau+1}\right\}. \tag{A-38}$$

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

$$\begin{aligned} p\left(r_{t+1} | M_i', \mathcal{D}^t\right) &= \int p\left(r_{t+1} | \boldsymbol{\theta}_{t+1}, h_{t+1}, \boldsymbol{\Theta}, \boldsymbol{\theta}^t, h^t, M_i', \mathcal{D}^t\right) \\ &\quad \times p\left(\boldsymbol{\theta}_{t+1}, h_{t+1} | \boldsymbol{\Theta}, \boldsymbol{\theta}^t, h^t, M_i', \mathcal{D}^t\right) \\ &\quad \times p\left(\boldsymbol{\Theta}, \boldsymbol{\theta}^t, h^t \middle| M_i', \mathcal{D}^t\right) d\boldsymbol{\Theta} d\boldsymbol{\theta}^{t+1} dh^{t+1}. \end{aligned} \tag{A-39}$$

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

1. Draws from $p\left(\boldsymbol{\Theta}, \boldsymbol{\theta}^t, h^t \middle| M_i', \mathcal{D}^t\right)$ are obtained from the Gibbs sampling algorithm described above;

2. Draws from $p\left(\boldsymbol{\theta}_{t+1}, h_{t+1} | \boldsymbol{\Theta}, \boldsymbol{\theta}^t, h^t, M_i', \mathcal{D}^t\right)$: having processed data up to time t, the next step is to simulate the future volatility, h_{t+1}, and the future parameters, $\boldsymbol{\theta}_{t+1}$. We have that

$$h_{t+1} | \boldsymbol{\Theta}, \boldsymbol{\theta}^t, h^t, M_i', \mathcal{D}^t \sim \mathcal{N}\left(\lambda_0 + \lambda_1 h_t, \sigma_\xi^2\right). \tag{A-40}$$

and

$$\boldsymbol{\theta}_{t+1} | \boldsymbol{\Theta}, \boldsymbol{\theta}^t, h^t, M_i', \mathcal{D}^t \sim \mathcal{N}\left(\boldsymbol{\gamma}_{\boldsymbol{\theta}}' \boldsymbol{\theta}_t, \boldsymbol{Q}\right). \tag{A-41}$$

3. Draws from $p\left(r_{t+1} | \boldsymbol{\theta}_{t+1}, h_{t+1}, \boldsymbol{\Theta}, \boldsymbol{\theta}^t, h^t, M_i', \mathcal{D}^t\right)$: we have that

$$r_{t+1} | \boldsymbol{\theta}_{t+1}, h_{t+1}, \boldsymbol{\Theta}, \boldsymbol{\theta}^t, h^t, M_i', \mathcal{D}^t \sim \mathcal{N}\left(\left(\mu + \mu_{t+1}\right) + \left(\beta + \beta_{t+1}\right) x_t, \exp\left(h_{t+1}\right)\right). \tag{A-42}$$

B Sequential combination

In this section, we summarize the prior elicitation and the posterior simulation for the 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.

B.1 Priors

First, we need to specify priors for σ_κ^{-2} and for the diagonal elements of $\mathbf{\Lambda}$. The prior for σ_κ^{-2}, the precision of our measure of incompleteness in the combination scheme, and the diagonal elements of $\mathbf{\Lambda}^{-1}$, the precision matrix of the process $\mathbf{z}_{t+1}$ governing the combination weights $\mathbf{w}_{t+1}$, are assumed to be gamma, $\mathcal{G}(\underline{s}_{\sigma_\kappa}^{-2}, \underline{v}_{\sigma_\kappa}(\underline{t}-1))$ and $\mathcal{G}(\underline{s}_{\mathbf{\Lambda}}^{-1}, \underline{v}_{\mathbf{\Lambda}}(\underline{t}-1))$, respectively. We set informative values on our prior beliefs regarding the incompleteness and the combination weights. Precisely, we set $\underline{v}_{\sigma_\kappa} = \underline{v}_{\mathbf{\Lambda}_i} = 1$ 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 $\underline{s}_{\mathbf{\Lambda}}^{-1} = 4$, allowing $\mathbf{z}_{t+1}$ to evolve freely over time and differ from the initial value $\mathbf{z}_0$, set to equal weights.[41]

B.2 Posterior simulation

Let $\boldsymbol{\varsigma}$ be the parameter vector of the combination model, that is $\boldsymbol{\varsigma} = (\sigma_\kappa^2, \mathbf{\Lambda})$. Assume that $\widetilde{\mathbf{r}}_\tau$, $\tau = 1, \ldots, t+1$ is computed using formulas from either the linear or TVP-SV models given in the previous section (recall that $\widetilde{\mathbf{r}}_\tau = (\widetilde{r}_{1,\tau}, ..., \widetilde{r}_{N,\tau})'$ is the $N \times 1$ vector of predictions made at time τ, and $p\left(\widetilde{\mathbf{r}}_\tau \middle| D^{\tau-1}\right)$ is its joint predictive density); define the vector of observable $\mathbf{r}_{1:t} = (r_1, \ldots, r_t)' \in D^t$, the augmented state vector $\mathbf{Z}_{t+1} = (\mathbf{w}_{t+1}, \mathbf{z}_{t+1}, \boldsymbol{\varsigma}_{t+1})$, where $\boldsymbol{\varsigma}_{t+1} = \boldsymbol{\varsigma}$, $\forall t$. We write the model combination in its state space form as

$$
\begin{aligned}
r_t &\sim p(r_t|\widetilde{\mathbf{r}}_t, \mathbf{Z}_t) && \text{(measurement density)} && \text{(B-1)}\\
\mathbf{Z}_t &\sim p(\mathbf{Z}_t|\mathbf{Z}_{t-1}, \mathbf{r}_{1:t}, \widetilde{\mathbf{r}}_t) && \text{(transition density)} && \text{(B-2)}\\
\mathbf{Z}_0 &\sim p(\mathbf{Z}_0) && \text{(initial density)} && \text{(B-3)}
\end{aligned}
$$

