Online Appendix for "Optimal Portfolio Choice under Decision-Based Model Combinations"

Davide Pettenuzzo[*]
Brandeis University

Francesco Ravazzolo[†]
Norges Bank, and
BI Norwegian Business School

November 4, 2015

Roadmap

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. Next, 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.

[*]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.

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 (A-1) are assumed to be normal and independent of σ_ε^2,[1]

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

with the hyperparameters $\underline{\mathbf{b}}$ and $\underline{\mathbf{V}}$ calibrated over the initial twenty years of data, January 1927 to December 1946.[2] 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}:

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

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 $\underline{v}_0 = 1$.[3]

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

[2] 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).

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

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:

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

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

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

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:

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

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 $\boldsymbol{\theta}_{\tau+1}=\left(\mu_{\tau+1},\beta_{\tau+1}\right)'$ 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 $\boldsymbol{\gamma_\theta}$ are restricted to lie between -1 and 1.[4] 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 (A-10)-(A-12) 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 $\boldsymbol{\gamma_\theta}$, λ_0, and λ_1. Using the decomposition $p\left(\boldsymbol{\theta}^t, \boldsymbol{\gamma_\theta}, \mathbf{Q}\right) = p\left(\boldsymbol{\theta}^t \middle| \boldsymbol{\gamma_\theta}, \mathbf{Q}\right) p\left(\boldsymbol{\gamma_\theta}\right) p\left(\mathbf{Q}\right)$, we note that (A-11) along with the assumption that $\boldsymbol{\theta}_1 = \mathbf{0}$ implies

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

with $\boldsymbol{\theta}_{\tau+1} | \boldsymbol{\gamma_\theta}, \boldsymbol{\theta}_\tau, \mathbf{Q} \sim \mathcal{N}\left(\boldsymbol{\gamma}'_{\boldsymbol{\theta}} \boldsymbol{\theta}_\tau, \mathbf{Q}\right)$, for $\tau = 1, ..., t-1$. To complete the prior elicitation for $p\left(\boldsymbol{\theta}^t, \boldsymbol{\gamma_\theta}, \mathbf{Q}\right)$, we specify priors for $\mathbf{Q}$ and $\boldsymbol{\gamma_\theta}$ 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 $\boldsymbol{\theta}_\tau$.[5] We set $\underline{k}_Q = 0.01$ to limit the extent 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 $\gamma_{\boldsymbol{\theta}}^i$

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

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

[4] 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.

[5] 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).

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 (A-12), 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:[6]

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

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

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

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 (A-10) 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 (A-10) 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,

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

where

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

As for $p\left(\boldsymbol{Q} \middle| \boldsymbol{\Theta}_{-\boldsymbol{Q}}, \boldsymbol{\theta}^t, h^t, M_i', \mathcal{D}^t\right)$, 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).[7] 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 (A-10) as

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

Squaring and taking logs on both sides of (A-26) yields a new state space system that replaces (A-10)-(A-12) with

$$\begin{aligned} r^{**}_{\tau+1} &= 2h_{\tau+1} + u^{**}_{\tau+1}, & \text{(A-27)} \\ h_{\tau+1} &= \lambda_0 + \lambda_1 h_\tau + \xi_{\tau+1}, & \text{(A-28)} \end{aligned}$$

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

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

where each state has probability

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

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:

$$\begin{aligned} r^{**}_{\tau+1} &= 2h_{\tau+1} + e_{\tau+1}, \\ h_{\tau+1} &= \lambda_0 + \lambda_1 h_\tau + \xi_{\tau+1}, & \text{(A-33)} \end{aligned}$$

[7] 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 $e_{\tau+1} \sim N\left(m_j - 1.2704, v_j^2\right)$ with probability $\Pr\left(s_{\tau+1} = j | \boldsymbol{\Theta}, \boldsymbol{\theta}^t, h^t, M_i', \mathcal{D}^t\right)$. 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 $p\left(\sigma_\xi^{-2}\middle| \mu, \beta, \boldsymbol{\theta}^t, \boldsymbol{Q}, h^t, \lambda_0, \lambda_1, \boldsymbol{\gamma_\theta}, M_i', \mathcal{D}^t\right)$ is readily available as,

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

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

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

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

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

where

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

and

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

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

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

and

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

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

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

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 $\boldsymbol{\Lambda}$. The prior for σ_{κ}^{-2}, the precision of our measure of incompleteness in the combination scheme, and the diagonal elements of $\boldsymbol{\Lambda}^{-1}$, 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 $\mathcal{G}(\underline{s}_{\boldsymbol{\Lambda}}^{-1}, \underline{v}_{\boldsymbol{\Lambda}}(\underline{t}-1))$, respectively. We set informative values on our prior beliefs regarding the incompleteness and the combination weights. Precisely, we set $\underline{v}_{\sigma_\kappa} = \underline{v}_{\boldsymbol{\Lambda}_i} = 1$ 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 $\underline{s}_{\boldsymbol{\Lambda}}^{-1} = 4$, allowing $\mathbf{z}_{t+1}$ to evolve freely over time and differ from the initial value $\mathbf{z}_0$, set to equal weights.[8]

B.2 Posterior simulation

Let $\boldsymbol{\varsigma}$ be the parameter vector of the combination model, that is $\boldsymbol{\varsigma} = (\sigma_{\kappa}^2, \boldsymbol{\Lambda})$. 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

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

made at time τ, and $p\left(\widetilde{\mathbf{r}}_{\tau}\right| D^{\tau-1}\big)$ 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 $\mathbf{Z}_{t+1} = (\mathbf{w}_{t+1}, \mathbf{z}_{t+1}, \varsigma_{t+1})$, where $\varsigma_{t+1} = \varsigma$, $\forall t$. We write the model combination in its state space form as

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

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

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

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

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

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 $\lambda = 0.9$.[9] 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

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

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

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

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 $p\left(\left. r_\tau\right| M_i, \mathcal{D}^{\tau-1}\right)$, 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

[9] 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.

density at time τ, $p(\mathbf{w}_\tau|\mathcal{D}^{\tau-1})$, 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

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

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,

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

where m denotes the model under consideration (either univariate or model combination), and $t^* = \bar{t} - \underline{t}$. 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.[10]

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 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 $\sigma_{\varepsilon}^{-2}$, 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

[10] 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.

