

Optimal Asset Allocation with Multivariate Bayesian Dynamic Linear Models

Carlos M. Carvalho **Jared D. Fisher**
University of Texas at Austin* University of Texas at Austin[†]

Davide Pettenuzzo
Brandeis University[‡]

September 25, 2018

Abstract

We introduce a simulation-free method to model and forecast multiple asset returns and employ it to investigate the optimal ensemble of features to include when jointly predicting monthly stock and bond excess returns. Our approach builds on the Bayesian Dynamic Linear Models of [West and Harrison \(1997\)](#), and it can objectively determine, through a fully automated procedure, both the optimal set of regressors to include in the predictive system and the degree to which the model coefficients, volatilities, and covariances should vary over time. When applied to a portfolio of five stock and bond returns, we find that our method leads to large forecast gains, both in statistical and economic terms. In particular, we find that relative to a standard no-predictability benchmark, the optimal combination of predictors, stochastic volatility, and time-varying covariances increases the annualized certainty equivalent returns of a leverage-constrained power utility investor by more than 500 basis points.

Keywords: Optimal asset allocation, Bayesian econometrics, Dynamic linear models

JEL Classifications: C11; C22; G11; G12

*McCombs Schools of Business, University of Texas, 2110 Speedway Stop, Austin, TX 78712. carlos.carvalho@mcombs.utexas.edu

[†]McCombs Schools of Business, University of Texas, 2110 Speedway Stop, Austin, TX 78712. jared.fisher@utexas.edu

[‡]Brandeis University, Sachar International Center, 415 South St, Waltham, MA 02453. dpettenu@brandeis.edu

1 Introduction

The study of portfolio theory and its implications for the asset allocation decisions of investors has and continues to play a central role in financial economics. Within this literature, a highly debated item over the years has been the question of whether asset returns are predictable and the extent to which this predictability affects the investor’s optimal allocation choices.

There is by now an extensive empirical literature that has found evidence for predictability in stock and bond returns by means of valuation ratios, interest rates, and macroeconomic quantities.¹ Prior to the turn of the century, much of this literature focused on identifying variables that had significant and robust in-sample predictive power when forecasting returns. However, thanks in part to evidence uncovered in studies such as [Bossaerts and Hillion \(1999\)](#), [Ang and Bekaert \(2007\)](#), and [Welch and Goyal \(2008\)](#), in recent years the emphasis has been gradually shifting from in-sample to out-of-sample predictability. A similar pattern has been observed for bond returns, where [Thornton and Valente \(2012\)](#) have shown that the information subsumed into forward rates and forward spreads, while quite successful in-sample, does not generate systematic economic value to investors out-of-sample.

The disparities between in-sample and out-of-sample evidence of return predictability can be in part explained by the presence of model instability in return prediction models. Due to the regular occurrence of a multitude of shocks to financial markets and the overall economy, investors are facing a constantly evolving, uncertain landscape and need to resort to highly adaptive methods when building their forecasts. By now, it is clear that not a single feature alone, but an ensemble of features is required to cope with the resulting uncertainty and instability as well as generate good predictions. This has been shown to be true for stock returns ([Johannes et al., 2014](#)) as well as for bond returns ([Gargano et al., 2017](#)). In particular, features that satisfy these out-of-sample needs include model and parameter uncertainty, time-varying volatility, time-varying parameters, and economically motivated constraints.

While there is ample evidence backing said ensemble of features when modeling returns on a single risky asset, surprisingly no study has examined how these features interact when

¹See for example [Fama and Schwert \(1977\)](#), [Campbell and Shiller \(1988\)](#), [Lettau and Ludvigson \(2001\)](#), [Lewellen \(2004\)](#), and [Ang and Bekaert \(2007\)](#).

jointly forecasting the returns of multiple risky assets. Yet, most investors hold many risky assets at once in their portfolios, which makes this an empirically relevant question. The primary contribution of this paper is to unify the features highlighted in the aforementioned papers into a single, computationally friendly framework capable of jointly handling multiple risky assets from different classes. Specifically, our framework builds on the Bayesian Dynamic Linear Models (DLMs) of [West and Harrison \(1997\)](#) and [Gruber and West \(2016\)](#) and examines a Bayesian agent who recursively updates her prior beliefs as new data is observed, therefore mimicking the real time decision making process of an investor.

The key element of our modeling approach is the ability to integrate a number of useful features into a flexible yet computationally simple method. First, our approach is well suited to integrate parameter uncertainty into the problem, as the DLMs yield predictive densities, rather than point forecasts, for each asset return. Second, the DLM framework allows for multivariate stochastic volatility. Both [Johannes et al. \(2014\)](#) and [Gargano et al. \(2017\)](#) find that stochastic volatility is a key feature to incorporate when modeling and forecasting stock and bond returns. The benefits of stochastic volatility are particularly pronounced during periods of very high market turmoil, such as the dot-com bubble as well as the most recent financial crisis. Given our emphasis on jointly modeling multiple risky assets, the key adjustment herein is how we model time-variation in the cross-asset covariances. We provide two alternative approaches to handle this. Our first method builds on the Wishart DLM (W-DLM, henceforth) of [West and Harrison \(1997\)](#). Two key restrictions of the W-DLM are that, first, it forces all the assets in the system to share the same vector of predictor variables, and second, that variances and covariances are modeled in the same structure and must time-varying jointly. While in some settings this requirement may be appropriate, it is likely not a desirable feature when working with returns from very heterogeneous asset classes, such as equity and fixed income. To alleviate these concerns, our second approach builds on the Simultaneous Graphical DLM (SG-DLM, henceforth) of [Gruber and West \(2016\)](#). The SG-DLM permits each asset to feature its own set of predictor variables. In addition, the SG-DLM can easily be modified to allow for separate degrees of time variation for variances and covariances, which we find to be a very useful feature with financial returns. Most importantly, both DLM methods, as we present them, yield closed-form solutions for all the moments of the

posterior distributions and predictive densities, and hence are computationally faster than the particle filter algorithm of [Johannes et al. \(2014\)](#) or the Markov chain Monte Carlo approaches of [Gargano et al. \(2017\)](#) and others.²

Third, our models allow for time-variation in the regression coefficients. It has been shown extensively that the regression coefficients of asset return predictive regressions change over time ([Viceira, 1997](#); [Pastor and Stambaugh, 2001](#); [Kim et al., 2005](#); [Paye and Timmermann, 2006](#); [Lettau and Van Nieuwerburgh, 2008](#); [Pettenuzzo and Timmermann, 2011](#)). Rather than allowing for discrete non-recurring shifts, we let the regression coefficients evolve over time by adopting a flexible time-varying parameters specification. In this regard, our work is similar to [Dangl and Halling \(2012\)](#), who model and forecast the S&P 500 index and find that time-varying parameter models are strongly preferable to predictive regression with constant coefficients.

Fourth, our approach controls for model uncertainty through model averaging. Thanks to the computational savings afforded by our approach, we are able to consider in reasonable computation time both the uncertainty regarding the degree to which parameters, volatilities, and covariances vary over time, as well as which predictors should be included in the model. We accomplish this by first fitting a separate DLM to each possible permutation of predictors and degrees of time-variation. Next, we compute the predictive densities implied by each of these permutations and combine them together using both equal-weighted and score-weighted combinations. The latter weights the different model permutations according to their historical statistical fit, as measured by their logarithmic predictive scores.

Our secondary contribution is to empirically test the roles played by these features when forecasting multiple stock and bond returns. More specifically, we evaluate the performance of the W-DLM and SG-DLM models by jointly modeling the monthly excess returns on the five- and ten-year Treasury bonds, as well as the excess returns on the size-sorted small-, mid-, and large-cap stock portfolios. As for the predictors, we include the 15 variables studied in [Welch and Goyal \(2008\)](#) as well as the three predictors for bond returns considered by [Gargano et al. \(2017\)](#), namely forward spreads, the [Cochrane and Piazzesi \(2005\)](#) factor and

²To have its forward filter be closed-form, SG-DLMs must assume an appropriate dependence structure across the asset returns in the system. Details are given in [Subsection 2.2](#).

the [Ludvigson and Ng \(2009\)](#) factor. We then estimate a W-DLM and a SG-DLM for each different combination of stock and bond predictors as well as combinations of different degrees of time variation in the regression coefficients, variances, and covariances. These individual DLMs are then averaged together in different groups to account for the aforementioned model uncertainty.

We evaluate the predictive performance of the various models and features over the 1985-2014 period against a simple no-predictability benchmark, and we find large statistical and economic benefits from using the appropriate ensemble of features. Among the features we consider, we find that W-DLMs and SG-DLMs with stochastic volatility bring the largest gains in terms of statistical predictability. In terms of economic predictability, which we quantify using certainty equivalent returns, we find that the optimal set of features includes SG-DLMs with stochastic volatility and time-varying covariances. In particular, we find that when using the optimal set of features our leverage-constrained power utility investor earns over 500 basis points (on an annualized basis) more than if she relied on the no-predictability benchmark.

Our paper relates to several branches of the literature. The papers most closely related to us are [Dangl and Halling \(2012\)](#), [Johannes et al. \(2014\)](#), and [Gargano et al. \(2017\)](#). All three papers focus on modeling and forecasting asset returns (stocks in the first two cases, treasury bonds in the last case) using flexible model specifications and building density forecasts that are robust to the presence of model instability and model uncertainty. In particular, [Dangl and Halling \(2012\)](#) use a DLM that is similar to what we employ here and allow for model uncertainty over different predictors and degrees of time-variation in the regression coefficients (but do not allow for stochastic volatility). In contrast, [Johannes et al. \(2014\)](#) and [Gargano et al. \(2017\)](#) allow for both time-varying regression coefficients and stochastic volatility, but because of their reliance on MCMC methods are forced to set a priori the degree to which parameters and volatility change over time. In addition, all three papers focus on univariate models and forecast a single financial asset at a time.³ Relative to their setup, our approach

³While the main focus in [Gargano et al. \(2017\)](#) is on univariate predictive regressions, they include a small application where they extend their setup to forecasting multiple treasury bond returns (differing in their maturities) at once.

jointly models multiple risky assets and takes into account the model uncertainty that arises from the availability of multiple predictors and from not knowing the degree of time variation in the regression coefficients, variances, and covariances.

There is also a small literature that has focused on forecasting multiple risky assets from different asset classes. [Brennan et al. \(1997\)](#) look at a portfolio that includes a stock index, a bond index, and cash, and forecast each asset return using a distinct predictor. This leads to a number of computational complexities, which they solve by estimating partial differential equations numerically. [Wachter and Warusawitharana \(2009\)](#) model the returns of both a stock index and a long-term bond using a single predictor variable but, because of their specific setup, need to rely on MCMC methods. [Gao and Nardari \(2018\)](#) model the returns of stocks, bonds, cash, and commodities by fitting multiple models with single predictors and averaging them with equal weights. They also allow for a time-varying covariance matrix, which they implement via the dynamic conditional correlation method of [Engle \(2002\)](#). Relative to these papers, ours provides the first attempt to objectively determine the optimal combination of features to include when modeling multiple risky assets at once, and does so by using a computationally efficient and simulation-free approach.

The remainder of the paper is organized as follows. [Section 2](#) introduces the W-DLM and SG-DLM model specifications, the set of features we control for and our approach for averaging across all permutations of predictors and model characteristics. Next, [Section 3](#) describes the data and priors we adopted, while [Section 4](#) summarizes our empirical analysis and the results we obtain. Finally, [Section 5](#) provides some concluding remarks.

2 Our Approach

In this section, we introduce the approach we rely on to estimate and forecast multiple risky asset returns. We begin by describing in [Subsection 2.1](#) and [Subsection 2.2](#) the two Bayesian dynamic linear models (DLMs) we work with, namely the Wishart Dynamic Linear Model and the Simultaneous Graphical Dynamic Linear Model. Both methods allow the regression coefficients, variances, and covariances to vary over time and are therefore capable of coping with the model instability that plagues the relationship between asset returns and predictor

variables. At the same time, both methods require the investor to know a priori the degree of time variation in the model parameters as well as the right combination of predictors to include in the regressions. In practice, the investor is likely unaware of what the optimal predictive model may look like, and is therefore facing uncertainty across all these dimensions. In [Subsection 2.3](#), we describe a fully-automated data-based approach that we use to resolve this uncertainty.

2.1 Wishart Dynamic Linear Model

One of the key advantages of DLMS, compared to other Bayesian approaches, is that they feature closed-form solutions for all parameter updates as well as model forecasts. This is accomplished by a simulation-free procedure, known as a deterministic forward filter, which simulates how most people think, i.e. modifying their prior beliefs in real time as new data becomes available. More specifically, the posterior distribution of all model parameters at time $t - 1$ becomes the prior at time t , and once time t data becomes available, a simple set of formulas merge time t priors and time t likelihood into time t posteriors. As part of this process, real time predictive densities and point forecasts can be obtained in a straightforward manner. This procedure is repeated throughout the sample, thus yielding a sequence of posterior distributions and predictive densities.

Our first approach builds on the Wishart DLM (W-DLM) of [West and Harrison \(1997\)](#), which allows for time-varying regression coefficients as well as time-varying variances and covariances. As its name suggests, the W-DLM assumes that the error covariance matrix follows an inverse-Wishart distribution (\mathcal{IW} , henceforth). This is paired with the additional restriction that all the equations in the system share the same predictor variables.⁴

Let \mathbf{r}_t denote a $q \times 1$ vector of log excess returns at time t ($t = 1, \dots, T$) and \mathbf{x}_{t-1} represent a $p \times 1$ vector of lagged predictor variables, common to all q risky assets (throughout, we use bold lower-case letters to represent vectors and bold capitalized letters to represent matrices).⁵

⁴The W-DLM is a generalization of the approach employed by [Dangl and Halling \(2012\)](#) to model and forecast stock returns. Relative to [Dangl and Halling \(2012\)](#), the W-DLM allows a modeler to model multiple risky assets at once and to include time varying variances and covariances. It is essentially a deterministic forward-filter analog of the approach of [Wachter and Warusawitharana \(2009\)](#).

⁵ \mathbf{x}_{t-1} may or may not include a constant/intercept term.

The W-DLM can be written as:

$$\mathbf{r}_t = \mathbf{B}_t' \mathbf{x}_{t-1} + \mathbf{v}_t \quad \mathbf{v}_t | \boldsymbol{\Sigma}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_t) \quad (1)$$

where \mathbf{B}_t is the $p \times q$ matrix of time-varying regression coefficients, which evolve over time according to pq random walk processes,

$$\text{vec}(\mathbf{B}_t) = \text{vec}(\mathbf{B}_{t-1}) + \boldsymbol{\omega}_t \quad \boldsymbol{\omega}_t | \boldsymbol{\Sigma}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_t \otimes \mathbf{W}_t) \quad (2)$$

with $\boldsymbol{\omega}_t$ denoting a $pq \times 1$ vector of zero-mean normally distributed error terms, $\text{vec}(\cdot)$ is the vectorization operator, and \otimes represents the Kronecker product.^{6,7}

Next, the $q \times 1$ error vector \mathbf{v}_t is independently and normally distributed over time with variance-covariance matrix $\boldsymbol{\Sigma}_t$, given by

$$\boldsymbol{\Sigma}_t = \begin{bmatrix} \sigma_{1,t}^2 & \sigma_{12,t} & \cdots & \sigma_{1q,t} \\ \sigma_{12,t} & \sigma_{2,t}^2 & \cdots & \sigma_{2q,t} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1q,t} & \sigma_{2q,t} & \cdots & \sigma_{q,t}^2 \end{bmatrix} \quad (3)$$

where both the variances $(\sigma_{1,t}^2, \dots, \sigma_{q,t}^2)$ and the covariances $\sigma_{ij,t}$ ($i, j = 1, \dots, q, j > i$) are allowed to vary over time. Finally, the $p \times p$ matrix \mathbf{W}_t controls the degree of time variation of the regression coefficient matrix \mathbf{B}_t , and we will specify its exact form below.