The state predictive and filtering densities, which provide the posterior densities of the combination weights, are

$$
\begin{aligned}
p(\mathbf{Z}_{t+1}|\mathbf{r}_{1:t}, \widetilde{\mathbf{r}}_{1:t}) &= \int p(\mathbf{Z}_{t+1}|\mathbf{Z}_t, \mathbf{r}_{1:t}, \widetilde{\mathbf{r}}_{1:t}) p(\mathbf{Z}_t|\mathbf{r}_{1:t}, \widetilde{\mathbf{r}}_{1:t}) d\mathbf{Z}_t && \text{(B-4)}\\
p(\mathbf{Z}_{t+1}|\mathbf{r}_{1:t+1}, \widetilde{\mathbf{r}}_{1:t+1}) &= \frac{p(r_{t+1}|\mathbf{Z}_{t+1}, \widetilde{\mathbf{r}}_{t+1}) p(\mathbf{Z}_{t+1}|\mathbf{r}_{1:t}, \widetilde{\mathbf{r}}_{1:t})}{p(r_{t+1}|\mathbf{r}_{1:t}, \widetilde{\mathbf{r}}_{1:t})} && \text{(B-5)}
\end{aligned}
$$

[41] 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 $\mathbf{\Lambda}$ allow the posterior weights on the individual models to differ substantially from equal weights.

and the marginal predictive density of the observable variables is then

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

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

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

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 $\widetilde{\mathbf{r}}_\tau$, $\tau = 1, \ldots, t+1$.

We initialize independently the M particle sets: $\Xi_0^j = \{\mathbf{Z}_0^{i,j}, \omega_0^{i,j}\}_{i=1}^N$, $j = 1, \ldots, M$. Each particle set Ξ_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 $\widetilde{\mathbf{r}}_1^j$, $j = 1, \ldots, M$, from $p(\widetilde{\mathbf{r}}_1|r_0)$ 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 $p(\widetilde{\mathbf{r}}_{t+1}|r_{1:t})$ 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 $\Xi_t^j = \{\mathbf{Z}_{1:t}^{i,j}, \omega_t^{i,j}\}_{i=1}^N$, $j = 1, \ldots, M$, is available at time t and that each particle set provides the approximation

$$p_{N,j}(\mathbf{z}_t|\mathbf{r}_{1:t}, \widetilde{\mathbf{r}}_{1:t}^j) = \sum_{i=1}^N \omega_t^{i,j} \delta_{\mathbf{z}_t^{i,j}}(\mathbf{z}_t) \tag{B-6}$$

of the filtering density, $p(\mathbf{Z}_t|\mathbf{y}_{1:t}, \widetilde{\mathbf{r}}_{1:t}^j)$, conditional on the j-th predictor realization, $\widetilde{\mathbf{r}}_{1:t}^j$. The prediction (including the weights $\mathbf{w}_{t+1}$) are computed using the state predictive $p(\mathbf{Z}_{t+1}|\mathbf{r}_{1:t}, \widetilde{\mathbf{r}}_{1:t})$. After collecting the results from the different particle sets, it is possible to obtain the following empirical predictive density for the stock returns

$$p_{M,N}(r_{t+1}|\mathbf{r}_{1:t}) = \frac{1}{MN} \sum_{j=1}^M \sum_{i=1}^N \omega_t^{i,j} \delta_{r_{t+1}^{i,j}}(r_{t+1}) \tag{B-7}$$

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.

C Robustness analysis

In this section we summarize the results of several robustness checks on the main results for the S&P500 index. First, we investigate the effect on the profitability analysis presented in sections 5.2 and 6 of altering the investor's relative risk aversion coefficient A. Next, we conduct a subsample analysis to shed light on the robustness of the results to the choice of the forecasting evaluation period. We next investigate the implications of altering the parameter λ controlling the degree of learning in the model combination weights. After that, 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.

C.1 Sensitivity to risk aversion

The economic predictability analysis we reported in sections 5.2 and 6 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 based on the prevaling mean (PM) benchmark are shown in Table C.2, while Table C.3 presents results based on the alternative PM benchmark with stochastic volatility, PM-SV.

Starting with Table C.2, we begin with the case $A = 2$, i.e., lower risk aversion compared to the baseline case. Under this scenario, the CER-based DeCo scheme generates CERDs 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' CERDs. 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 CERD 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 CERD 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 CERD of the linear models.

We next consider the case with $A = 10$. In this case we find an overall decrease in CERD values, both for the individual models and the model combinations. However, the CER-based 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 CERD for the CER-based DeCo combination scheme averaging across the TVP-SV models is still quite large, at 126 basis points.

Moving on to the PM-SV benchmark, a quick comparison between Table C.2 and Table C.3 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 CER-based DeCo results remain quite strong even after replacing the benchmark model, especially for the case of TVP-SV models. In particular, when $A = 2$ the CER-based DeCo CERD under the TVP-SV models is as high as 116 basis points, while when $A = 10$ it reaches 85 basis points.

C.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 C.4 show CERD 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 (2015) document a similar pattern in forecasting macroeconomic variables. Interestingly, the CER-based 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 C.4 show CERD 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 C.4 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 CER-based DeCo CERDs are still fairly large, as high as 87 basis points in the case of linear models, and as high as 167 basis points in the TVP-SV case.

C.3 Sensitivity to the learning dynamics

When specifying the learning mechanism for the CER-based DeCo in equations (7)-(9), we introduced the smoothing parameter λ, where $\lambda \in (0,1)$. Our main analysis of the economic value of equity premium forecasts in Sections 5.2 and 6 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 $\lambda = 0.9$.[42] Table C.5 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 λ by investigating the economic impact of such choice across different risk aversion coefficients ($A = 2$, 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 CER-based DeCo models based on $\lambda = 0.95$ generating, on average, slightly higher CERDs.

C.4 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

$$U\left(W_{i,\tau}\right) = E\left[W_{i,\tau}\middle|\,\mathcal{D}^{\tau-1}\right] - \frac{A}{2}Var\left[W_{i,\tau}\middle|\,\mathcal{D}^{\tau-1}\right] \tag{C-1}$$

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

$$W_{i,\tau} = \left(1-\omega_{i,\tau-1}\right)\exp\left(r^f_{\tau-1}\right) + \omega_{i,\tau-1}\exp\left(r^f_{\tau-1} + r_\tau\right) \tag{C-2}$$

[42] As for the case of a larger discount factor, note that when $\lambda = 1$ equation (8) implies that the CER-based 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.