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.

References

Banbura M, Giannone D, Reichlin L. 2010. Large Bayesian vector auto regressions. *Journal of Applied Econometrics* **25**: 71–92. ISSN 1099-1255.

Billio M, Casarin R, Ravazzolo F, van Dijk HK. 2013. Time-varying combinations of predictive densities using nonlinear filtering. *Journal of Econometrics* **177**: 213–232.

Carter CK, Kohn R. 1994. On gibbs sampling for state space models. *Biometrika* **81**: pp. 541–553. ISSN 00063444.

Clark TE. 2011. Real-time density forecasts from bayesian vector autoregressions with stochastic volatility. *Journal of Business & Economic Statistics* **29**: 327–341.

Clark TE, Ravazzolo F. 2015. Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics* **30**: 551–575. ISSN 1099-1255.

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

Durbin J, Koopman SJ. 2002. A simple and efficient simulation smoother for state space time series analysis. *Biometrika* **89**: pp. 603–615. ISSN 00063444.

Henkel SJ, Martin JS, Nardari F. 2011. Time-varying short-horizon predictability. *Journal of Financial Economics* **99**: 560 – 580. ISSN 0304-405X.

Kim S, Shephard N, Chib S. 1998. Stochastic volatility: Likelihood inference and comparison with arch models. *The Review of Economic Studies* **65**: 361–393.

Koop G. 2003. *Bayesian Econometrics*. John Wiley & Sons, Ltd.

Poirier DJ. 1995. *Intermediate Statistics and Econometrics: A Comparative Approach*. MIT Press.

Primiceri GE. 2005. Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies* **72**: 821–852.

Rapach DE, Strauss JK, Zhou G. 2010. Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies* **23**: 821–862.

Stock JH, Watson MW. 1996. Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics* **14**: 11–30.

Stock JH, Watson MW. 2004. Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting* **23**: 405–430. ISSN 1099-131X.

Welch I, Goyal A. 2008. A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies* **21**: 1455–1508.

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

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 %** ***

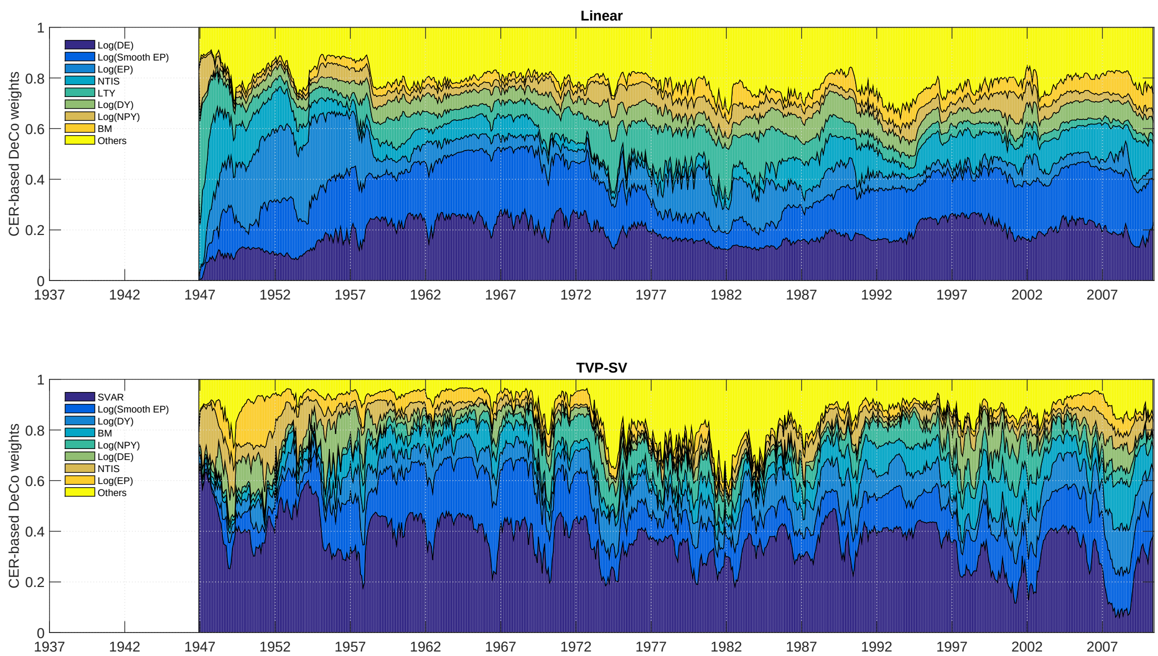
This table reports the out-of-sample R^2 ("OoS R^2"), the average cumulative rank probability score differentials ("CRPSD"), and the average log predictive score differentials ("LSD") for the combination schemes and the individual prediction models of monthly excess returns. The out-of-sample R^2 are 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

Predictor	Panel A: vs. PM		Panel B: vs. PM-SV	
	Linear	TVP-SV	Linear	TVP-SV
Individual models				
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.