The model in (1)-(3) is completed by specifying the initial states for both the regression coefficients and the variance-covariance matrix at time $t = 0$. These are given by the following distributions, where \mathcal{D}_0 denotes the information set available at time $t = 0$

$$\begin{aligned} \text{vec}(\mathbf{B}_0) | \boldsymbol{\Sigma}_0, \mathcal{D}_0 &\sim \mathcal{N}(\text{vec}(\mathbf{M}_0), \boldsymbol{\Sigma}_0 \otimes \mathbf{C}_0) \\ \boldsymbol{\Sigma}_0 | \mathcal{D}_0 &\sim \mathcal{IW}(n_0, \mathbf{S}_0) \end{aligned} \quad (4)$$

Here, the $p \times q$ matrix \mathbf{M}_0 denotes the mean of the coefficient matrix \mathbf{B}_0 , while the $p \times p$ matrix \mathbf{C}_0 summarizes the degree of confidence in \mathbf{M}_0 . Similarly, \mathbf{S}_0 represents an estimate of the $q \times q$ error covariance matrix $\boldsymbol{\Sigma}_0$, which follows an Inverse-Wishart distribution with n_0

⁶Specifically, the vectorization of an $m \times n$ matrix \mathbf{A} , denoted $\text{vec}(\mathbf{A})$, is the $mn \times 1$ column vector obtained by stacking the columns of the matrix \mathbf{A} on top of one another.

⁷The W-DLM in (1) can also be written using the matrix-normal distribution, i.e. $\mathbf{B}_t = \mathbf{B}_{t-1} + \boldsymbol{\Omega}_t$, $\boldsymbol{\Omega}_t | \boldsymbol{\Sigma}_t \sim \mathcal{MN}(\mathbf{0}, \mathbf{W}_t, \boldsymbol{\Sigma}_t)$. Here $\boldsymbol{\Omega}_t$ follows a matrix-normal distribution \mathcal{MN} with left variance matrix \mathbf{W}_t and right variance matrix $\boldsymbol{\Sigma}_t$. This is the notation adopted by West and Harrison (1997, section 16.2). See also Dawid (1981) for a description of the matrix-normal distribution and its properties.

degrees of freedom. n_0 , in turn, can be interpreted as the effective sample size of the initial state.

In practice, (4) can also be interpreted as the posterior distribution of the parameters at time $t = 0$. We use this initial posterior in a process called *evolution*, where at any point in time t ($t = 1, \dots, T$) we use the posterior distribution from time $t - 1$ to compute the prior distribution of the parameters at time t . This is given by

$$\begin{aligned} \text{vec}(\mathbf{B}_t) | \boldsymbol{\Sigma}_t, \mathcal{D}_{t-1} &\sim \mathcal{N}(\text{vec}(\mathbf{M}_{t-1}), \boldsymbol{\Sigma}_t \otimes \hat{\mathbf{C}}_t) \\ \boldsymbol{\Sigma}_t | \mathcal{D}_{t-1} &\sim \mathcal{IW}(\hat{n}_t, \mathbf{S}_{t-1}) \end{aligned} \quad (5)$$

where $\hat{\mathbf{C}}_t$ and \hat{n}_t are modified versions of \mathbf{C}_{t-1} and n_{t-1} and are used as estimates of \mathbf{C}_t and n_t . In particular, we set

$$\hat{\mathbf{C}}_t = \frac{1}{\delta_\beta} \mathbf{C}_{t-1} \quad (6)$$

and

$$\hat{n}_t = \delta_v n_{t-1} \quad (7)$$

where $\delta_\beta \in (0, 1]$ and $\delta_v \in (0, 1]$ denote discount factors. δ_β is incorporated into the model (and hence we can control the degree of time variation of the regression coefficient matrix \mathbf{B}_t) by rewriting the $p \times p$ matrix \mathbf{W}_t in (2) as

$$\mathbf{W}_t = \frac{1 - \delta_\beta}{\delta_\beta} \mathbf{C}_{t-1}, \quad (8)$$

which suggests that the smaller the discount factor δ_β is, the larger the elements of the covariance matrix \mathbf{W}_t will be, thus increasing the variance/uncertainty around time t regression coefficients and allowing \mathbf{B}_t to move further away from \mathbf{B}_{t-1} . In the extreme case of $\delta_\beta = 1$ we have that $\hat{\mathbf{C}}_t = \mathbf{C}_{t-1}$ and $\mathbf{W}_t = 0$, which means that when $\delta_\beta = 1$ the regression coefficient matrix \mathbf{B}_t does not vary over time. As for δ_v , note that (5) implies that

$$\mathbb{E}(\boldsymbol{\Sigma}_t | \mathcal{D}_{t-1}) = \frac{1}{\hat{n}_t - q - 1} \mathbf{S}_{t-1} \quad (9)$$

which means that the smaller δ_v is, the larger the expected value of all elements in the error covariance matrix will be. Also, it can be shown that for large t , $0 < \delta_v < 1$ implies that the posterior estimates of the variances and covariances across series essentially become

exponentially weighted moving averages of the past sample variances and sample covariances, with weights that decay over time as a function of δ_v . This, in turn, suggests that the smaller the discount factor δ_v is, the quicker Σ_t can adapt to the new data and the more it can move away from Σ_{t-1} . Finally, in the extreme case of $\delta_v = 1$, we obtain a model where there is no discounting of the old data and thus Σ_t is assumed constant, i.e. a constant volatility model.

With (5) in hand, it becomes possible to compute the predictive distribution of \mathbf{r}_t , conditional on the information set available at time $t - 1$. In particular, we have that

$$\mathbf{r}_t | \delta_\beta, \delta_v, \mathcal{D}_{t-1} \sim \mathcal{T}_{\hat{n}_t} \left(\mathbf{M}'_{t-1} \mathbf{x}_{t-1}, \quad \mathbf{S}_{t-1} (1 + \mathbf{x}'_{t-1} \hat{\mathbf{C}}_t \mathbf{x}_{t-1}) \right). \quad (10)$$

where $\mathcal{T}_{\hat{n}_t}$ denotes a Student's t-distribution with \hat{n}_t degrees of freedom.⁸ This implies that the conditional forecast of the mean vector and variance-covariance matrix of \mathbf{r}_t will be given by

$$\mathbb{E}[\mathbf{r}_t | \delta_\beta, \delta_v, \mathcal{D}_{t-1}] = \mathbf{M}'_{t-1} \mathbf{x}_{t-1} \quad (11)$$

$$\text{Cov}[\mathbf{r}_t | \delta_\beta, \delta_v, \mathcal{D}_{t-1}] = \frac{\hat{n}_t}{\hat{n}_t - 2} \mathbf{S}_{t-1} (1 + \mathbf{x}'_{t-1} \hat{\mathbf{C}}_t \mathbf{x}_{t-1}). \quad (12)$$

After observing the actual returns for time period t , we can update the prior for time t from (5) into the posterior for time t . We provide the details of the closed-form updating equations in [Appendix A](#), where we show how from the initial states in (4), the sequence of regression coefficients $\{\mathbf{B}_t\}_{t=1}^T$ and variance-covariance matrices $\{\Sigma_t\}_{t=1}^T$ can be obtained by a simple and very fast forward filter. Thus, the W-DLM deterministically gives the posterior distribution of the model parameters at each time step, avoiding the need for computationally expensive Markov chain Monte Carlo simulation methods.

While computationally very fast, the W-DLM presents three key drawbacks. First, very much like [Wachter and Warusawitharana \(2009\)](#)'s model, the W-DLM uses the same predictors for each asset. The severity of this restriction will depend on the particular assets being modeled, but it is not hard to imagine situations where this restriction may not be desirable. Second, the conjugate inverse Wishart prior, while computationally very

⁸Note that we have opted for a notation where we make explicit the dependence of the predictive distribution for \mathbf{r}_t (and its moments) to the choices made with respect to the two discount factors, δ_β and δ_v .

convenient, is notoriously inflexible and may not adapt well to underlying data.⁹ Finally, by construction the W-DLM features a single discount factor for the entire covariance matrix, which means that both the variances and covariances will be discounted in the same way. In the next section, we present a more general approach that will permit us to relax all three drawbacks of the W-DLM.

2.2 Simultaneous Graphical Dynamic Linear Model

Our second approach builds on the simultaneous graphical dynamic linear model (SG-DLM) of Gruber and West (2016). Relative to the W-DLM method described in the previous section, one of the key advantages of the SG-DLM is that it can accommodate asset-specific regressors, while still allowing for time-varying regression coefficients, variances, and covariances. This is accomplished through a modeling strategy that “decouples” the joint dynamic system into separate univariate models for each of the risky assets, taking into full account the contemporaneous dependencies across assets. In turn, these univariate models can be updated with great computational speed, thus preserving the closed-form forward filter nature of the algorithm. We begin by re-writing the joint dynamic system for the q excess returns \mathbf{r}_t as follows:

$$\mathbf{r}_t = \begin{pmatrix} \mathbf{x}'_{1,t-1}\boldsymbol{\beta}_{1t} \\ \vdots \\ \mathbf{x}'_{q,t-1}\boldsymbol{\beta}_{qt} \end{pmatrix} + \begin{pmatrix} \mathbf{r}'_{-1,t}\boldsymbol{\gamma}_{1t} \\ \vdots \\ \mathbf{r}'_{-q,t}\boldsymbol{\gamma}_{qt} \end{pmatrix} + \boldsymbol{\nu}_t \quad \boldsymbol{\nu}_t | \boldsymbol{\Omega}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_t) \quad (13)$$

where $\mathbf{x}_{j,t-1}$ ($j = 1, \dots, q$) denotes the $p_j \times 1$ vector of asset j ’s specific lagged predictors (possibly including an intercept), while $\mathbf{r}_{-j,t}$ represents the contemporaneous log excess returns of all assets other than asset j . Similarly, $\boldsymbol{\beta}_{jt}$ denotes the $p_j \times 1$ vector of the predictors’ coefficients while $\boldsymbol{\gamma}_{jt}$ is the $(q-1) \times 1$ vector of coefficients capturing the contemporaneous correlations between asset j ’s log excess return and the remaining $q-1$ log excess returns. Finally, $\boldsymbol{\Omega}_t = \text{diag}(\sigma_{1t}^2, \dots, \sigma_{qt}^2)$ is a $q \times q$ matrix with the assets’ error variances on the diagonal.

Note that relative to the W-DLM specification in (1), which models the contemporaneous correlations across asset returns through the full variance-covariance matrix $\boldsymbol{\Sigma}_t$, the system

⁹See for example Barnard et al. (2000) or Gelman and Hill (2006). One simple point is that the inverse Wishart has only a single parameter governing the variability about all of its elements, thus, for a distribution of a covariance matrix, your uncertainty about all the variances and covariances must be the same.

in (13) handles the contemporaneous correlations by introducing the γ_{jt} parameters and the $\mathbf{r}_{-j,t}$ regressors ($j = 1, \dots, q$) while leaving all elements of the error term $\boldsymbol{\nu}_t$ contemporaneously uncorrelated, i.e. $\nu_{it} \perp \nu_{jt}$ for all $i \neq j$. This modeling choice, as we will show shortly, is what allows the SG-DLM to continue working with a closed-form forward-filter even after relaxing the restrictions enforced by the W-DLM.

We proceed by combining all elements of γ_{1t} to γ_{qt} into the $q \times q$ zero-diagonal matrix $\boldsymbol{\Gamma}_t$ as follows,

$$\boldsymbol{\Gamma}_t = \begin{bmatrix} 0 & \gamma_{12,t} & \cdots & \gamma_{1q-1,t} & \gamma_{1q,t} \\ \gamma_{21,t} & 0 & \cdots & \gamma_{2q-1,t} & \gamma_{2q,t} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma_{q1,t} & \gamma_{q2,t} & \cdots & \gamma_{qq-1,t} & 0 \end{bmatrix} \quad (14)$$

which in turn allows us to rewrite (13) as

$$\mathbf{r}_t = \begin{pmatrix} \mathbf{x}'_{1,t-1} \boldsymbol{\beta}_{1t} \\ \vdots \\ \mathbf{x}'_{q,t-1} \boldsymbol{\beta}_{qt} \end{pmatrix} + \boldsymbol{\Gamma}_t \mathbf{r}_t + \boldsymbol{\nu}_t \quad \boldsymbol{\nu}_t | \boldsymbol{\Omega}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_t). \quad (15)$$

It is easy to show that we can further rearrange (15) to write

$$\mathbf{r}_t = (I - \boldsymbol{\Gamma}_t)^{-1} \begin{pmatrix} \mathbf{x}'_{1,t-1} \boldsymbol{\beta}_{1t} \\ \vdots \\ \mathbf{x}'_{q,t-1} \boldsymbol{\beta}_{qt} \end{pmatrix} + \mathbf{u}_t \quad \mathbf{u}_t | \boldsymbol{\Sigma}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_t) \quad (16)$$

where $\boldsymbol{\Sigma}_t = (I - \boldsymbol{\Gamma}_t)^{-1} \boldsymbol{\Omega}_t ((I - \boldsymbol{\Gamma}_t)^{-1})'$ is now a full variance-covariance matrix capturing the contemporaneous correlations among the q assets. As shown by Gruber and West (2016), the presence of the $(I - \boldsymbol{\Gamma}_t)^{-1}$ term in (16) significantly complicates the inference, as the joint posterior of the parameters is now proportional to the determinant $|I - \boldsymbol{\Gamma}_t|$ times the product of q univariate normal densities, i.e.

$$p(\mathbf{r}_t | \boldsymbol{\beta}_{1t}, \dots, \boldsymbol{\beta}_{qt}, \boldsymbol{\Gamma}_t, \boldsymbol{\Omega}_t) \propto |I - \boldsymbol{\Gamma}_t| \prod_{j=1}^q p(r_{jt} | \boldsymbol{\beta}_{jt}, \gamma_{jt}, \sigma_{jt}^2). \quad (17)$$

The obvious exception to this rule is the case where $|I - \boldsymbol{\Gamma}_t| = 1$. In this case, as we will show below, it becomes possible to derive the multivariate distribution of all assets using fast and reliable univariate forward filters similar to those introduced by West and Harrison (1997).

This is indeed the avenue we explore here.¹⁰

¹⁰In particular, we build on Zhao et al. (2016), who present a forward filter algorithm for a dynamic linear system with a fully recursive triangular specification (where (14) is a triangular matrix), where the parameters within each equation of the system are updated individually.

In particular, we follow [Primiceri \(2005\)](#), [Carriero et al. \(2016\)](#), and [Koop et al. \(2017\)](#) and assume that the dynamic system in (13) is fully recursive. This, in turn, implies that the $\mathbf{\Gamma}_t$ matrix in (15) becomes lower triangular, still featuring zeros on its main diagonal. Next, we write r_{jt} , the log excess return of risky asset j at time t , as a linear combination of a $p_j \times 1$ vector of asset-specific lagged predictors $\mathbf{x}_{j,t-1}$ as well as the contemporaneous log excess returns from the previous $j - 1$ assets, which we denote with $\mathbf{r}_{< j, t}$,

$$r_{jt} = \mathbf{x}'_{j,t-1} \boldsymbol{\beta}_{jt} + \mathbf{r}'_{< j, t} \boldsymbol{\gamma}_{< j, t} + \nu_{jt} \quad \nu_{jt} \sim N(0, \sigma_{jt}^2) \quad (18)$$

where $\boldsymbol{\gamma}_{< j, t}$ is the $(j - 1) \times 1$ vector of coefficients associated with the contemporaneous excess returns $\mathbf{r}_{< j, t}$. We now specify the law of motion for the regression coefficients $\boldsymbol{\beta}_{jt}$ and $\boldsymbol{\gamma}_{< j, t}$:

$$\begin{pmatrix} \boldsymbol{\beta}_{jt} \\ \boldsymbol{\gamma}_{< j, t} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}_{j,t-1} \\ \boldsymbol{\gamma}_{< j, t-1} \end{pmatrix} + \boldsymbol{\omega}_{jt} \quad \boldsymbol{\omega}_{jt} \sim N(\mathbf{0}, \mathbf{W}_{jt}). \quad (19)$$

where $\boldsymbol{\omega}_{jt}$ is the $(p_j + j - 1) \times 1$ vector of evolution errors with covariance matrix \mathbf{W}_{jt} .