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

$$\omega^*_{i,\tau-1} = \frac{\exp\left(\widehat{\mu}_{i,\tau} + \frac{\widehat{\sigma}^2_{i,\tau}}{2}\right) - 1}{A\exp\left(r^f_{\tau-1}\right)\exp\left(2\widehat{\mu}_{i,\tau} + \widehat{\sigma}^2_{i,\tau}\right)\left(\exp\left(\widehat{\sigma}^2_{i,\tau}\right) - 1\right)}. \tag{C-3}$$

where $\widehat{\mu}_{i,\tau}$ and $\widehat{\sigma}^2_{i,\tau}$ are shorthands for the mean and variance of $p\left(r_\tau | M_i, \mathcal{D}^{\tau-1}\right)$, the predictive density of r_τ 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 $M_1, ..., M_N$, but also on the overall statistical and economic predictability of the CER-based DeCo combination scheme. In fact, as we have discussed in subsection 3.2, the combination weight conditional density at time τ, $p(\mathbf{w}_\tau | \widetilde{\mathbf{r}}_\tau, \mathcal{D}^{\tau-1})$, depends on the history of profitability of the individual models M_1 to M_N through equations (7)–(9).

Note next that in the case of a mean variance investor, time τ CER is simply equal to the investor's realized utility $W^*_{i,\tau}$, hence equation (9) is replaced by

$$f\left(r_\tau, \widetilde{\mathbf{r}}_{i,\tau}\right) = U\left(W^*_{i,\tau}\right), \tag{C-4}$$

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

$$W_{i,\tau} = \left(1 - \omega^*_{i,\tau-1}\right)\exp\left(r^f_{\tau-1}\right) + \omega^*_{i,\tau-1}\exp\left(r^f_{\tau-1} + r_\tau\right). \tag{C-5}$$

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,

$$CER_m = 12 \times \frac{1}{t^*}\sum_{\tau=\underline{t}+1}^{\overline{t}} U\left(W^*_{m,\tau}\right) \tag{C-6}$$

where m denotes the model under consideration (either univariate or model combination), and $t^* = \overline{t} - \underline{t}$. Table C.6 presents differential certainty equivalent return estimates, relative to the benchmark prevailing mean model PM,

$$CERD_m = CER_m - CER_{PM} \tag{C-7}$$

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 Table 2 in the paper and Table C.6 reveals that the economic gains for power utility and mean variance utility are quite similar in magnitude, and the overall takeaways from sections 5.2 and 6 remain unchanged. In particular, the CER-based DeCo combination scheme generates sizable CERDs, especially when combining TVP-SV models. For the benchmark case of $A = 5$, the CERD is as high as 220 basis points. Altering the risk aversion coefficients produces CERDs for the CER-based DeCo model ranging from 115 basis points ($A = 10$) to 436 basis points ($A = 2$).

C.5 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 CER-based DeCo combination scheme as the key prior hyperparameters change.

First, we investigate the impact of changing the prior hyperparameter $\underline{s}_{\mathbf{\Lambda}}^{-1}$ in (7) controlling the degree of time variation in the CER-based DeCo combination weights, which was set to $\underline{s}_{\mathbf{\Lambda}}^{-1} = 4$ in our baseline results. As sensitivities, we experiment with $\underline{s}_{\mathbf{\Lambda}}^{-1} = 0.2$ and $\underline{s}_{\mathbf{\Lambda}}^{-1} = 1000$, which imply more volatile combination weights (in the case of $\underline{s}_{\mathbf{\Lambda}}^{-1} = 0.2$), or smoother combination weights (in the case of $\underline{s}_{\mathbf{\Lambda}}^{-1} = 1000$). In the former case, the annualized CERD of the CER-based 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 CER-based DeCo. On the other hand, setting $\underline{s}_{\mathbf{\Lambda}}^{-1} = 1000$ yields a much larger reduction in the CER-based DeCo CERD, which decreases to 0.27%. It thus appears that too large a value for $\underline{s}_{\mathbf{\Lambda}}^{-1}$ produces combination weights that are far too smooth, affecting the economic performance of CER-based DeCo.[43]

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 μ and β, 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

[43] We also investigate the sensitivity of our baseline results to the choice of $\underline{s}_{\sigma_\kappa}^{-2}$, the prior hyperparameter controlling the degree of model incompleteness, and find that the performance of CER-based 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 $\underline{s}_{\sigma_\kappa}^{-2}$ is too large the estimation algorithm seems to converge very slowly.

the case of $\underline{\psi} = 10$) or more concentrated prior distributions (in the case of $\underline{\psi} = 0.01$) for μ and β. Similarly, the prior hyperparameter $\underline{v}_0$ controls the tightness of the prior on σ_ε^{-2}, and our baseline results are based on $\underline{v}_0 = 1$, which correspond to an hypothetical prior sample size of 20 years. As sensitivities, we experiment with $\underline{v}_0 = 0.1$ and $\underline{v}_0 = 100$, which imply, respectively, an hypothetical prior sample of two years (in the case of $\underline{v}_0 = 0.1$) or as large as 2,000 years (in the case of $\underline{v}_0 = 100$). Table C.7 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 2 in the paper 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 CER-based 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 CER-based DeCo combination scheme still performs quite adequately, with an annualized CERD of 48 basis points.

D Additional results

In this section, we present a number of supplementary tables and charts, including results for a shorter evaluation sample ending in 2007 before the onset of the latest recession, and a graphical summary of the time dynamics of the CER-based DeCo combination weights.

Table D.1 and Table D.2 are the analog of tables 1 and 2 in the paper for the shorter evaluation sample ending in December 2007, before the onset of the latest recession. Table D.1 presents the results on the statistical predictability of the individual models as well as the various model combination schemes, while Table D.2 reports their annualized CERD, relative to the prevailing mean benchmark.

Finally, Figure D.1 displays the posterior means of the CER-based DeCo combination weights for the top linear models (top panel) and TVP-SV models (bottom panel) over the whole evaluation period, January 1947 to December 2010.

Table C.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

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 C.2. 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 %	**1.10 %**	-0.16 %	**0.46 %**
Log earning price ratio	**0.12 %**	**1.51 %**	**0.13 %**	**0.59 %**
Log smooth earning price ratio	-1.36 %	**1.19 %**	-0.19 %	**0.47 %**
Log dividend-payout ratio	**0.99 %**	**1.07 %**	**0.21 %**	**0.46 %**
Book-to-market ratio	-1.37 %	**1.30 %**	-0.28 %	**0.31 %**
T-Bill rate	-0.66 %	**1.32 %**	-0.13 %	**0.43 %**
Long-term yield	-0.86 %	**0.74 %**	-0.17 %	**0.27 %**
Long-term return	-0.81 %	**0.86 %**	-0.19 %	**0.38 %**
Term spread	**0.47 %**	**1.68 %**	**0.06 %**	**0.42 %**
Default yield spread	-0.49 %	**1.10 %**	-0.10 %	**0.45 %**
Default return spread	-0.06 %	**1.15 %**	-0.09 %	**0.32 %**
Stock variance	**0.02 %**	**1.31 %**	**0.02 %**	**0.52 %**
Net equity expansion	**0.54 %**	**1.16 %**	-0.08 %	**0.41 %**
Inflation	-0.41 %	**0.88 %**	-0.07 %	**0.40 %**
Log total net payout yield	-1.07 %	**0.48 %**	-0.18 %	**0.23 %**
Model Combinations				
Equal weighted combination	**0.06 %**	**1.20 %**	**0.02 %**	**0.55 %**
BMA	-0.09 %	**1.28 %**	-0.02 %	**0.52 %**
Optimal prediction pool	-1.02 %	**1.28 %**	-0.41 %	**0.51 %**
CER-based linear pool	**0.04 %**	**1.44 %**	**0.01 %**	**0.56 %**
DeCo	0.00 %	**1.83 %**	**0.01 %**	**0.90 %**
CER-based DeCo	**2.63 %**	**2.33 %**	**0.50 %**	**1.26 %**

This table reports the certainty equivalent return differentials (CERD) 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 models "CER-based linear pool" and "CER-based DeCo" refer to the case with A matching the values in the headings ($A = 2$, 10) and, in the case of "CER-based DeCo", $\lambda = 0.95$. 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_τ: $r_{\tau+1} = \mu + \beta x_\tau + \varepsilon_{\tau+1}$, 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_τ: $r_{\tau+1} = (\mu + \mu_{\tau+1}) + (\beta+\beta_{\tau+1})\, x_\tau + \exp(h_{\tau+1})\, u_{\tau+1}$, and combination of these N time-varying parameter and stochastic volatility individual models. CERD 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 CERD is greater than zero. All results are based on the whole forecast evaluation period, January 1947 - December 2010.

Table C.3. Effect of risk aversion on economic performance measures and alternative benchmark

	A=2		A=10	
	Linear	TVP-SV	Linear	TVP-SV
Individual models				
Log dividend yield	-2.16 %	-0.07 %	-0.57 %	**0.05 %**
Log earning price ratio	-1.06 %	**0.34 %**	-0.29 %	**0.18 %**
Log smooth earning price ratio	-2.54 %	**0.02 %**	-0.60 %	**0.06 %**
Log dividend-payout ratio	-0.18 %	-0.10 %	-0.21 %	**0.05 %**
Book-to-market ratio	-2.55 %	**0.13 %**	-0.69 %	-0.10 %
T-Bill rate	-1.84 %	**0.15 %**	-0.54 %	**0.01 %**
Long-term yield	-2.03 %	-0.44 %	-0.58 %	-0.15 %
Long-term return	-1.99 %	-0.31 %	-0.61 %	-0.03 %
Term spread	-0.70 %	**0.50 %**	-0.35 %	**0.01 %**
Default yield spread	-1.66 %	-0.07 %	-0.51 %	**0.04 %**
Default return spread	-1.24 %	-0.03 %	-0.50 %	-0.10 %
Stock variance	-1.16 %	**0.14 %**	-0.40 %	**0.11 %**
Net equity expansion	-0.64 %	-0.02 %	-0.49 %	-0.01 %
Inflation	-1.58 %	-0.29 %	-0.48 %	-0.01 %
Log total net payout yield	-2.25 %	-0.69 %	-0.60 %	-0.18 %
Model Combinations				
Equal weighted combination	-1.12 %	**0.03 %**	-0.40 %	**0.14 %**
BMA	-1.27 %	**0.11 %**	-0.43 %	**0.11 %**
Optimal prediction pool	-2.20 %	**0.11 %**	-0.82 %	**0.10 %**
CER-based linear pool	-1.13 %	**0.27 %**	-0.40 %	**0.15 %**
DeCo	-1.18 %	**0.65 %**	-0.41 %	**0.49 %**
CER-based DeCo	**1.46 %**	**1.16 %**	**0.09 %**	**0.85 %**

This table reports the certainty equivalent return differentials (CERD) 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 models "CER-based linear pool" and "CER-based DeCo" refer to the case with A matching the values in the headings ($A = 2$, 10) and, in the case of "CER-based DeCo", $\lambda = 0.95$. 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_τ: $r_{\tau+1} = \mu + \beta x_\tau + \varepsilon_{\tau+1}$, 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_τ: $r_{\tau+1} = (\mu + \mu_{\tau+1}) + (\beta + \beta_{\tau+1})\, x_\tau + \exp(h_{\tau+1})\, u_{\tau+1}$, and combination of these N time-varying parameter and stochastic volatility individual models. CERD are annualized and are measured relative to the prevailing mean model with stochastic volatility which assumes a constant equity premium. Bold figures indicate all instances in which the CERD is greater than zero. All results are based on the whole forecast evaluation period, January 1947 - December 2010.