The SG-DLM is completed by specifying the initial states of the model parameters, that is, regression coefficients $\boldsymbol{\beta}_{jt}$, contemporaneous returns coefficients $\boldsymbol{\gamma}_{< j, t}$, and variance term σ_{jt}^2 . For each asset j , we write

$$\begin{aligned} \begin{pmatrix} \boldsymbol{\beta}_{j0} \\ \boldsymbol{\gamma}_{< j, 0} \end{pmatrix} \Big| \sigma_{j0}^2, \mathcal{D}_0 &\sim \mathcal{N} \left(\mathbf{m}_{j0}, \frac{\sigma_{j0}^2}{s_{j0}} \mathbf{C}_{j0} \right) \\ \sigma_{j0}^{-2} | \mathcal{D}_0 &\sim \mathcal{G} \left(\frac{n_{j0}}{2}, \frac{n_{j0} s_{j0}}{2} \right) \end{aligned} \quad (20)$$

where \mathbf{m}_{j0} is a $(p_j + j - 1) \times 1$ vector denoting the mean of the coefficients $(\boldsymbol{\beta}'_{j0}, \boldsymbol{\gamma}'_{< j, 0})'$, while \mathbf{C}_{j0} is a $(p_j + j - 1) \times (p_j + j - 1)$ covariance matrix factor summarizing the uncertainty surrounding the mean estimates \mathbf{m}_{j0} . The initial error precision $1/\sigma_{j0}^2$ follows a Gamma distribution with mean $1/s_{j0}$ and degrees of freedom n_{j0} . n_{j0} can be interpreted as the effective sample size of this initial posterior. We further abbreviate these two distributions using the joint Normal-Gamma distribution

$$\begin{pmatrix} \boldsymbol{\beta}_{j0} \\ \boldsymbol{\gamma}_{< j, 0} \end{pmatrix}, \sigma_{j0}^2 \Big| \mathcal{D}_0 \sim \mathcal{NG}(\mathbf{m}_{j0}, \mathbf{C}_{j0}, n_{j0}, s_{j0}). \quad (21)$$

As with the initial conditions for the W-DLM in (4), (21) can be interpreted as the posterior distribution of the model parameters at time $t = 0$. Once this process is initialized,

at any given point in time t ($t = 1, \dots, T$) we can use the posterior distribution from time $t-1$ to compute the prior distributions of the model parameters at time t . These are given by

$$\left(\begin{matrix} \beta_{jt} \\ \gamma_{<j,t} \end{matrix} \right), \sigma_{jt}^2 \Big| \mathcal{D}_{t-1} \sim \mathcal{NG}(\mathbf{m}_{j,t-1}, \hat{\mathbf{C}}_{jt}, \hat{n}_{jt}, s_{j,t-1}). \quad (22)$$

where $\hat{\mathbf{C}}_{jt}$ and \hat{n}_{jt} are modified versions of $\mathbf{C}_{j,t-1}$ and $n_{j,t-1}$, and are used as estimates of $\mathbf{C}_{j,t}$ and $n_{j,t}$. In particular, we set

$$\hat{\mathbf{C}}_{j,t} = \begin{bmatrix} \mathbf{C}_{\beta\beta j,t-1}/\delta_{\beta j} & \mathbf{C}_{\beta\gamma j,t-1} \\ \mathbf{C}_{\gamma\beta j,t-1} & \mathbf{C}_{\gamma\gamma j,t-1}/\delta_{\gamma j} \end{bmatrix}. \quad (23)$$

and

$$\hat{n}_{jt} = \delta_{vj} n_{j,t-1} \quad (24)$$

where $\delta_{\beta j} \in (0, 1]$, $\delta_{\gamma j} \in (0, 1]$, and $\delta_{vj} \in (0, 1]$ denote asset-specific discount factors. In particular, as shown in (23), the updated variance term $\hat{\mathbf{C}}_{j,t}$ features different blocks, separating asset j 's predictor coefficients β_{jt} from asset j 's correlation factors $\gamma_{<j,t}$. In turn, this gives the user the freedom to introduce, asset by asset, a separate discount factor for the correlations ($\delta_{\gamma j}$) and the predictor coefficients ($\delta_{\beta j}$), allowing each asset's dynamic regression coefficients and correlation factors to evolve over time at potentially different paces.¹¹ It is possible to show that

$$\mathbf{W}_{jt} = \begin{bmatrix} (\frac{1}{\delta_{\beta j}} - 1)\mathbf{C}_{\beta\beta j,t-1} & 0 \\ 0 & (\frac{1}{\delta_{\gamma j}} - 1)\mathbf{C}_{\gamma\gamma j,t-1} \end{bmatrix} \quad (25)$$

which suggests that the smaller the discount factors $\delta_{\beta j}$ and $\delta_{\gamma j}$ are, the larger the elements in the respective blocks of the covariance matrix \mathbf{W}_{jt} will be, thus increasing the chances that β_{jt} and $\gamma_{<j,t}$ will move further away from $\beta_{j,t-1}$ and $\gamma_{<j,t-1}$.¹² As for δ_{vj} , much like δ_v with the W-DLM, we have that small values of δ_{vj} lead to large variability (and thus flexibility) in the volatilities, with σ_{jt}^2 allowed to move further away from $\sigma_{j,t-1}^2$. In contrast, when $\delta_{vj} = 1$ there is no discounting of past data and, as a result, σ_{jt}^2 does not vary over time.¹³

¹¹This mimics the block discounting approach introduced by West and Harrison (1997, Section 6.3.2).

¹²Note that the zero off-diagonal blocks in (25) represent an assumption (stemming from West and Harrison (1997, Section 6.3.2)), namely that the correlations between the predictor coefficients β_{jt} and the correlation factors $\gamma_{<j,t}$ are constant (but not zero). This assumption, in turn, leads to having no discount factors in the denominators of the off-diagonal blocks of $\hat{\mathbf{C}}_{j,t}$ in (23).

¹³We note that while in principle the SG-DLM permits each asset to have its own degree of time variation in coefficients, variances, and covariances, it is also quite easy to introduce restrictions in the model setup. For example, one could imagine a situation where all assets within a given class (e.g., bonds or stocks) share the same discount factors, or even a situation where, as it was the case with the W-DLM, all the assets in the system share the same discount factors.

Once (22) is available, it becomes possible to derive the predictive distribution for \mathbf{r}_t , conditional on the information set available at time $t - 1$. Thanks to the fully recursive identification strategy we adopted, we can proceed sequentially through the q equations of the dynamic system. Starting with the first asset in the system, we have that

$$\mathbb{E}[r_{1t}|\boldsymbol{\delta}_j, \mathcal{D}_{t-1}] = \mathbf{x}'_{1,t-1} \mathbf{m}_{1,t-1} \quad (26)$$

$$Var[r_{1t}|\boldsymbol{\delta}_j, \mathcal{D}_{t-1}] = \frac{\hat{n}_{1t}}{\hat{n}_{1t} - 2} (\mathbf{x}'_{1,t-1} \hat{\mathbf{C}}_{1t} \mathbf{x}_{1,t-1} + s_{1,t-1}). \quad (27)$$

where, as with the W-DLM forecasts, we have highlighted the dependence of these predictive moments on the choices made regarding the discount factors, that is $\boldsymbol{\delta}_j = (\delta_{\beta j}, \delta_{\gamma j}, \delta_{vj})$. As for the generic asset j in the system ($1 < j \leq q$), we begin by separating the elements of the coefficient mean vector $\mathbf{m}_{j,t-1}$ according to whether they relate to the lagged predictor variables or the contemporaneous returns, i.e. $\mathbf{m}_{j,t-1} = (\mathbf{m}'_{\beta j,t-1}, \mathbf{m}'_{\gamma < j,t-1})'$.¹⁴ It then follows that

$$\mathbb{E}[r_{jt}|\boldsymbol{\delta}_j, \mathcal{D}_{t-1}] = \mathbf{x}'_{j,t-1} \mathbf{m}_{\beta j,t-1} + \mathbb{E}[\mathbf{r}_{< j,t}|\boldsymbol{\delta}_j, \mathcal{D}_{t-1}]' \mathbf{m}_{\gamma < j,t-1}, \quad (28)$$

$$Var[r_{jt}|\boldsymbol{\delta}_j, \mathcal{D}_{t-1}] = \frac{\hat{n}_{jt}}{\hat{n}_{jt} - 2} \left\{ tr \left(\hat{\mathbf{C}}_{\gamma < j,t} Cov[\mathbf{r}_{< j,t}|\boldsymbol{\delta}_j, \mathcal{D}_{t-1}] \right) + c_{jt} + s_{j,t-1} \right\} \\ + \mathbf{m}'_{\gamma < j,t-1} Cov[\mathbf{r}_{< j,t}|\boldsymbol{\delta}_j, \mathcal{D}_{t-1}] \mathbf{m}_{\gamma < j,t-1} \quad (29)$$

and

$$Cov[r_{jt}, \mathbf{r}_{< j,t}|\boldsymbol{\delta}_j, \mathcal{D}_{t-1}] = \mathbf{m}'_{\gamma < j,t-1} Cov[\mathbf{r}_{< j,t}|\boldsymbol{\delta}_j, \mathcal{D}_{t-1}] \quad (30)$$

where $\mathbb{E}[\mathbf{r}_{< j,t}|\boldsymbol{\delta}_j, \mathcal{D}_{t-1}]$ and $Cov[\mathbf{r}_{< j,t}|\boldsymbol{\delta}_j, \mathcal{D}_{t-1}]$ are known, $tr()$ stands for the trace of a matrix, and

$$c_{jt} = \left(\begin{matrix} \mathbf{x}_{j,t-1} \\ \mathbb{E}[\mathbf{r}_{< j,t}|\boldsymbol{\delta}_j, \mathcal{D}_{t-1}] \end{matrix} \right)' \hat{\mathbf{C}}_{jt} \left(\begin{matrix} \mathbf{x}_{j,t-1} \\ \mathbb{E}[\mathbf{r}_{< j,t}|\boldsymbol{\delta}_j, \mathcal{D}_{t-1}] \end{matrix} \right) \quad (31)$$

Applied iteratively, equations (27) through (30) yield the mean vector and covariance matrix of the predictive density of \mathbf{r}_t .

After observing the actual returns for time period t , we can update the prior for time t into the posterior for time t . We provide the details of these updating formulas in [Appendix](#)

¹⁴In particular, $\mathbf{m}_{\beta j,t-1}$ denotes the $p_j \times 1$ vector of coefficients for predictor variables $\mathbf{x}_{j,t-1}$, while $\mathbf{m}_{\gamma < j,t-1}$ is the $(j-1) \times 1$ vector of coefficients for the vector of contemporaneous returns $\mathbf{r}_{< j,t}$.

B, where we show how from the initial states (20), the sequence of regression coefficients and variance-covariance matrices can be obtained by a simple and very fast forward filter. Thus, the SG-DLM, like the W-DLM, deterministically gives the posterior distribution of the model parameters and the predictive densities of the q risky assets at each time step and avoids the need for computationally expensive Markov chain Monte Carlo methods.

2.3 Model Averaging

As we mentioned at the outset, both the W-DLM and the SG-DLM require the investor to know a priori the degree of time variation in the model parameters as well as the right combination of predictors to include in the model. In practice, the investor is unaware of what the optimal combination of these features may look like, and she is therefore facing significant uncertainty along these dimensions. To address this issue, we turn to model combinations.

For both the W-DLM and SG-DLM specifications, we estimate a different version of each model for every possible combination of predictor variables and discount factors. We defer the discussion of the predictors to the next section, where we will provide a detailed list of all the stock and bond predictors we consider in this study. As for the discount factors for the W-DLM, we consider values from two equally-spaced grids: $\delta_\beta \in \{0.98, 0.99, 1.0\}$ and $\delta_v \in \{0.95, 0.975, 1.0\}$.¹⁵ For the SG-DLM, we consider values from three equally-spaced grids, namely $\delta_\beta, \delta_\gamma \in \{0.98, 0.99, 1.0\}$, and $\delta_v \in \{0.95, 0.975, 1.0\}$. We have dropped the j subscript here to indicate that, in our empirical application, all assets in a particular SG-DLM will share the same discount factors.¹⁶

Next, at each point in time, we combine the forecast distributions obtained from all the permutations of predictors and discount factors. We do this separately for both the W-DLM and SG-DLM models. Note that, while we could have also chosen to combine the resulting

¹⁵While we could explore more of the model space by increasing the number of points used within these ranges, three values of each suffice to demonstrate the effects of model averaging and time-variation. We find no notable changes when increasing to ten values within each grid. Likewise, Dangi and Halling (2012) use $\delta_\beta \in \{0.96, 0.98, 1.00\}$, and find no notable changes by doubling the granularity to $\delta_\beta \in \{0.96, 0.97, 0.98, 0.99, 1.00\}$.

¹⁶As we mentioned before, the SG-DLM can allow each asset to have its own set of discount factors. However, for the specific empirical application considered in this paper we have found that a model with separate discount factors for each asset class does not outperform the simpler specification where the discount factors are constant across assets. Therefore, in what follows we will restrict our attention to the special case where $\delta_{\beta 1} = \dots = \delta_{\beta q}$, $\delta_{\gamma 1} = \dots = \delta_{\gamma q}$, and $\delta_{v 1} = \dots = \delta_{v q}$.

predictive densities across the two model specifications, we have elected to keep the two methods separated to better isolate the impact of the aforementioned W-DLM restrictions, and to empirically quantify the importance of relaxing such constraints with the SG-DLM approach. Also, in an attempt to slightly ease the notation, below we will use \mathcal{M}_i to denote the model with the i -th permutation of predictors and discount factors considered, where $i = 1, \dots, K_W$ in the case of the W-DLMs and $i = 1, \dots, K_{SG}$ in the case of the SG-DLMs, and K_W (K_{SG}) denotes the total number of model permutations we consider. We will then generally refer to the time t predictive mean and covariance matrix that come out of the i -th permutation of predictors and discount factors with $\mathbb{E}(\mathbf{r}_t|\mathcal{M}_i, \mathcal{D}_{t-1})$ and $Cov(\mathbf{r}_t|\mathcal{M}_i, \mathcal{D}_{t-1})$.

We explore two alternative combination schemes, as both have seen empirical success in the stock and bond predictability literatures. Our first combination scheme allows the weights on individual forecasting models to reflect their past predictive accuracy, and is therefore inspired by the optimal prediction pool approach of [Geweke and Amisano \(2011\)](#) and its good performance in settings similar to ours, as documented by [Pettenuzzo et al. \(2014\)](#) and [Gargano et al. \(2017\)](#). Specifically, at each point in time t , we compute model \mathcal{M}_i 's weight ($i = 1, \dots, K_W$ in the case of the W-DLMs and $i = 1, \dots, K_{SG}$ in the case of the SG-DLMs) by looking at its historical statistical performance up through time $t - 1$, as determined by the multivariate log score:

$$w_{i,t} \propto \sum_{\tau=1}^{t-1} \ln(S_{i,\tau}) \quad (32)$$

Here $S_{i,\tau}$ denotes the recursively computed score for model i at time τ , which we obtain by evaluating a Gaussian density with mean vector and covariance matrix equal to $\mathbb{E}(\mathbf{r}_\tau|\mathcal{M}_i, \mathcal{D}_{\tau-1})$ and $Cov(\mathbf{r}_\tau|\mathcal{M}_i, \mathcal{D}_{\tau-1})$ at the realized log excess returns \mathbf{r}_τ . This approach rewards the high-performing combinations of predictors and discount factors, assigning them more weight in the model combination. Our second combination scheme is the equal-weighted pool, which weight each of the K_W (or K_{SG}) models equally and has been shown by [Rapach et al. \(2010\)](#) to work well at least in the case of stock returns.

3 Data and Priors

3.1 Data

This section describes how we construct our portfolio of risky assets as well as which predictors we consider in our analysis.

3.1.1 Asset Returns

As for our pool of risky assets, we focus on a portfolio of monthly stock and bond returns, and, in particular, we consider: (i) the value-weighted return of the largest 20% of firms listed on the Center for Research in Security Prices’s database (CRSP); (ii) the value-weighted return of the CRSP firms in between the median and 80th percentile in size; (iii) the value-weighted return of the smallest 50% of CRSP firms; (iv) the five-year Treasury bond return; (v) the 10-year Treasury bond return. In addition, we collect data on the one-month Treasury bill rate (from Ibbotson Associates), which we use in our analysis to denote the returns of a risk-free investment strategy and to compute excess returns. All returns are continuously compounded, and the stock returns come from the CRSP’s monthly cap-based portfolios file.¹⁷ In contrast, monthly returns on five- and ten-year Treasury bonds are computed using the two-step procedure described in Gargano et al. (2017). In particular, in the first step we start from the daily yield curve parameter estimates of Gurkaynak et al. (2007) and use them to reconstruct the entire yield curve at the daily frequency. Next, focusing on the last day of each month’s estimated log yields, we combine the interpolated log yields to generate non-overlapping monthly bond returns for various maturities.¹⁸ Excess returns are obtained by subtracting the continuously compounded monthly T-bill rate from the previously computed asset returns.

¹⁷The T-bill rate comes from the research factors file, which is made available by Kenneth French at <http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/datalibrary.html>.

¹⁸Let n be the bond maturity in years. Time $t + 1$ holding period bond returns are given by the following formula

$$r_{t+1}^{(n)} = ny_t^{(n)} - (n - h/12)y_{t+1}^{(n-h/12)},$$

where $y_t^{(n)}$ is the log yield of the time t bond with n periods to maturity. To obtain five- and ten-year bond returns, we set $n = 60$ and $n = 120$ respectively.

3.1.2 Predictors

As for the predictors considered in this analysis, we start by including the equity predictors studied in [Welch and Goyal \(2008\)](#).¹⁹ These variables can be divided into three groups, namely stock, treasury, and corporate bond market variables. Stock market variables include the dividend-price ratio, dividend-payout ratio, stock variance, book-to-market ratio, and net equity expansion. Treasury market variables include the Treasury bill rate, long-term yield, term spread, and inflation rate. Finally, the default yield spread incorporates information from the corporate bond market. We augment this list of variables with the three predictors for bond returns considered by [Gargano et al. \(2017\)](#). Specifically, we consider forward spreads as proposed by [Fama and Bliss \(1987\)](#), a linear combination of forward rates as proposed by [Cochrane and Piazzesi \(2005\)](#), and a linear combination of macro factors, as proposed by [Ludvigson and Ng \(2009\)](#). We give both an equity and bond predictor to each to every DLM, which, for 15 stock predictors and three bond predictors, yields 45 different combinations. Hence, we specify 45 DLMs per combination of discount factors, yielding a total of 405 W-DLMs (due to $3^2 = 9$ different combinations of discount factors per each of the 45 predictor combinations) and 1215 SG-DLMs (from $3^3 = 27$ combinations of discount factors).²⁰

Once combined, our sample of monthly excess returns and predictors spans from January 1962 to December 2014, for a total of 636 observations (635 observations once we lag the predictor values). We provide summary statistics for both excess returns and predictors in [Table 1](#).

3.2 Initial States

As we described in [Section 2](#), we use \mathcal{D}_0 to denote the information set that we rely on to initialize the W-DLM and SG-DLM forward filters. We set aside the first 120 months of data to initialize/train our models, hence \mathcal{D}_0 in our case denotes the time period ranging from January 1962 to January 1972. We center the initial states for both the W- and SG-DLM

¹⁹We refer to [Welch and Goyal \(2008\)](#) for a detailed description of the construction of the individual predictors, which are available at <http://www.hec.unil.ch/agoyal/>.

²⁰To make the comparison across models easier, each equation in a predictive system/DLM will include the same two predictors, that is, one bond and one stock predictor, together for all assets. This will be true regardless of whether we work with the W- or SG-DLM methods.

specifications on the models' OLS estimates obtained over \mathcal{D}_0 . Specifically, in the W-DLM we set \mathbf{M}_0 , the conditional mean of the initial state in (4), to the coefficient estimates from an OLS multivariate predictive regression over the training dataset, and set \mathbf{S}_0 to the corresponding sample covariance matrix of the OLS residuals. Next, we specify $\mathbf{C}_0 = 100I_p$, which effectively renders the prior on the initial state \mathbf{B}_0 uninformative. Finally, we set the degrees of freedom n_0 to 10, therefore down-weighting the prior on $\mathbf{\Sigma}_0$ and rendering it flat and uninformative.

As for the SG-DLM, separately for each of the q equations in the system, we set the vectors \mathbf{m}_{j0} in (20) to the corresponding vectors of OLS estimates obtained over the training sample, while we set s_{j0} to the sample variance obtained from the OLS residuals ($j = 1, \dots, q$). Next, we let $\mathbf{C}_{j0} = 100s_{j0}I_{p_j+j-1}$, which renders the prior on \mathbf{C}_{j0} uninformative and also guarantees that the implied prior moments on the initial SG-DLM regression parameters are equivalent to those from the W-DLM. Lastly, as with the W-DLM, we set the degrees of freedom n_{j0} to 10, effectively making the prior on σ_{j0}^{-2} flat and uninformative.

4 Empirical Results

In this section, we describe our empirical results. We will begin with an investigation of the role played by the various key features of our approach, with a particular emphasis on the importance of time variation in the first and second moments of asset returns and the strength of the predictability stemming from the various regressors we consider. Next, we will turn to examining the quality and accuracy of the W-DLM and SG-DLM forecasts, with an eye towards both statistical and economic measures of predictability. More specifically, we will evaluate the forecast accuracy of these models over the last 360 months of data in our sample, January 1985 through December 2014. In this way, we explicitly remove from the forecast evaluation sample the period of time characterized by the oil shocks of 1973-1974 and the bond market experimentation of the early 1980's.

4.1 A close look at the role of the various modeling features

As we discussed in Subsection 2.3, one of the key advantages of our approach is the ability to take into account the model uncertainty arising from both the availability of different predictor variables and the presence of multiple discount factors controlling the time variation

in regression coefficients, variances, and covariances. Thanks to the closed-form nature of the forward filters we rely on, this can be accomplished in a very timely manner, without the need to resort on expensive MCMC simulations. In this section, we take a close look at the role of predictor uncertainty and time-variation in both the W-DLM and SG-DLM models.

In order to disentangle the relative importance of these features, for both the W-DLM and SG-DLM models we compute four variations of the score-weighted and equal-weighted model combinations described in [Subsection 2.3](#). Our first model combination, which we label LIN, constrains $\delta_\beta = \delta_v = 1$ for the W-DLMs and $\delta_{\beta j} = \delta_{vj} = \delta_{\gamma j} = 1$ for the SG-DLMs ($j = 1, \dots, q$), thus completely removing time variation in both the regression coefficients and variance-covariance matrix. In other words, the LIN specifications control for the uncertainty arising solely from the choice of which predictors to include in the model. Our second variant, which we denote as TVP to reference its time-varying parameters, is obtained by selectively combining only the subset of W-DLMs or SG-DLMs with $\delta_v = 1$ (as well as, in the case of the SG-DLMs, $\delta_\gamma = 1$) and $\delta_\beta < 1$. In this case, we are focusing on all those models with a constant variance covariance matrix, taking into account the uncertainty pertaining to the choice of which predictors to include and how much the regression coefficients should be allowed to vary over time.²¹ Our third variant, which we denote as SV to reference stochastic volatility, is similarly obtained by selectively combining only the subset of models with $\delta_\beta = 1$ and $\delta_v < 1$ (as well as, in the case of the SG-DLMs, either $\delta_v < 1$ or $\delta_\gamma < 1$). Thus, we are removing altogether time variation in the regression coefficients, while controlling for the uncertainty arising from which predictors to include and how much time variation to afford to variances and covariances. Lastly, our fourth model combination variant, which we denote with TVP-SV, is obtained by combining all the W-DLMs or SG-DLMs that set $\delta_\beta < 1$ and $\delta_v < 1$ (as well as, in the case of the SG-DLMs, $\delta_\gamma < 1$). This is therefore our most flexible model combination scheme, where the regression coefficients, variances, and covariances all vary over time.

We start by looking at the importance of time variation in the first and second moments of stock and bond returns. The score-based model combination weights in [\(32\)](#) are a combination

²¹In this sense, we can think of the TVP variant of our model combinations as the multivariate extension of the approach first proposed by [Dangl and Halling \(2012\)](#) to forecast stock returns.

of accuracy forecasting the first and second moments. To demonstrate how these weights change over time, instead of plotting curves for all 405 W-DLMs and 1215 SG-DLMs (as there are 45 different pairs of predictors matched with 9 and 27 different combinations of discount factors for the W-DLMs and SG-DLMs respectively), we summarize by plotting the percent of the total aggregate weight that a typical model from one of our model combinations receives. This is simply calculated as the percent of total model weight assigned to each one of the four groups (LIN/TVP/SV/TVP-SV) and dividing it by the total number of models within that group. [Figure 1](#) shows the evolution over time of these weights for our four model combination variants of the W-DLMs (top panel) and SG-DLMs (bottom panel). We also report, in the legend of both panels, the total number of models within each group. As it can be seen from both panels, the assumption of constant variances/covariances appears to be strongly rejected by the data, with the average model combination weights of the models belonging to the LIN and TVP only receiving marginal support by the data. In contrast, allowing for variation in the variances and covariances produces much larger model weights, with the models within the SV and TVP-SV model combinations receiving, on average, weights that are two to three times larger.

Next, [Figure 2](#) and [Figure 3](#) plot the time series of asset volatilities and cross-correlations associated with the score-weighted LIN, TVP, SV, and TVP-SV SG-DLM model specifications.²² For comparability, we have also included in each panel the (constant) sample volatility or correlation, computed over the 1972–2014 period and depicted with a thin black line. For all five assets, we see that the conditional volatilities of all five assets for the SV and TVP-SV models vary significantly over time, with long spells of time characterized by above-average volatility (see in particular the late 1980’s, early 2000’s, and the most recent financial crisis) as well as shorter periods with below-average volatilities. We can also recognize a marked difference between the pattern of time-variation of the three equity returns and those of the two bond returns, and strong similarities in the volatilities of the 5- and 10-year bond returns, which are seen in the different magnitudes of the vertical axes. As for the correlations, we see long stretches of time with conditional correlations that are either above or below their average counterparts. Again, we observe different patterns of

²²We provide similar plots for the four W-DLM model combination variants in [Appendix C](#).

time-variation in the three equity returns from those of the two bond returns.

We conclude this section with a look at the relative importance of the predictor variables we consider in this study. [Figure 4](#) depicts the evolution over time of the score-based model combination weights in [\(32\)](#) for both the W- and SG-DLM over the evaluation period, January 1985 through December 2014. More specifically, the left panels of the figure focus on the equity predictors from [Welch and Goyal \(2008\)](#), and the right panels of the figure repeat the same calculations for the three bond predictors from [Gargano et al. \(2017\)](#). The construction of these curves is the same as in [Figure 1](#), but instead of combining into different feature sets, weights are combined based on the predictors used. Starting with the left panels of the figure, we observe that among all the equity predictors, the stock variance, default yield spread, and default return spread take turns having the largest weight in the model combination, and are always among the most important variables. Conversely, the earning/price ratio predictor appears to not be favored in the model combination, consistently scoring among the lowest weights. Moving on to the right panel of the figure and the bond predictors, we find that the Cochrane-Piazzesi factor and the Fama-Bliss spreads receive the highest weights in the model combination, while surprisingly the Ludvigson-Ng macro factor appears to be less favored, at least in terms of log-scores. This result, which at first appears to be contradicting the results in [Gargano et al. \(2017\)](#), is due to the fact that our combination weights are driven by the predictors' relative log-scores, and as [Gargano et al. \(2017\)](#) show (see their Figures 6 and 7) the advantage of the Ludvigson-Ng macro factor is particularly apparent when focusing on point predictability. However, it is worthwhile pointing out that the performance gap between the three predictors shrinks over time. The general patterns in the score-weights over time for W- and SG-DLMs are largely similar, showing that the relative importance of the variables holds regardless of DLM type.

4.2 Out-of-Sample Performance

We now turn to evaluating the relative predictive accuracy of the various W-DLM and SG-DLM specifications over the period spanning from January 1985 to December 2014. Throughout, our benchmark model will be a no-predictability SG-DLM with constant mean and constant variance-covariance matrix (that is, the specification in [\(13\)](#) with $\mathbf{x}_{j,t-1} = 1$

and $\delta_{\beta_j} = \delta_{\gamma_j} = \delta_{v_j} = 1$, for all j), in line with what is customary in both the stock and bond return predictability literatures. We will provide results separately for each of the five risky assets that we are focusing on, as well jointly for the whole system of equations.

4.2.1 Measures of Predictive Accuracy

Starting with the point forecast accuracy, for each of the five asset returns we summarize the precision of the point forecasts for model i , relative to that from the benchmark model, by means of the ratio of mean-squared forecast errors (“MSFEs”):

$$MSFE_{ij} = \frac{\sum_{\tau=\underline{t}}^T e_{ij,\tau}^2}{\sum_{\tau=\underline{t}}^T e_{bcmk,j,\tau}^2} \quad (33)$$

where \underline{t} denotes the beginning of the out-of-sample period, i refers to the W-DLM or SG-DLM model under consideration (i.e. LIN, TVP, SV, TVP-SV), $e_{ij,\tau} = r_{j\tau} - \mathbb{E}(r_{j\tau} | \mathcal{M}_i, \mathcal{D}_{\tau-1})$ and $e_{bcmk,j,\tau} = r_{j\tau} - \mathbb{E}(r_{j\tau} | \mathcal{M}_{bcmk}, \mathcal{D}_{\tau-1})$ are the forecast errors of asset return j at time τ associated with model i and the benchmark model, respectively. Values of $MSFE_{ij}$ below one suggest that for asset j , model i produces more accurate point forecasts than the no-predictability benchmark.

We also measure the point-forecast accuracy of the various method by looking jointly at all five assets. Following [Christoffersen and Diebold \(1998\)](#), we compute the ratio between the weighted multivariate mean squared forecast error (WMSFE, also known as the squared Mahalanobis distance) of model i and the no-predictability benchmark as follows:

$$WMSFE_i = \frac{\sum_{\tau=\underline{t}}^T \mathbf{e}'_{i\tau} \left[\widehat{Cov}(\mathbf{r}_t) \right]^{-1} \mathbf{e}_{i\tau}}{\sum_{\tau=\underline{t}}^T \mathbf{e}'_{bcmk,\tau} \left[\widehat{Cov}(\mathbf{r}_t) \right]^{-1} \mathbf{e}_{bcmk,\tau}} \quad (34)$$

where $\mathbf{e}_{i\tau} = (e_{i1,\tau}, \dots, e_{iq,\tau})'$ and $\mathbf{e}_{bcmk,\tau} = (e_{bcmk,1,\tau}, \dots, e_{bcmk,q,\tau})'$ are the $q \times 1$ vectors of forecast errors at time τ associated with model i and the benchmark model, while $\widehat{Cov}(\mathbf{r}_t)$ denotes the sample estimates of the asset returns unconditional variance-covariance matrix, computed over the evaluation period.²³

²³The role of this covariance matrix is to standardize the distances in multivariate space, and weight the assets' forecast errors differently depending on the variability of the underlying assets and correlation across assets. All things equal, there will be less penalty for the forecast errors of a highly volatile asset than from those of a less-volatile asset, but also less reward when accurate. At the same time, there will be more penalty for forecast errors in directions not implied by the correlations in the empirical sample covariance matrix, meaning higher penalties for forecast errors in opposite directions for correlated assets and high penalties for forecast errors in the same direction for negatively correlated assets.