Table C.4. 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 %	**0.68 %**	**3.31 %**	**1.92 %**	**0.14 %**	**1.72 %**	-0.81 %	**0.07 %**
Log earning price ratio	**0.14 %**	**1.53 %**	**0.73 %**	-0.75 %	**0.58 %**	**1.39 %**	-0.10 %	**0.82 %**
Log smooth earning price ratio	-1.07 %	**1.21 %**	**2.80 %**	-0.45 %	-0.40 %	**1.50 %**	-0.37 %	**0.30 %**
Log dividend-payout ratio	**0.96 %**	**1.87 %**	-2.05 %	-3.16 %	**0.46 %**	**1.35 %**	**0.35 %**	**0.52 %**
Book-to-market ratio	-0.94 %	**1.29 %**	**1.11 %**	-2.44 %	-0.18 %	**1.21 %**	-0.99 %	-0.01 %
T-Bill rate	-0.52 %	**0.88 %**	**0.96 %**	**0.81 %**	**0.09 %**	**1.77 %**	-0.61 %	-0.05 %
Long-term yield	-0.64 %	**0.31 %**	**1.02 %**	**1.48 %**	-0.06 %	**1.45 %**	-0.63 %	-0.43 %
Long-term return	-0.32 %	**1.02 %**	-0.85 %	-0.51 %	-0.88 %	**0.96 %**	**0.06 %**	**0.52 %**
Term spread	-0.01 %	**1.03 %**	**0.87 %**	-0.18 %	**0.14 %**	**1.42 %**	**0.16 %**	**0.19 %**
Default yield spread	-0.27 %	**1.40 %**	**0.10 %**	-1.55 %	-0.24 %	**1.53 %**	-0.16 %	**0.18 %**
Default return spread	-0.12 %	**1.16 %**	-0.28 %	-1.80 %	-0.34 %	**0.55 %**	**0.06 %**	**0.69 %**
Stock variance	-0.13 %	**1.90 %**	**0.62 %**	-3.15 %	-0.14 %	**1.49 %**	**0.14 %**	**0.43 %**
Net equity expansion	**0.76 %**	**1.89 %**	-4.06 %	-4.03 %	**0.06 %**	**1.51 %**	-0.34 %	**0.06 %**
Inflation	-0.19 %	**1.30 %**	-0.11 %	-1.45 %	-0.16 %	**1.27 %**	-0.19 %	**0.30 %**
Log total net payout yield	-0.73 %	**0.65 %**	**1.28 %**	-0.40 %	**0.24 %**	**1.54 %**	-0.98 %	-0.64 %
Model Combinations								
Equal weighted combination	-0.18 %	**1.34 %**	**0.89 %**	-0.22 %	**0.13 %**	**1.66 %**	-0.10 %	**0.45 %**
BMA	-0.27 %	**1.32 %**	**0.96 %**	-0.27 %	**0.13 %**	**1.66 %**	-0.24 %	**0.39 %**
Optimal prediction pool	-0.39 %	**1.87 %**	-2.75 %	-3.02 %	-0.48 %	**1.48 %**	-1.17 %	**0.43 %**
CER-based linear pool	-0.25 %	**1.39 %**	**0.99 %**	-0.04 %	**0.09 %**	**1.71 %**	-0.15 %	**0.54 %**
DeCo	-0.25 %	**1.71 %**	**1.05 %**	**2.10 %**	**0.10 %**	**2.77 %**	-0.14 %	**0.78 %**
CER-based DeCo	**0.67 %**	**2.64 %**	**2.15 %**	**1.69 %**	**1.00 %**	**3.25 %**	**0.87 %**	**1.67 %**

This table reports the certainty equivalent return differentials (CERD) 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 models "CER-based linear pool" and "CER-based DeCo" refer to the case with $A = 5$ and, in the case of "CER-based DeCo", $\lambda = 0.95$. 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_τ: $r_{\tau+1} = \mu + \beta x_\tau + \varepsilon_{\tau+1}$, 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_τ: $r_{\tau+1} = (\mu + \mu_{\tau+1}) + (\beta + \beta_{\tau+1}) x_\tau + \exp(h_{\tau+1}) u_{\tau+1}$, and combination of these N time-varying parameter and stochastic volatility individual models. CERD 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 CERD is greater than zero.

Table C.5. Economic performance measures: alternative learning dynamics

	Full sample		NBER expansions		NBER recessions		1947-1978		1979-2010	
λ	Linear	TVP-SV	Linear	TVP-SV	Linear	TVP-SV	Linear	TVP-SV	Linear	TVP-SV
					A=2					
λ=0.9	**2.31 %**	**2.16 %**	**1.67 %**	**2.14 %**	**5.09 %**	**2.26 %**	**2.31 %**	**2.29 %**	**2.32 %**	**2.03 %**
λ=0.95	**2.63 %**	**2.33 %**	**1.89 %**	**2.28 %**	**5.86 %**	**2.56 %**	**2.66 %**	**2.35 %**	**2.60 %**	**2.31 %**
					A=5					
λ=0.9	**0.93 %**	**2.50 %**	**0.67 %**	**2.71 %**	**2.09 %**	**1.53 %**	**1.02 %**	**3.23 %**	**0.83 %**	**1.76 %**
λ=0.95	**0.94 %**	**2.46 %**	**0.67 %**	**2.64 %**	**2.15 %**	**1.69 %**	**1.00 %**	**3.25 %**	**0.87 %**	**1.67 %**
					A=10					
λ=0.9	**0.50 %**	**1.19 %**	**0.36 %**	**1.26 %**	**1.14 %**	**0.85 %**	**0.58 %**	**1.56 %**	**0.42 %**	**0.81 %**
λ=0.95	**0.50 %**	**1.26 %**	**0.37 %**	**1.30 %**	**1.13 %**	**1.07 %**	**0.57 %**	**1.63 %**	**0.43 %**	**0.88 %**

This table reports the certainty equivalent return differentials (CERD) for portfolio decisions based on recursive out-of-sample forecasts of monthly excess returns, under two alternative parametrizations for the learning dynamics, λ=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_τ: $r_{\tau+1} = \mu + \beta x_\tau + \varepsilon_{\tau+1}$, 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_τ: $r_{\tau+1} = (\mu + \mu_{\tau+1}) + (\beta+\beta_{\tau+1})\, x_\tau + \exp(h_{\tau+1})\, u_{\tau+1}$, and combination of these N time-varying parameter and stochastic volatility individual models. CERD 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 CERD 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 C.6. Economic performance measures: Mean Variance preferences

	A=2		A=5		A=10	
	Linear	TVP-SV	Linear	TVP-SV	Linear	TVP-SV
Individual models						
Log dividend yield	-0.80 %	**1.60 %**	-0.33 %	**0.80 %**	-0.16 %	**0.41 %**
Log earning price ratio	**0.31 %**	**2.24 %**	**0.20 %**	**1.07 %**	**0.10 %**	**0.54 %**
Log smooth earning price ratio	-0.99 %	**2.01 %**	-0.39 %	**0.84 %**	-0.20 %	**0.42 %**
Log dividend-payout ratio	**1.06 %**	**2.03 %**	**0.39 %**	**0.88 %**	**0.20 %**	**0.46 %**
Book-to-market ratio	-1.48 %	**1.85 %**	-0.57 %	**0.57 %**	-0.29 %	**0.30 %**
T-Bill rate	-0.59 %	**2.05 %**	-0.24 %	**0.82 %**	-0.13 %	**0.41 %**
Long-term yield	-0.81 %	**1.38 %**	-0.33 %	**0.53 %**	-0.16 %	**0.26 %**
Long-term return	-0.96 %	**1.54 %**	-0.38 %	**0.72 %**	-0.20 %	**0.35 %**
Term spread	**0.37 %**	**2.68 %**	**0.13 %**	**0.88 %**	**0.06 %**	**0.44 %**
Default yield spread	-0.46 %	**1.88 %**	-0.19 %	**0.84 %**	-0.09 %	**0.43 %**
Default return spread	-0.22 %	**1.71 %**	-0.15 %	**0.61 %**	-0.08 %	**0.30 %**
Stock variance	**0.01 %**	**2.06 %**	**0.01 %**	**1.00 %**	-0.01 %	**0.52 %**
Net equity expansion	**0.46 %**	**2.06 %**	-0.07 %	**0.82 %**	-0.04 %	**0.43 %**
Inflation	-0.37 %	**1.77 %**	-0.14 %	**0.78 %**	-0.09 %	**0.38 %**
Log total net payout yield	-0.91 %	**0.86 %**	-0.36 %	**0.40 %**	-0.18 %	**0.21 %**
Model Combinations						
Equal weighted combination	**0.03 %**	**2.35 %**	**0.01 %**	**1.00 %**	**0.00 %**	**0.51 %**
BMA	-0.11 %	**2.28 %**	-0.05 %	**0.99 %**	-0.03 %	**0.49 %**
Optimal prediction pool	-1.66 %	**2.03 %**	-0.74 %	**0.99 %**	-0.38 %	**0.51 %**
CER-based linear pool	**0.00 %**	**2.34 %**	-0.01 %	**1.01 %**	**0.00 %**	**0.50 %**
DeCo	-0.04 %	**3.52 %**	-0.02 %	**1.64 %**	-0.02 %	**0.84 %**
CER-based DeCo	**2.55 %**	**4.36 %**	**0.85 %**	**2.20 %**	**0.45 %**	**1.15 %**

This table reports the certainty equivalent return differentials (CERD) 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 models "CER-based linear pool" and "CER-based DeCo" refer to the case with A matching the values in the column headings ($A = 2$, 5, 10) and, in the case of 'CER-based DeCo", $\lambda = 0.95$. 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_τ: $r_{\tau+1} = \mu + \beta x_\tau + \varepsilon_{\tau+1}$, 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_τ: $r_{\tau+1} = (\mu + \mu_{\tau+1}) + (\beta + \beta_{\tau+1})\, x_\tau + \exp(h_{\tau+1})\, u_{\tau+1}$, and combination of these N time-varying parameter and stochastic volatility individual models. CERD 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 CERD is greater than zero. All results are based on the whole forecast evaluation period, January 1947 - December 2010.

Table C.7. Prior sensitivity analysis: economic performance

	$\psi = 10$, $\underline{v}_0 = 0.1$	$\psi = 0.01$, $\underline{v}_0 = 100$
Individual models		
Log dividend yield	-0.25 %	-0.19 %
Log earning price ratio	**0.27 %**	**0.08 %**
Log smooth earning price ratio	-0.29 %	-0.16 %
Log dividend-payout ratio	**0.30 %**	**0.06 %**
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	**0.33 %**	-0.23 %
Default yield spread	**0.00 %**	-0.11 %
Default return spread	**0.01 %**	-0.03 %
Stock variance	**0.27 %**	**0.00 %**
Net equity expansion	-0.02 %	-0.03 %
Inflation	-0.01 %	-0.12 %
Log total net payout yield	-0.23 %	-0.23 %
Model Combinations		
Equal weighted combination	**0.16 %**	-0.08 %
BMA	**0.17 %**	-0.06 %
Optimal prediction pool	-0.56 %	-0.07 %
CER-based linear pool	**0.18 %**	-0.04 %
DeCo	-0.06 %	**0.00 %**
CER-based DeCo	**0.74 %**	**0.48 %**

This table reports the certainty equivalent return differentials (CERD) 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. refer to the case with $A = 5$ and, in the case of "CER-based DeCo", $\lambda = 0.95$. Predictive return distributions are based on a linear regression of monthly excess returns on an intercept and a lagged predictor variable, x_τ: $r_{\tau+1} = \mu + \beta x_\tau + \varepsilon_{\tau+1}$, and combination of these N linear individual models for two different set of priors. The prior set with $\psi = 10$ and $\underline{v}_0 = 0.1$ refers to a diffuse prior assumption and $\psi = 10$ and $\underline{v}_0 = 0.1$ to an informative prior assumption. CERD 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 CERD is greater than zero. All results are based on the whole forecast evaluation period, January 1947 - December 2010.