As for the quality of the density forecasts, we compute the average log score (ALS) differential between model i and the no-predictability benchmark,

$$ALS_{ij} = \frac{1}{T - \underline{t} + 1} \sum_{\tau=\underline{t}}^T (\ln(S_{ij,\tau}) - \ln(S_{bcmk,j,\tau})) \quad (35)$$

where $S_{ij,\tau}$ ($S_{bcmk,j,\tau}$) denotes model i 's (benchmark's) log score at time τ , which we obtain by evaluating a univariate Gaussian density with mean vector $\mathbb{E}(\mathbf{r}_{j\tau}|\mathcal{M}_i, \mathcal{D}_{\tau-1})$ ($\mathbb{E}(\mathbf{r}_{j\tau}|\mathcal{M}_{bcmk}, \mathcal{D}_{\tau-1})$) and variance $Var(\mathbf{r}_{j\tau}|\mathcal{M}_i, \mathcal{D}_{\tau-1})$ ($Var(\mathbf{r}_{j\tau}|\mathcal{M}_{bcmk}, \mathcal{D}_{\tau-1})$) at the realized excess returns $\mathbf{r}_{j\tau}$. Positive values of ALS_{ij} indicate that model i produces on average more accurate density forecasts for variable j than the benchmark. Finally, we consider the multivariate average log score differentials (MVALS) between model i and the benchmark,

$$MVALS_i = \frac{1}{T - \underline{t} + 1} \sum_{\tau=\underline{t}}^T (\ln(S_{i,\tau}) - \ln(S_{bcmk,\tau})) \quad (36)$$

where $S_{i,\tau}$ ($S_{bcmk,\tau}$) are computed as described in [Subsection 2.3](#).²⁴

4.2.2 Results

We begin by inspecting the point and density forecast predictability of both W-DLM and SG-DLM models on an asset-by-asset basis, as summarized by the *MSFE* and *ALS* metrics. [Table 2](#) reports the *MSFE* ratios of the LIN, SV, TVP, and TVP-SV variants of the W-DLM and SG-DLM models, individually for the five asset returns and relative to the no-predictability benchmark. Across the columns of the table, we report the average predictive improvements obtained by either relying on the equal-weighted or score-weighted combinations. As for the three equity returns, we see that stochastic volatility plays a small role in improving the SG-DLM predictions, while time-varying parameters hurt both W- and SG-DLMs. In fact, the TVP and TVP-SV specifications do worse at point prediction than the benchmark. Results for the two bond returns are stronger, with ample and widespread evidence of point-predictability. This appears to be true regardless of the combination scheme and the set of features considered, though time-varying parameters are

²⁴This measure penalizes wrong return predictions based on the variance of the prediction. If the model is highly confident in an inaccurate prediction, it scores very low. If highly confident and correct, it receives a high score. If the model is unconfident in the prediction, and hence has high variance and a relatively flat pdf, then there is little penalty for being wrong but also little bonus for being correct.

again never preferable. This widespread predictability is consistent with the findings of both Thornton and Valente (2012) and Gargano et al. (2017). Next, Table 3 inspects the asset-specific density forecast predictability of the same models depicted in Table 2. Here, in line with the results reported in Figure 1, we find that for all five assets’ SG-DLMs, the SV and TVP-SV model combinations always lead to the most accurate predictive densities. However, we see this does not hold for W-DLMs on mid- and small-cap stocks, where LIN is best. Furthermore, the mid- and small-cap stocks see large drops in performance when adding time-varying parameters, which benefit 10-Year bonds, across models and weighting schemes.

Next, we turn to the joint point and density forecast predictability, as measured by the *WMSFE* and *MVALS* metrics. Starting with Table 4, we find that in term of point forecast accuracy the best performing models feature constant coefficients, and adding time-varying parameters to the model set appear to slightly increase the forecasting error, across the board. Moving on to the joint accuracy of the density forecasts, Table 5 reports the average (multivariate) log score improvements that are brought in by the different sets of feature and model combination schemes. The SV and TVP-SV feature sets lead to the largest gains in accuracy, with the largest gains being associated with the SG-DLM TVP-SV model. Interestingly, the comparison of log scores between W-DLM and SG-DLM seem to favor the latter model specification when volatilities and correlations are stochastic, while when constant, the DLM types are comparable. To shed further light on where the SV and TVP-SV model are most successful, Figure 5 plots the cumulative sum of the multivariate log score differentials, $CSMVLS D_{it} = \sum_{\tau=\underline{t}}^t (\ln(S_{i,\tau}) - \ln(S_{bcmk,\tau}))$ for SG-DLMs with different feature sets. The figure clearly shows how, starting around the mid 1990’s and continuing all the way to the end of the sample, the SV and TVP-SV models consistently generate significantly more accurate density forecasts than all the alternative model specifications. Furthermore, we note that TVP-SV gains a step over SV during the housing bubble, which step persists through the end of the sample.

The previous tables and figures indicate that time-varying volatility and correlation play a very important role in generating sharp density forecasts. This appears to be true both

at the individual asset level as well as when focusing on all stock and bond returns jointly. To offer additional insights on the mechanics behind this result, [Figure 6](#) shows the heat-map of the joint density forecast accuracy, as measured by the *MVALS* metric, for all the possible combinations of discount factors considered, both for the W-DLMs (top left panel) and the SG-DLMs (remaining panels).²⁵ This plot permits to pinpoint exactly the mix of features that lead to the largest predictive gains in our model combinations, and to provide further clarity on the drivers of the results summarized in [Tables 2](#) through [5](#). Starting with the W-DLM case, we observe that the most successful models feature a large degree of variation across the board, in volatilities, correlations, and regression coefficients. This demonstrates why there are high score values for the TVP-SV model in [Table 5](#), as it is the only feature set incorporating models from this lower left quadrant. The remaining three panels show that the best performing SG-DLM models include high degrees of time variation in the volatilities, moderate degrees of time-variation in correlations, and ambivalence toward the degree of movement in the regression coefficients (compared to the W-DLMs). Note specifically, however, that the ideal degree of time-variation in the correlations ($\delta_\gamma \approx 0.986$) is much higher than that of volatility ($\delta_v \leq 0.95$). This further validates the importance of allowing for a different degree of variation in the on- and off-diagonal elements of the variance covariance matrix.

4.3 Portfolio Analysis

We now turn to evaluating the portfolio implications and economic predictability implied by the W-DLM and SG-DLM predictive densities that we derived.

4.3.1 Framework

We focus on the problem of a Bayesian investor endowed with power utility (see, among others, [Johannes et al., 2014](#); [Gargano et al., 2017](#); [Gao and Nardari, 2018](#)). At each point in time, the investor chooses her optimal asset allocations by distributing her total wealth between q (equity and bond) risky assets and one risk-free asset, under the constraint that the

²⁵In particular, each point in the heat-maps corresponds to the average *MVALS* associated to a given combination of discount factors, averaged over all 45 permutations of predictors (in the case of the SG-DLMs, also averaged over all possible values of the discount factor not shown on the axes).

sum of her long and short positions does not exceed 300% of her wealth or fall below -200% of her wealth, and that none of her individual positions (including the weight on the risk-free asset) falls outside the same range.²⁶ Assuming that the excess returns on the q risky assets are jointly log-normally distributed, we can follow [Campbell et al. \(2003\)](#) and approximate $r_{p,it}$, the log return of the portfolio implied by model i at time t , with the following formula

$$r_{p,it} = r_{f,t-1} + \boldsymbol{\omega}'_{i,t-1}(\mathbf{r}_t - r_{f,t-1}\mathbf{1}) + \frac{1}{2}\boldsymbol{\omega}'_{i,t-1}\text{diag}(\widehat{\boldsymbol{\Sigma}}_{i,t|t-1}) - \frac{1}{2}\boldsymbol{\omega}'_{i,t-1}\widehat{\boldsymbol{\Sigma}}_{i,t|t-1}\boldsymbol{\omega}_{i,t-1} \quad (37)$$

where $\boldsymbol{\omega}_{i,t-1}$ is a vector of portfolio weights, $\widehat{\boldsymbol{\Sigma}}_{i,t|t-1} = \text{Cov}(\mathbf{r}_t | \mathcal{M}_i, \mathcal{D}_{t-1})$ denotes the risky assets' forecasted variance-covariance matrix at time t based on the estimates given by model \mathcal{M}_i and conditional on the information set at time $t-1$, $r_{f,t-1}$ represents the continuously compounded risk-free rate, and $\mathbf{1}$ is a vector of ones the same length as \mathbf{r}_t . Let A denote the investor's relative degree of risk aversion, then the optimal weights on the q risky assets implied by model i are given by the solution of the following constrained maximization problem,

$$\begin{aligned} \arg \max_{\boldsymbol{\omega}_{i,t-1}} \quad & \boldsymbol{\omega}'_{i,t-1} \left(\widehat{\boldsymbol{\mu}}_{i,t|t-1} + \frac{1}{2}\text{diag}(\widehat{\boldsymbol{\Sigma}}_{i,t|t-1}) \right) - \frac{A}{2}\boldsymbol{\omega}'_{i,t-1}\widehat{\boldsymbol{\Sigma}}_{i,t|t-1}\boldsymbol{\omega}_{i,t-1} \\ \text{s.t.} \quad & \boldsymbol{\omega}'_{i,t-1}\mathbf{1} \in [-2, 3] \\ & \boldsymbol{\omega}_{i,j,t-1} \in [-2, 3], \quad j = 1, \dots, q \end{aligned} \quad (38)$$

where $\widehat{\boldsymbol{\mu}}_{i,t|t-1} = \mathbb{E}(\mathbf{r}_t | \mathcal{M}_i, \mathcal{D}_{t-1})$ is the mean of the predictive density of the vector of risky assets \mathbf{r}_t , computed using the information set available at time $t-1$ and under model \mathcal{M}_i .

We next use the sequence of portfolio weights $\{\boldsymbol{\omega}_{i,t-1}\}_{t=\underline{t}}^T$ computed under the various W- and SG-DLM models as well as under the benchmark model to compute the investor's certainty equivalent returns (CERs), which can be further expressed in percentage annualized terms as:

$$CER_i = 100 \times \left(\left[\frac{1}{T - \underline{t} + 1} \sum_{t=\underline{t}}^T \widehat{W}_{it}^{1-A} \right]^{\frac{12}{1-A}} - 1 \right) \quad (39)$$

where $\widehat{W}_{it} = \exp(r_{p,it})$ denotes the realized wealth at time t , as implied by model i .

4.3.2 Results

Table 6 presents the annualized CERs over the whole out-of-sample period for the various W-DLM and SG-DLM model combinations for an investor with power utility and coefficient of

²⁶Similarly, [Gao and Nardari \(2018\)](#) consider an investor who is constrained and will not be allowed to short risky assets and/or borrow more than a certain amount of cash.

relative risk aversion $A = 5$. In particular, the table reports the CER gains relative to the no-predictability benchmark model, i.e. $CER_i - CER_{bcmk}$ (as a reference point, the annualized CER for the benchmark model over the same period is equal to 5.896%.) As it can be inferred from the table, all feature sets and all model averaging weighting schemes produce higher CERs than the no-predictability benchmark. This is true regardless of whether we focus on the W- or SG-DLM models. In addition, we make the following observations. First, as it was the case with the log score measures, the largest gains in CERs occur when volatility is allowed to vary over time. Across the board, the inclusion of stochastic volatilities and correlations always lead to larger CERs, and this is true whether or not one also allows for time-variation in the regression coefficients. That is, SV produces higher CERs than LIN, and so does TVP-SV compared to TVP. In particular, we find that the SV model combination of SG-DLMs produces CER gains of about 5.9% over the benchmark. The role of SV in the case of the W-DLM models is slightly less pronounced, with the improvements ranging from 4.6% to 5.4%. We attribute this difference to the additional restrictions that the W-DLM models impose, most notably the requirement that a single discount factor must control the degree of time-variation in both variances and covariances. Second, the inclusion of time-varying coefficients in the model always decreases CERs. Third, equal-weighting holds a slight edge over the comparable score-weighted models, except for TVP-SV SG-DLMs. While the differences between the two weighting schemes are only marginal, this result goes in the opposite direction of what we found in [Table 5](#) when focusing on the log score differentials, and is likely due to the way we computed the model combination weights in [\(32\)](#). Fourth, SG-DLMs improve over the comparable W-DLMs, except for the equal-weighting LIN specification. Again, we believe this result is particularly due to the added flexibility that the SG-DLMs bring to the SV dynamics, and this is consistent with the largest differences between W- and SG-DLMs occurring when the SV feature is include in the model set.

As with our investigation of the joint statistical predictability of the W- and SG-DLM models, we conclude this section with a closer look at how exactly the various model features help achieve large CERs. In particular, [Figure 7](#) shows the heat-map of the CER gains associated with all the possible combinations of discount factors considered, both for the W-DLMs (top left panel) and the SG-DLMs (remaining panels). Starting with the W-DLM

case, we observe that there are two regions of highly profitable models. The first features a modest degree of time-variation in the volatilities and correlations and very little variation in the regression coefficients. The second features no time-variation in the regression coefficients to go along with high degrees of time-variation in the covariance matrix elements. As for the more flexible SG-DLMs, the remaining three panels show that the best performing models also feature minimal (but some) variation in the regression coefficients, combined with as much time-variation in the volatilities and correlations as we will allow. While the differences between W- and SG-DLM’s optimal combinations for CER performance are not as stark as those for score, we do see that the magnitude of the SG-DLM CER noticeably exceed those of the W-DLM.

5 Conclusion

In this paper, we build on the Wishart dynamic linear model (W-DLM) of [West and Harrison \(1997\)](#) and the simultaneous graphical dynamic linear model (SG-DLM) of [Gruber and West \(2016\)](#) to introduce a flexible approach to model and forecast multiple asset returns. This approach allows us to integrate a number of useful features into a predictive system, namely model and parameter uncertainty, time-varying parameters, stochastic volatility, and time varying covariances. We combine these DLM methods with a fully automated data-based model-averaging procedure to objectively determine the optimal set of said features and employ it to jointly forecast monthly stock and bond excess returns. This is made possible by the computational speed of the DLMs.

When evaluated over the January 1985 – December 2014 period, we find large statistical and economic benefits from using the appropriate ensemble of features in predicting stock and bond returns. In particular, we find that W-DLMs and SG-DLMs with stochastic volatility and time-varying covariances bring the largest gains in terms of statistical predictability, and that time-varying parameters can enhance the ensemble when forecasting distributions, though not for point predictions. Lastly, SG-DLM models with predictors, stochastic volatility and time-varying correlations lead to the largest economic gains. We show that when using this optimal set of features, a leverage-constrained power utility investor earns over 500 basis

points (on an annualized basis) more than if she relied on the no-predictability benchmark.