Table D.1. Out-of-sample forecast performance, 1947-2007

Predictor	Panel A: OoS R^2		Panel B: CRPSD		Panel C: LSD	
	Linear	TVP-SV	Linear	TVP-SV	Linear	TVP-SV
Individual models						
Log dividend yield	-0.64 %	**0.95 %**	-0.42 %	**8.59 %** ***	-0.19 %	**11.17 %** ***
Log earning price ratio	-1.55 %	**0.37 %**	-0.46 %	**8.80 %** ***	-0.03 %	**11.65 %** ***
Log smooth earning price ratio	-1.98 %	**0.55 %**	-0.71 %	**8.87 %** ***	-0.02 %	**11.88 %** ***
Log dividend-payout ratio	-1.79 %	-1.74 %	-0.36 %	**7.01 %** ***	-0.14 %	**9.51 %** ***
Book-to-market ratio	-2.16 %	-0.35 %	-0.66 %	**8.74 %** ***	-0.13 %	**12.05 %** ***
T-Bill rate	-0.09 %	**0.55 %**	-0.09 %	**7.38 %** ***	-0.10 %	**9.31 %** ***
Long-term yield	-1.00 %	-1.07 %	-0.40 %	**6.66 %** ***	-0.22 %	**8.78 %** ***
Long-term return	-1.81 %	-0.78 %	-0.52 %	**6.89 %** ***	-0.16 %	**9.10 %** ***
Term spread	**0.17 %**	**0.28 %**	**0.08 %**	**7.25 %** ***	**0.00 %**	**9.16 %** ***
Default yield spread	-0.21 %	-0.08 %	-0.07 %	**7.57 %** ***	-0.08 %	**8.41 %** ***
Default return spread	-0.31 %	-0.49 %	-0.09 %	**7.37 %** ***	-0.05 %	**9.42 %** ***
Stock variance	-0.13 %	-0.80 %	-0.04 %	**8.90 %** ***	-0.05 %	**12.20 %** ***
Net equity expansion	**0.10 %**	-0.10 %	**0.28 %**	**7.80 %** ***	**0.18 %** *	**10.12 %** ***
Inflation	-0.16 %	**0.08 %**	-0.05 %	**8.06 %** ***	-0.14 %	**10.42 %** ***
Log total net payout yield	-0.68 %	**0.41 %**	-0.29 %	**7.64 %** ***	**0.08 %**	**10.05 %** ***
Model combinations						
Equal weighted combination	**0.59 %**	**0.71 %** **	**0.10 %**	**8.21 %** ***	-0.10 %	**10.50 %** ***
BMA	**0.51 %**	**0.73 %** **	**0.13 %**	**8.25 %** ***	**0.04 %**	**10.63 %** ***
Optimal prediction pool	-1.16 %	-0.77 %	-0.18 %	**8.92 %** ***	-0.01 %	**12.22 %** ***
DeCo	**0.54 %**	**1.53 %** ***	**0.10 %**	**9.00 %** ***	**0.02 %**	**11.40 %** ***
CER-based DeCo	**2.55 %** ***	**2.33 %** ***	**0.76 %** ***	**9.73 %** ***	**0.24 %** ***	**12.20 %** ***

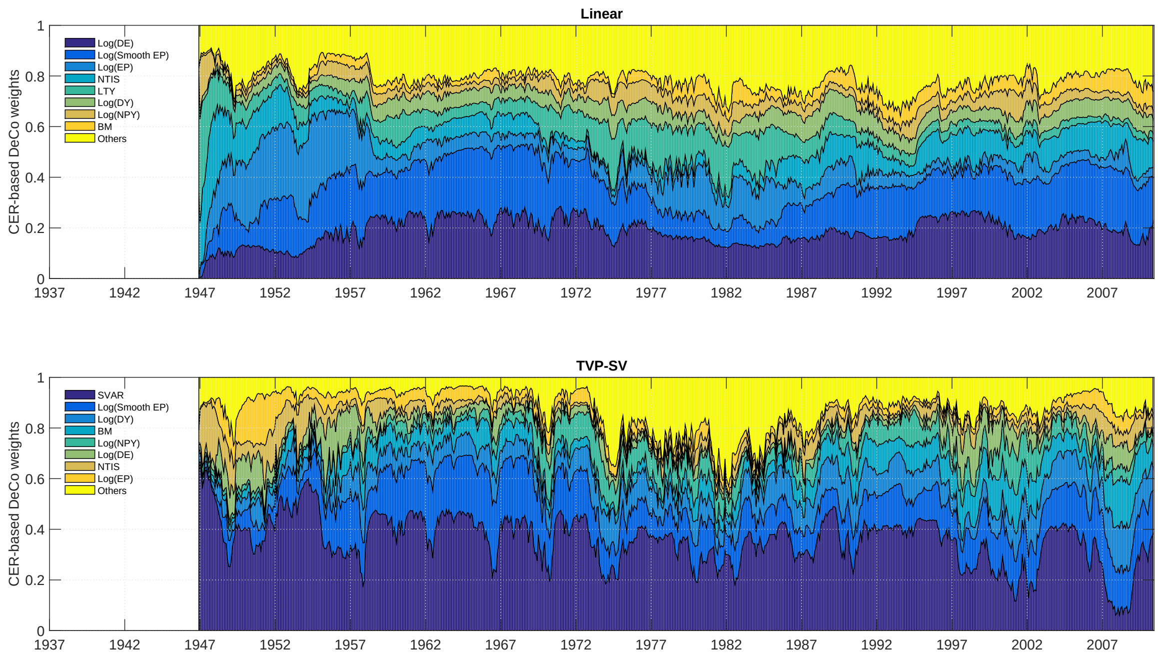
This table reports the out-of-sample R^2 ("OoS R^2"), the average cumulative rank probability score differentials ("CRPSD"), and the average log predictive score differentials ("LSD") for the combination schemes and the individual prediction models of monthly excess returns. The out-of-sample R^2 are measured relative to the prevailing mean (PM) model as: $R^2_{m,OoS} = 1 - \left[\sum_{\tau=\underline{t}+1}^{\bar{t}} e^2_{m,\tau} / \sum_{\tau=\underline{t}+1}^{\bar{t}} e^2_{PM,\tau}\right]$ where m denotes either an individual model or a model combination, $\tau \in \{\underline{t}+1, ..., \bar{t}\}$, and $e_{m,\tau}$ ($e_{PM,\tau}$) stands for model m's (PM's) prediction error from the forecasts made at time τ,obtained by synthesizing the predictive density into a point forecast. The average CRPS differentials are expressed as percentage point differences relative to the PM model as $CRPSD_m = \sum_{\tau=\underline{t}+1}^{\bar{t}} \left(CRPS_{PM,\tau} - CRPS_{m,\tau}\right) / \sum_{\tau=\underline{t}+1}^{\bar{t}} CRPS_{PM,\tau}$, where $CRPS_{m,\tau}$ ($CRPS_{PM,\tau}$) denotes model m's (PM's) CRPS from the density forecasts made at time τ. The average log predictive score differentials are expressed as percentage point differences relative to the PM model as: $LSD_m = \sum_{\tau=\underline{t}+1}^{\bar{t}} \left(LS_{m,\tau} - LS_{PM,\tau}\right) / \sum_{\tau=\underline{t}+1}^{\bar{t}} LS_{PM,\tau}$, where $LS_{m,\tau}$ ($LS_{PM,\tau}$) denotes model m's (PM's) log predictive score from the density forecasts made at time τ. 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_τ: $r_{\tau+1} = \mu + \beta x_\tau + \varepsilon_{\tau+1}$, 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_τ: $r_{\tau+1} = (\mu + \mu_{\tau+1}) + (\beta + \beta_{\tau+1}) x_\tau + \exp(h_{\tau+1}) u_{\tau+1}$, and combination of these N TVP-SV individual models. The model "CER-based DeCo" refers to the case with $A = 5$ and $\lambda = 0.95$. We measure statistical significance relative to the prevailing mean model using the Diebold and Mariano (1995) t-tests for equality of the average loss. 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 the forecast accuracy metrics are greater than zero. All results are based on an evaluation period that extends from January 1947 to December 2007.