References

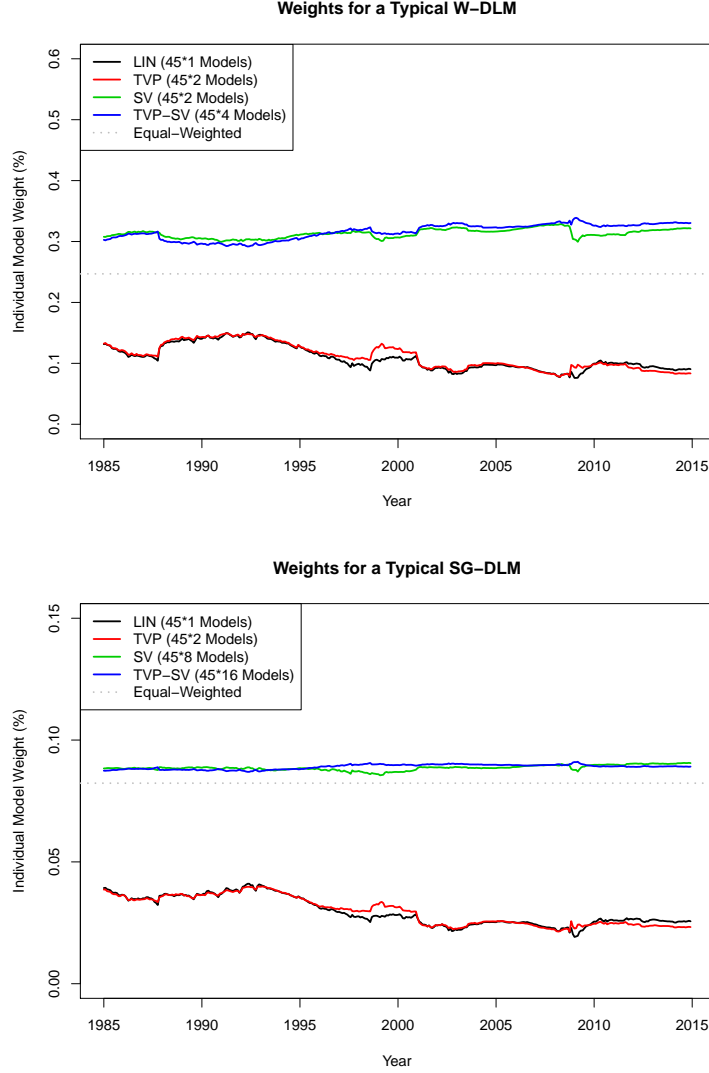
- Ang, A. and Bekaert, G. (2007). Stock return predictability: Is it there? *Review of Financial Studies*, 20(3):651–707.
- Barnard, J., McCulloch, R., and Meng, X.-L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, 10(4):1281–1311.
- Bossaerts, P. and Hillion, P. (1999). Implementing statistical criteria to select return forecasting models: what do we learn? *Review of Financial Studies*, 12(2):405–428.
- Brennan, M. J., Schwartz, E. S., and Lagnado, R. (1997). Strategic asset allocation. *Journal of Economic Dynamics and Control*, 21(8):1377 – 1403.
- Campbell, J. and Shiller, R. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1(3):195–228.
- Campbell, J. Y., Chan, Y. L., and Viceira, L. M. (2003). A multivariate model of strategic asset allocation. *Journal of Financial Economics*, 67(1):41 – 80.
- Carriero, A., Clark, T. E., and Marcellino, M. (2016). Large vector autoregressions with stochastic volatility and flexible priors. *Federal Reserve Bank of Cleveland Working Paper*.
- Christoffersen, P. F. and Diebold, F. X. (1998). Cointegration and long-horizon forecasting. *Journal of Business & Economic Statistics*, 16(4):450–458.
- Cochrane, J. H. and Piazzesi, M. (2005). Bond risk premia. *The American Economic Review*, 95(1):138–160.
- Dangl, T. and Halling, M. (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics*, 106(1):157 – 181.
- Dawid, A. P. (1981). Some matrix-variate distribution theory: Notational considerations and a bayesian application. *Biometrika*, 68(1):265–274.

- Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350.
- Fama, E. F. and Bliss, R. R. (1987). The information in long-maturity forward rates. *The American Economic Review*, 77(4):680–692.
- Fama, E. F. and Schwert, G. (1977). Asset returns and inflation. *Journal of Financial Economics*, 5(2):115 – 146.
- Gao, X. and Nardari, F. (2018). Do commodities add economic value in asset allocation? New evidence from time-varying moments. *Journal of Financial and Quantitative Analysis*, 53.
- Gargano, A., Pettenuzzo, D., and Timmermann, A. G. (2017). Bond return predictability: Economic value and links to the macroeconomy. *Management Science*, forthcoming.
- Gelman, A. and Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Analytical Methods for Social Research. Cambridge University Press.
- Geweke, J. and Amisano, G. (2011). Optimal prediction pools. *Journal of Econometrics*, 164(1):130 – 141.
- Gruber, L. and West, M. (2016). Gpu-accelerated bayesian learning and forecasting in simultaneous graphical dynamic linear models. *Bayesian Analysis*, 11(1):125–149.
- Gurkaynak, R. S., Sack, B., and Wright, J. H. (2007). The u.s. treasury yield curve: 1961 to the present. *Journal of Monetary Economics*, 54(8):2291 – 2304.
- Johannes, M., Korteweg, A., and Polson, N. (2014). Sequential learning, predictability, and optimal portfolio returns. *The Journal of Finance*, 69(2):611–644.
- Kim, C.-J., Morley, J. C., and Nelson, C. R. (2005). The structural break in the equity premium. *Journal of Business & Economic Statistics*, 23(2):181–191.
- Koop, G., Korobilis, D., and Pettenuzzo, D. (2017). Bayesian compressed vector autoregressions. *Journal of Econometrics*, page forthcoming. Working paper.

- Lettau, M. and Ludvigson, S. (2001). Consumption, aggregate wealth, and expected stock returns. *The Journal of Finance*, 56(3):815–849.
- Lettau, M. and Van Nieuwerburgh, S. (2008). Reconciling the return predictability evidence. *The Review of Financial Studies*, 21(4):1607–1652.
- Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics*, 74(2):209 – 235.
- Ludvigson, S. C. and Ng, S. (2009). Macro factors in bond risk premia. *The Review of Financial Studies*, 22(12):5027–5067.
- Pastor, L. and Stambaugh, R. F. (2001). The equity premium and structural breaks. *The Journal of Finance*, 56(4):1207–1239.
- Paye, B. S. and Timmermann, A. (2006). Instability of return prediction models. *Journal of Empirical Finance*, 13(3):274–315.
- Pettenuzzo, D. and Timmermann, A. (2011). Predictability of stock returns and asset allocation under structural breaks. *Journal of Econometrics*, 164(1):60 – 78. Annals Issue on Forecasting.
- Pettenuzzo, D., Timmermann, A., and Valkanov, R. (2014). Forecasting stock returns under economic constraints. *Journal of Financial Economics*, 114(3):517 – 553.
- Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, 72(3):821–852.
- Rapach, D. E., Strauss, J. K., and Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *The Review of Financial Studies*, 23(2):821–862.
- Thornton, D. L. and Valente, G. (2012). Out-of-sample predictions of bond excess returns and forward rates: An asset allocation perspective. *Review of Financial Studies*, 25(10):3141 – 3168.

- Viceira, L. (1997). Testing for structural change in the predictability of asset returns. *Unpublished*.
- Wachter, J. A. and Warusawitharana, M. (2009). Predictable returns and asset allocation: Should a skeptical investor time the market? *Journal of Econometrics*, 148(2):162 – 178.
- Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 21(4):1455–1508.
- West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*. Springer Series in Statistics. Springer, 2nd edition.
- Zhao, Z. Y., Xie, M., and West, M. (2016). Dynamic dependence networks: Financial time series forecasting and portfolio decisions. *Applied Stochastic Models in Business and Industry*, 32(3):311–332.

Figure 1. Time series of score-based weights by feature set

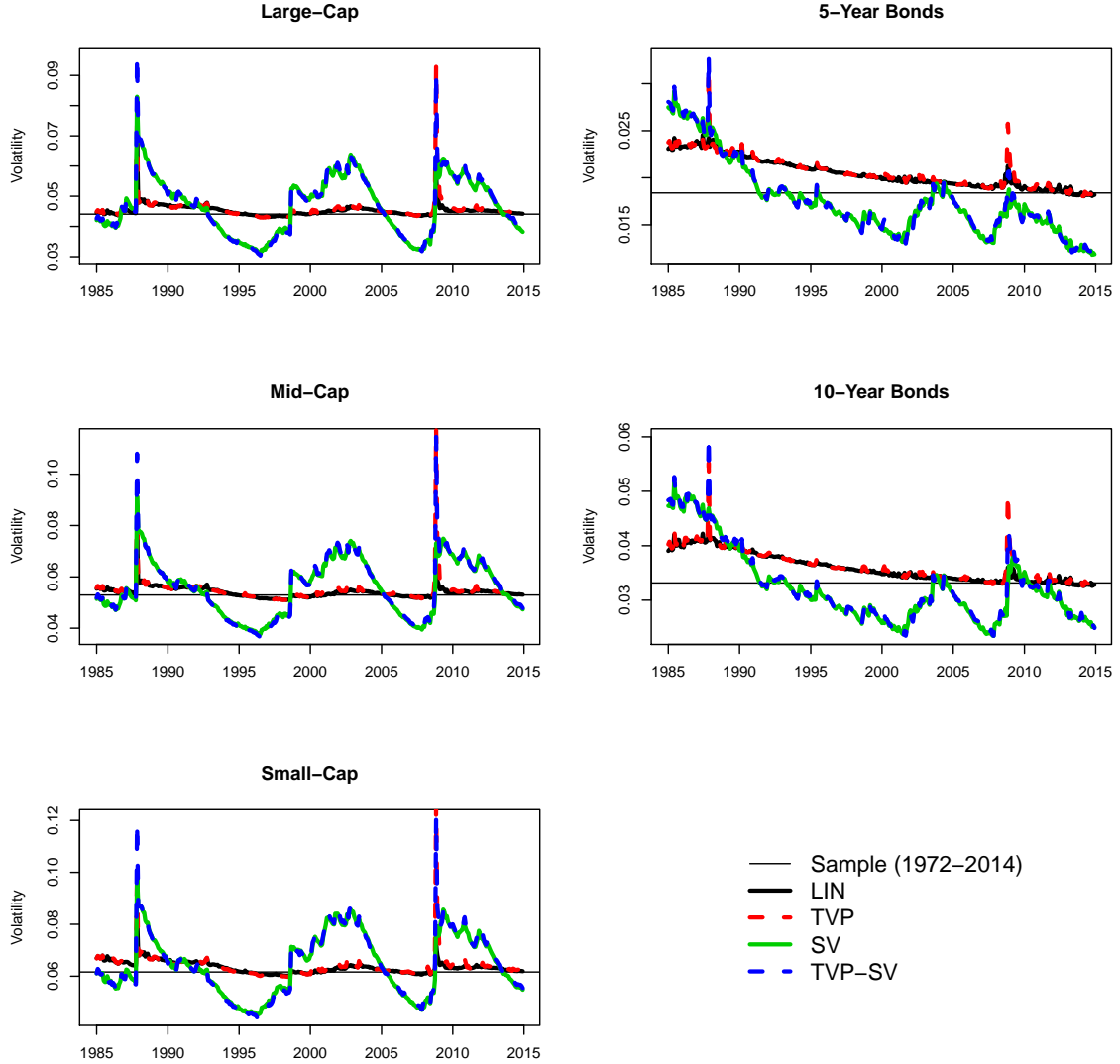


This figure shows the evolution over time of the score-based model combination weights for the four variants of the W-DLM (top panel) and SG-DLM (bottom panel) models, namely LIN, TVP, SV, and TVP-SV. At each point in time t , we compute model \mathcal{M}_i 's weight ($i = 1, \dots, K_W$ in the case of the W-DLM models and $i = 1, \dots, K_{SG}$ in the case of the SG-DLM models) by looking at its historical statistical performance up through time $t - 1$, as determined by the log score:

$$w_{i,t} \propto \sum_{\tau=1}^{t-1} \ln(S_{i,\tau})$$

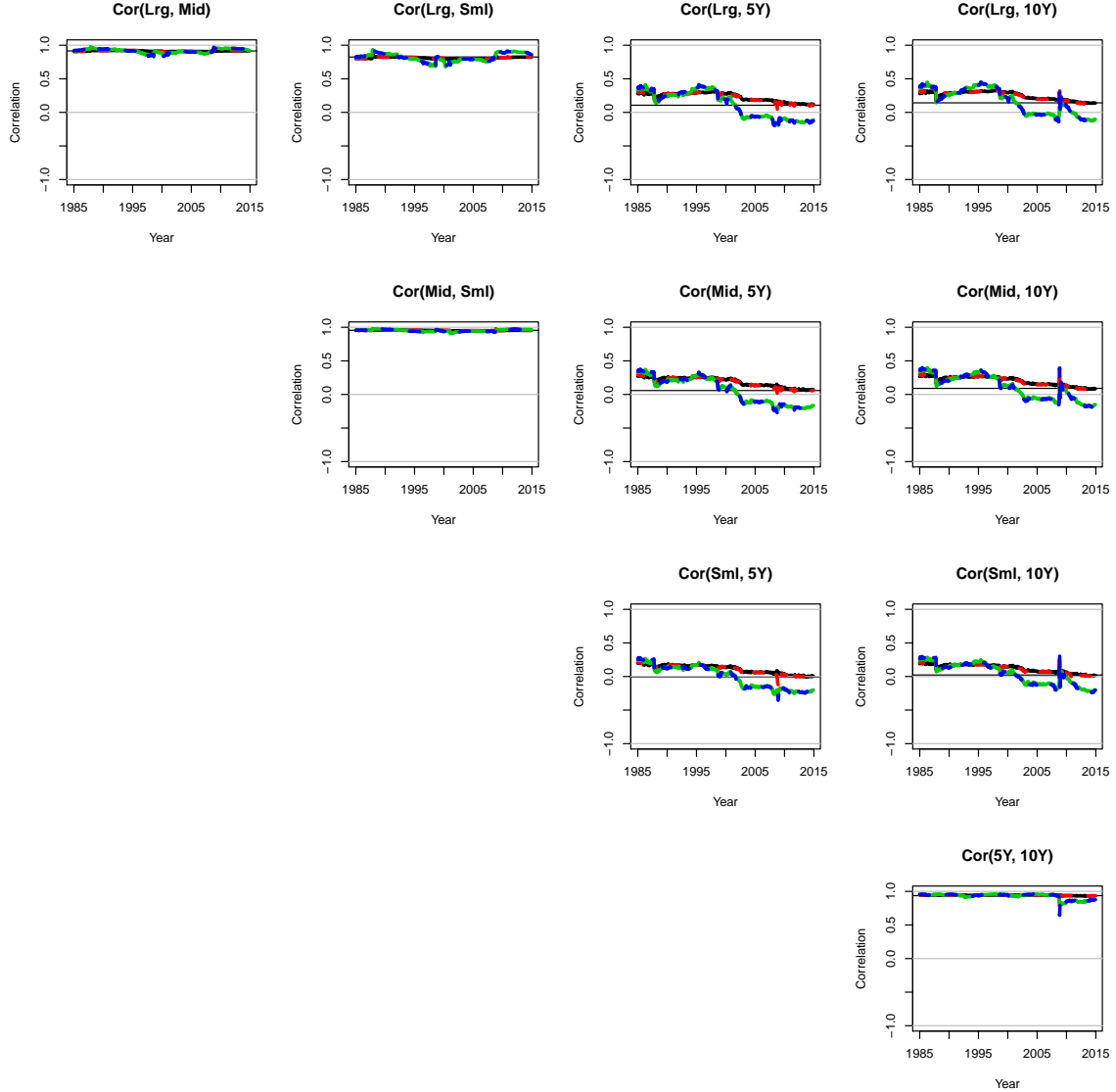
where $S_{i,\tau}$ denotes the recursively computed score for model i at time τ , which we obtain by evaluating a Gaussian density with mean vector and covariance matrix equal to $\mathbb{E}(\mathbf{r}_\tau | \mathcal{M}_i, \mathcal{D}_{\tau-1})$ and $Cov(\mathbf{r}_\tau | \mathcal{M}_i, \mathcal{D}_{\tau-1})$ at the realized log excess returns \mathbf{r}_τ . Next, we normalize the model weights across all models such that $\sum_{i=1}^{K_W} w_{i,t} = 1$ for W-DLMs and $\sum_{i=1}^{K_{SG}} w_{i,t} = 1$ for SG-DLMs. Then, models are aggregated to the appropriate feature set, whether it be LIN, TVP, SV, or TVP-SV. The curve shown here is the mean model weight as a percentage of the aggregate model weight over time, where the mean is taken over a given feature set. The evaluation period is January 1985 – December 2014.

Figure 2. Time series of predicted volatilities for SG-DLM models



The figure shows the time-series of predicted volatilities of excess returns for the four variants of the SG-DLM score-based model combinations, namely LIN, TVP, SV, and TVP-SV. Each panel represents a different asset, as labeled. Note that the scales of the vertical axes are different for each asset in order to compare patterns of change over time, as opposed comparing the magnitude of volatilities across assets. The solid black line represents the LIN model; the dotted red line tracks the TVP model; the solid green line depicts the SV model, while the blue dotted line displays the TVP-SV model. In each panel we also display, as a reference, the level of the unconditional standard deviation of each asset, computed over the whole evaluation period, January 1972 – December 2014.

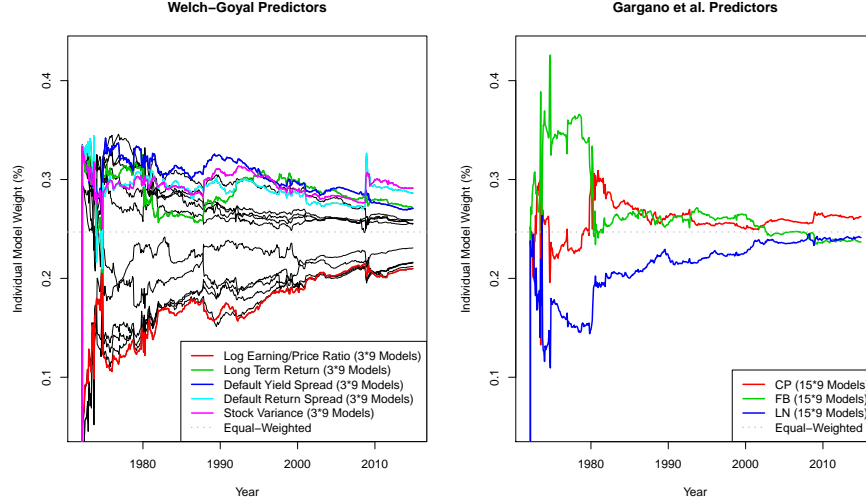
Figure 3. Time series of predicted correlations for SG-DLM models



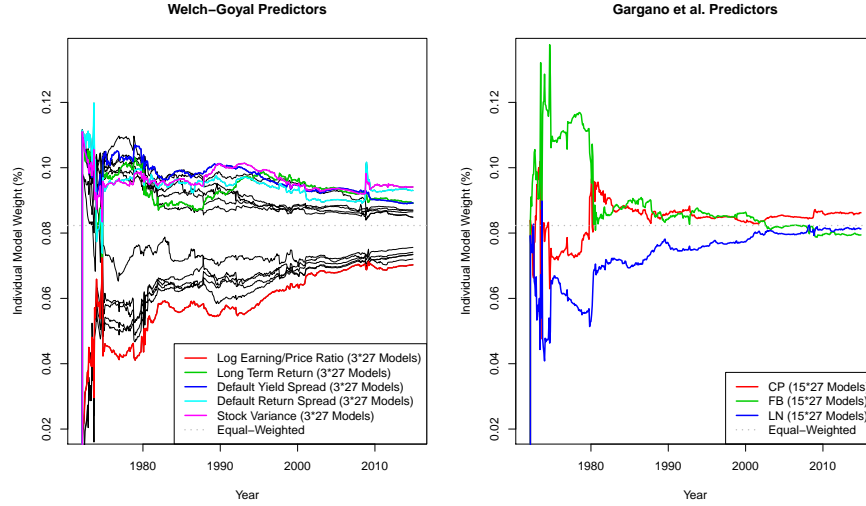
The figure shows the time-series of predicted correlations of excess returns for the four variants of the SG-DLM score-based model combinations, namely LIN, TVP, SV, and TVP-SV. Each panel represents a different pair of asset returns, as labeled. The solid black line represents the LIN model; the dotted red line tracks the TVP model; the solid green line depicts the SV model, while the blue dotted line displays the TVP-SV model. In each panel we also display, as a reference, the level of the unconditional correlation between each pair of asset returns, computed over the whole evaluation period, January 1972 – December 2014.

Figure 4. Time-series of score-based weights by predictor

Panel A: W-DLM



Panel B: SG-DLM

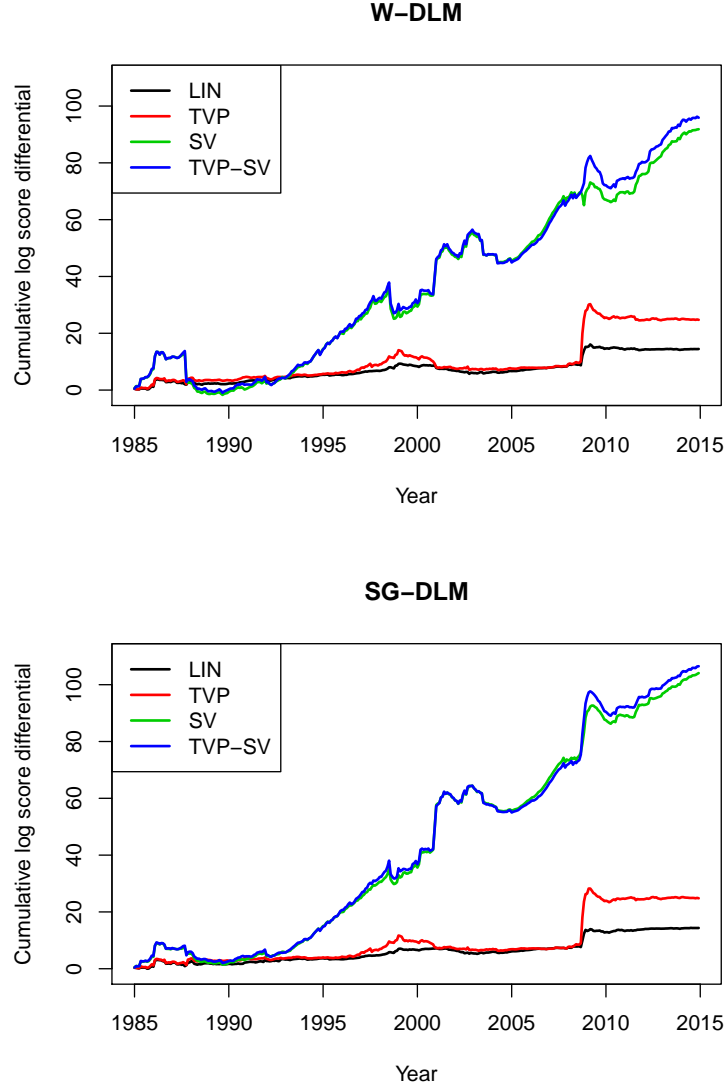


This figure shows the evolution over time of the score-based model combination weights for the various stock (left panels) and bond (right panels) predictors. The top two panels display results for the W-DLM models, while the bottom panels show the model combinations of the SG-DLM models. At each point in time t , we compute model \mathcal{M}_i weight ($i = 1, \dots, K_W$ in the case of the W-DLMs and $i = 1, \dots, K_{SG}$ in the case of the SG-DLMs) by looking at its historical statistical performance up through time $t - 1$, as determined by the log score:

$$w_{i,t} \propto \sum_{\tau=1}^{t-1} \ln(S_{i,\tau})$$

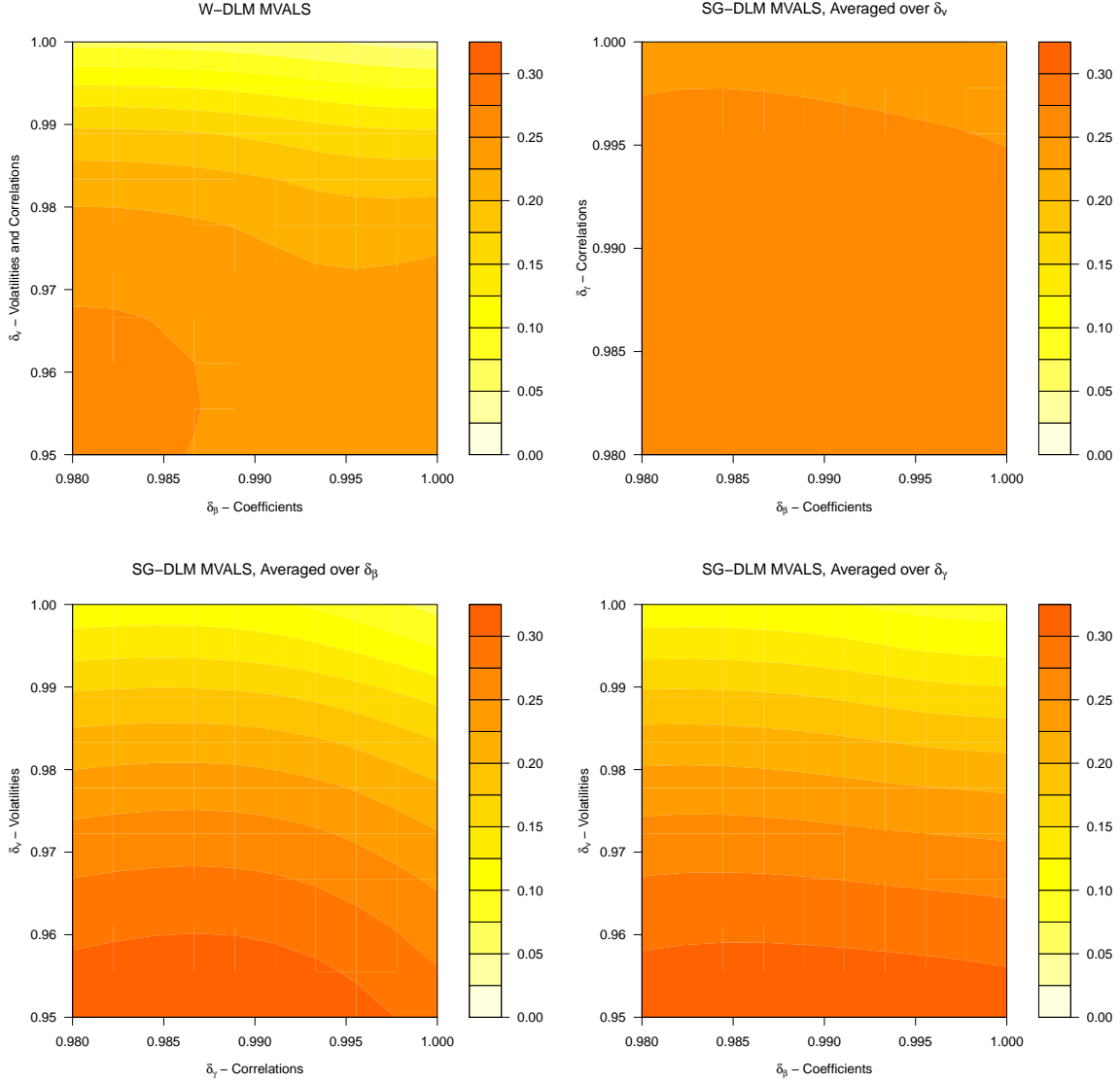
where $S_{i,\tau}$ denotes the recursively computed score for model i at time τ , which we obtain by evaluating a Gaussian density with mean vector and covariance matrix equal to $\mathbb{E}(\mathbf{r}_\tau | \mathcal{M}_i, \mathcal{D}_{\tau-1})$ and $\text{Cov}(\mathbf{r}_\tau | \mathcal{M}_i, \mathcal{D}_{\tau-1})$ at the realized log excess returns \mathbf{r}_τ . Next, we normalize the model weights across all models such that $\sum_{i=1}^{K_W} w_{i,t} = 1$ for W-DLMs and $\sum_{i=1}^{K_{SG}} w_{i,t} = 1$ for SG-DLMs. Then, models are aggregated according to the predictors included in each model. The curve shown here is the mean model weight as a percentage of the aggregate model weight over time, where the mean is taken over all models with the listed predictor. The evaluation period is January 1985 – December 2014.

Figure 5. Cumulative sum of the multivariate log score differentials for W-DLM and SG-DLM models



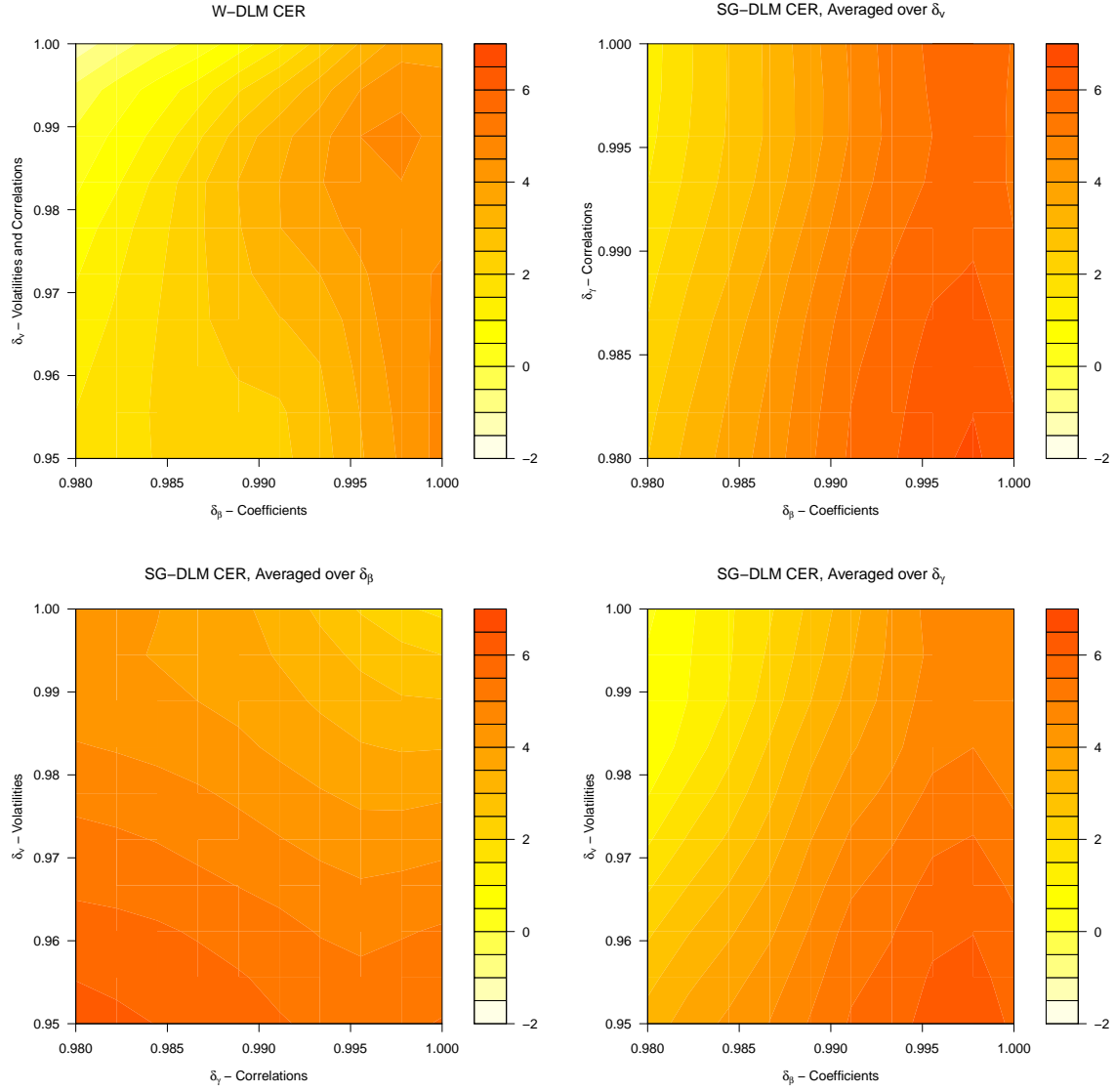
This figure plots the cumulative sum of the multivariate log score differentials, $CSMVLS_{it} = \sum_{\tau=t}^t (\ln(S_{i,\tau}) - \ln(S_{bcmk,\tau}))$ over time, where $S_{i,\tau}$ ($S_{bcmk,\tau}$) denote the model i 's (benchmark's) log score and is computed as described in [Subsection 2.3](#). The log score measures how accurate the multivariate distribution forecasts are given the realized observations. Within each panel, the solid black line represents the LIN model; the dotted red line tracks the TVP model; the solid green line depicts the SV model, while the blue dotted line displays the TVP-SV model. The evaluation period is January 1985 – December 2014.