Table D.2. Economic performance of portfolios based on out-of-sample return forecasts, 1947-2007

Individual models				
Predictor	Panel A: vs. PM		Panel B: vs. PM-SV	
	Linear	TVP-SV	Linear	TVP-SV
Log dividend yield	-0.42 %	**0.88 %**	-1.40 %	-0.09 %
Log earning price ratio	0.10 %	**1.12 %**	-0.87 %	**0.15 %**
Log smooth earning price ratio	-0.55 %	**0.92 %**	-1.53 %	-0.05 %
Log dividend-payout ratio	**0.49 %**	**1.11 %**	-0.48 %	**0.14 %**
Book-to-market ratio	-0.70 %	**0.70 %**	-1.67 %	-0.27 %
T-Bill rate	-0.24 %	**1.04 %**	-1.22 %	**0.07 %**
Long-term yield	-0.32 %	**0.59 %**	-1.30 %	-0.38 %
Long-term return	-0.47 %	**0.87 %**	-1.44 %	-0.10 %
Term spread	**0.20 %**	**1.03 %**	-0.77 %	**0.06 %**
Default yield spread	-0.18 %	**1.05 %**	-1.15 %	**0.08 %**
Default return spread	-0.10 %	**0.71 %**	-1.07 %	-0.26 %
Stock variance	-0.10 %	**1.05 %**	-1.07 %	**0.08 %**
Net equity expansion	**0.56 %**	**1.36 %**	-0.41 %	**0.38 %**
Inflation	-0.15 %	**1.00 %**	-1.12 %	**0.03 %**
Log total net payout yield	-0.26 %	**0.70 %**	-1.23 %	-0.27 %
Model Combinations				
Equal weighted combination	**0.03 %**	**1.23 %**	-0.94 %	**0.26 %**
BMA	-0.02 %	**1.23 %**	-0.99 %	**0.26 %**
Optimal prediction pool	-0.38 %	**1.04 %**	-1.36 %	**0.07 %**
CER-based linear pool	-0.02 %	**1.24 %**	-0.99 %	**0.27 %**
DeCo	**0.02 %**	**1.90 %**	-0.95 %	**0.93 %**
CER-based DeCo	**0.95 %**	**2.58 %**	-0.02 %	**1.61 %**

This table reports the annualized certainty equivalent return differentials (CERD) for portfolio decisions based on recursive out-of-sample forecasts of excess returns. Each period an investor with power utility and coefficient of relative risk aversion $A = 5$ selects stocks and T-bills based on a different predictive density, based either on a combination scheme or on an individual prediction model of the monthly excess returns. The columns "Linear" refers to predictive return distributions based on a linear regression of monthly excess returns on an intercept and a lagged predictor variable, x_τ: $r_{\tau+1} = \mu + \beta x_\tau + \varepsilon_{\tau+1}$, 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_τ: $r_{\tau+1} = (\mu + \mu_{\tau+1}) + (\beta+\beta_{\tau+1}) x_\tau + \exp(h_{\tau+1}) u_{\tau+1}$, and combination of these N TVP-SV individual models. The models "CER-based linear pool" and "CER-based DeCo" refer to the case with $A = 5$ and, in the case of "CER-based DeCo", $\lambda = 0.95$. Panel A reports CERD that are measured relative to the prevailing mean (PM) benchmark, while panel B presents CERD that are computed relative to the prevailing mean model with stochastic volatility (PM-SV) benchmark. Bold figures indicate all instances in which the CERD is greater than zero. All results are based on an evaluation period that extends from January 1947 to December 2007.

Figure D.1. Predictor weights for the CER-based DeCo combination scheme



This figure plots the posterior means of the CER-based DeCo weights for the top individual linear models (top panel) and TVP-SV models (bottom panel) over the out-of-sample period. The individual predictors showed are Log(DP): log dividend price ratio, Log(DY): log dividend yield, Log(EP): log earning price ratio, Log(Smooth EP): log smooth earning price ratio, Log(DE): log dividend-payout ratio, BM: book-to-market ratio, TBL: T-Bill rate, LTY: long-term yield, LTR: long-term return, TMS: term spread, DFY: default yield spread, DFR: default return spread, SVAR: stock variance, NTIS: net equity expansion, INFL: inflation, and Log(NPY): log total net payout yield. The out of sample period starts in January 1947 and ends in December 2010.

Centre for Applied Macro - and Petroleum economics (CAMP)
will bring together economists working on applied macroeconomic issues, with special emphasis on petroleum economics.

BI Norwegian Business School
Centre for Applied Macro - Petroleum economics (CAMP)
N-0442 Oslo

http://www.bi.no/camp

CAMP Working Paper Series
ISSN: 1892-2198