Figure 6. Heat map of multivariate average log scores for different discount factors



This figure shows the multivariate average log scores (*MVALS*) of 100 different combinations of discount factors. The smaller a discount factor (δ) is, the more dynamic a feature is over time. δ_v controls the degree to which volatility is stochastic. δ_γ controls the degree to which correlations may time-vary. δ_β controls the time-variation in the regression coefficients. In order to create a two-dimensional plot, each SG-DLM pane model-averages over one of the three discount factors. The evaluation period is January 1985 – December 2014.

Figure 7. Heat map of certainty equivalent returns for different discount factors



This figure shows the certainty equivalent return differentials of 100 different combinations of discount factors. The smaller a discount factor (δ) is, the more dynamic a feature is over time. δ_v controls the degree to which volatility is stochastic. δ_γ controls the degree to which correlations may time-vary. δ_β controls the time-variation in the regression coefficient parameters. In order to create a two-dimensional plot, each SG-DLM pane model-averages over one the three discount factors. The evaluation period runs from 1985 through 2014.

Table 1. Summary Statistics

	Mean	StDev	Min	P_{25}	P_{75}	Max	SR
Panel A: Excess Returns							
5 Year Bond	0.002	0.018	-0.089	-0.007	0.011	0.094	0.340
10 Year Bond	0.002	0.032	-0.119	-0.014	0.019	0.163	0.240
Large-Cap Stocks	0.004	0.043	-0.235	-0.019	0.031	0.153	0.293
Mid-Cap Stocks	0.005	0.052	-0.294	-0.023	0.040	0.200	0.348
Small-Cap Stocks	0.006	0.061	-0.342	-0.029	0.043	0.259	0.312
Panel B: Bond Predictors							
Cochrane-Piazzesi factor	0.079	0.701	-3.630	-0.276	0.381	4.691	
Fama-Bliss spread, 5 Year	0.145	0.134	-0.374	0.047	0.251	0.423	
Fama-Bliss spread, 10 Year	0.183	0.160	-0.332	0.062	0.313	0.524	
Ludvigson-Ng factor	0.106	0.468	-1.919	-0.201	0.354	3.037	
Panel C: Stock Predictors							
Log dividend price ratio	-3.582	0.403	-4.524	-3.919	-3.310	-2.753	
Log dividend yield	-3.577	0.403	-4.531	-3.914	-3.306	-2.751	
Log earning price ratio	-2.825	0.439	-4.836	-2.993	-2.584	-1.899	
Log smooth earning price ratio	-3.075	0.341	-3.911	-3.275	-2.855	-2.274	
Log dividend payout ratio	-0.757	0.319	-1.244	-0.939	-0.601	1.379	
Book to market ratio	0.508	0.264	0.121	0.297	0.683	1.207	
T-Bill rate	0.049	0.031	0.000	0.030	0.064	0.163	
Long term yield	0.068	0.026	0.021	0.048	0.082	0.148	
Long term return	0.006	0.030	-0.112	-0.010	0.023	0.152	
Term spread	0.018	0.015	-0.036	0.007	0.031	0.045	
Default yield spread	0.010	0.005	0.003	0.007	0.012	0.034	
Default return spread	0.000	0.015	-0.098	-0.005	0.006	0.074	
Stock variance	0.002	0.004	0.000	0.001	0.002	0.065	
Net equity expansion	0.012	0.019	-0.058	0.003	0.025	0.051	
Inflation	0.003	0.003	-0.018	0.002	0.005	0.018	

This table provides summary statistics for the excess returns and predictors we consider in our analysis. Specifically, for each series we report the mean, standard deviation, minimum, quartiles, maximum. In the case of the excess returns, we also report the annualized Sharpe ratio. Panel A reports summary statistics for the stock and bond excess returns. All returns are continuously compounded. Monthly data on the stocks come from CRSP Cap-based portfolios file, where Large-cap stocks are the largest 20% of firms, Mid-cap are the 50th to 80th size percentile of firms, and Small-cap are the smallest half of firms. Monthly returns on five- and ten-year Treasury bonds are computed using the two-step procedure described in [Gargano et al. \(2017\)](#). Panel B reports summary statistics for the three bond predictors considered in [Gargano et al. \(2017\)](#), namely [Fama and Bliss \(1987\)](#) forward spreads for 5- and 10-year maturities, the [Cochrane and Piazzesi \(2005\)](#) factor, and the [Ludvigson and Ng \(2009\)](#) macro factor. Finally, panel C provides summary statistics for the 15 stock predictors considered in [Welch and Goyal \(2008\)](#). The sample period ranges from January 1962 to December 2014.

Table 2. Mean-squared forecast errors of W-DLM and SG-DLM models by asset

Features	W-DLM		SG-DLM	
	Equal	Score	Equal	Score
Panel A: Large-Cap				
LIN	0.998	0.998	0.998	0.997
TVP	1.008	1.007	1.009	1.007
SV	0.998	0.997	0.993	0.993
TVP-SV	1.008	1.007	1.008	1.007
Panel B: Mid-Cap				
LIN	0.991	0.993	0.991	0.993
TVP	1.015	1.013	1.016	1.014
SV	0.991	0.993	0.986	0.987
TVP-SV	1.015	1.013	1.015	1.014
Panel C: Small-Cap				
LIN	0.989	0.993	0.990	0.993
TVP	1.016	1.015	1.016	1.015
SV	0.989	0.993	0.986	0.987
TVP-SV	1.016	1.014	1.016	1.015
Panel D: 5-Year Bonds				
LIN	0.962	0.970	0.962	0.968
TVP	0.965	0.970	0.965	0.968
SV	0.962	0.964	0.962	0.963
TVP-SV	0.965	0.966	0.965	0.965
Panel E: 10-Year Bonds				
LIN	0.965	0.970	0.965	0.968
TVP	0.981	0.988	0.981	0.987
SV	0.965	0.966	0.967	0.967
TVP-SV	0.981	0.987	0.981	0.983

This table reports, for each of the five asset returns we considered, the ratio of mean-squared forecast errors (“MSFEs”) between a given model and the no-predictability benchmark, computed as

$$MSFE_{ij} = \frac{\sum_{\tau=\underline{t}}^T e_{ij,\tau}^2}{\sum_{\tau=\underline{t}}^T e_{bcmk,j,\tau}^2}$$

where \underline{t} denotes the beginning of the out-of-sample period, i refers to the model under consideration (i.e. LIN, TVP, SV, TVP-SV W-DLMs or SG-DLMs), $e_{ij,\tau} = r_{j\tau} - \mathbb{E}(r_{j\tau}|\mathcal{M}_i, \mathcal{D}_{\tau-1})$ and $e_{bcmk,j,\tau} = r_{j\tau} - \mathbb{E}(r_{j\tau}|\mathcal{M}_{bcmk}, \mathcal{D}_{\tau-1})$ are the forecast errors of asset return j at time τ associated with model i and the benchmark model, respectively. Values of $MSFE_{ij}$ below one suggest that for asset j , model i produces more accurate point forecasts than the no-predictability benchmark. Bold-faced values indicate the best performing models within each asset class and DLM type, while Equal and Score denote, respectively, equal-weighted and score-weighted model combinations. The evaluation period is January 1985 – December 2014.

Table 3. Average log score differentials of W-DLM and SG-DLM models by asset

Features	W-DLM		SG-DLM	
	Equal	Score	Equal	Score
Panel A: Large-Cap				
LIN	0.007	0.006	0.006	0.006
TVP	0.006	0.005	0.006	0.005
SV	0.016	0.015	0.012	0.010
TVP-SV	0.014	0.014	0.010	0.007
Panel B: Mid-Cap				
LIN	0.009	0.007	0.008	0.007
TVP	0.002	0.002	0.002	0.001
SV	0.007	0.004	0.014	0.015
TVP-SV	-0.002	-0.002	0.006	0.006
Panel C: Small-Cap				
LIN	0.010	0.008	0.008	0.006
TVP	0.003	0.002	0.002	0.001
SV	0.005	0.001	0.017	0.019
TVP-SV	-0.006	-0.007	0.008	0.010
Panel D: 5-Year Bonds				
LIN	0.016	0.016	0.016	0.016
TVP	0.014	0.015	0.014	0.015
SV	0.095	0.093	0.078	0.091
TVP-SV	0.094	0.094	0.077	0.089
Panel E: 10-Year Bonds				
LIN	0.018	0.018	0.018	0.019
TVP	0.018	0.018	0.018	0.019
SV	0.062	0.061	0.058	0.070
TVP-SV	0.062	0.063	0.059	0.070

This table reports, for each of the five asset returns we considered, the average log score (ALS) differential between model i and the no-predictability benchmark,

$$ALS_{ij} = \frac{1}{T - \underline{t} + 1} \sum_{\tau=\underline{t}}^T (\ln(S_{ij,\tau}) - \ln(S_{bcmk,j,\tau}))$$

where $S_{ij,\tau}$ ($S_{bcmk,j,\tau}$) denotes model i 's (benchmark's) log score at time τ , which we obtain by evaluating a univariate Gaussian density with mean vector $\mathbb{E}(\mathbf{r}_{j\tau}|\mathcal{M}_i, \mathcal{D}_{\tau-1})$ ($\mathbb{E}(\mathbf{r}_{j\tau}|\mathcal{M}_{bcmk}, \mathcal{D}_{\tau-1})$) and variance $Var(\mathbf{r}_{j\tau}|\mathcal{M}_i, \mathcal{D}_{\tau-1})$ ($Var(\mathbf{r}_{j\tau}|\mathcal{M}_{bcmk}, \mathcal{D}_{\tau-1})$) at the realized excess returns $\mathbf{r}_{j\tau}$. Positive values of ALS_{ij} indicate that model i produces on average more accurate density forecasts for variable j than the benchmark. Bold-faced values indicate the best performing models within each asset class and DLM type, while Equal and Score denote, respectively, equal-weighted and score-weighted model combinations. The evaluation period is January 1985 – December 2014.

Table 4. Weighted Mean-squared forecast errors of W-DLM and SG-DLM models

Features, Weighting:	W-DLM		SG-DLM	
	Equal	Score	Equal	Score
LIN	0.989	0.991	0.989	0.991
TVP	1.004	1.006	1.004	1.006
SV	0.989	0.990	0.990	0.990
TVP-SV	1.004	1.006	1.004	1.004

This table reports the ratio between the weighted multivariate mean squared forecast error (WMSFE, also known as the squared Mahalanobis distance) between model i and the no-predictability benchmark, computed as follows:

$$WMSFE_i = \frac{\sum_{\tau=\underline{t}}^T \mathbf{e}'_{i\tau} [\widehat{Cov}(\mathbf{r}_t)]^{-1} \mathbf{e}_{i\tau}}{\sum_{\tau=\underline{t}}^T \mathbf{e}'_{bckmk,\tau} [\widehat{Cov}(\mathbf{r}_t)]^{-1} \mathbf{e}_{bckmk,\tau}}$$

where $\mathbf{e}_{i\tau} = (e_{i1,\tau}, \dots, e_{iq,\tau})'$ and $\mathbf{e}_{bckmk,\tau} = (e_{bckmk,1,\tau}, \dots, e_{bckmk,q,\tau})'$ are the $q \times 1$ vector of forecast errors at time τ associated with model i and the benchmark model, while $\widehat{Cov}(\mathbf{r}_t)$ denotes the sample estimates of the asset returns unconditional variance-covariance matrix, computed over the evaluation period. Values of $WMSFE_i$ below one suggest that model i produces more accurate point forecasts than the no-predictability benchmark. Bold-faced values indicate the best performing models within DLM type, while Equal and Score denote, respectively, equal-weighted and score-weighted model combinations. The evaluation period is January 1985 – December 2014.

Table 5. Multivariate Average log score differentials of W-DLM and SG-DLM models

Features, Weighting:	W-DLM		SG-DLM	
	Equal	Score	Equal	Score
LIN	0.043	0.040	0.042	0.040
TVP	0.066	0.069	0.066	0.069
SV	0.266	0.255	0.271	0.289
TVP-SV	0.265	0.266	0.279	0.296

This table reports the multivariate average log score differentials (MVALS) between model i and the benchmark,

$$MVALS_i = \frac{1}{T - \underline{t} + 1} \sum_{\tau=\underline{t}}^T (\ln(S_{i,\tau}) - \ln(S_{bckmk,\tau}))$$

where $S_{i,\tau}$ ($S_{bckmk,\tau}$) are computed as described in [Subsection 2.3](#). Values of $MVALS_i$ above zero suggest that model i produces more accurate density forecasts than the no-predictability benchmark. Bold-faced values indicate the best performing models within DLM type, while Equal and Score denote, respectively, equal-weighted and score-weighted model combinations. The evaluation period is January 1985 – December 2014.

Table 6. Annualized certainty equivalent returns of W-DLM and SG-DLM models

Features, Weighting:	W-DLM		SG-DLM	
	Equal	Score	Equal	Score
LIN	4.608	3.804	4.526	3.814
TVP	0.361	0.262	0.513	0.460
SV	5.429	4.623	5.944	5.892
TVP-SV	2.940	2.628	3.772	3.947

This table reports the annualized CERs (in percentage terms) over the whole out-of-sample period for the various W-DLM and SG-DLM model combinations for an investor with power utility and coefficient of relative risk aversion $A = 5$. In particular, the table reports the CER gains relative to the no-predictability benchmark model, i.e. $CER_i - CER_{bcmk}$, where

$$CER_i = 100 \times \left(\left[\frac{1}{T - \underline{t} + 1} \sum_{t=\underline{t}}^T \widehat{W}_{it}^{1-A} \right]^{\frac{12}{1-A}} - 1 \right)$$

and where $\widehat{W}_{it} = \exp(r_{p,it})$ denotes the realized wealth at time t , as implied by model i (the CER of the benchmark model, 5.896, is computed in an analogous manner). Bold-faced values indicate the best performing models within DLM type, while Equal and Score denote, respectively, equal-weighted and score-weighted model combinations. The evaluation period is January 1985 – December 2014.

Appendix A The Wishart DLM

In this Appendix we describe our implementation of the W-DLM model of [West and Harrison \(1997, Section 16.4\)](#).

A.1 Basic Equations

For convenience, we reproduce here the key equations of the model. The W-DLM can be written as:

$$\mathbf{r}_t = \mathbf{B}_t' \mathbf{x}_{t-1} + \mathbf{v}_t \quad \mathbf{v}_t | \boldsymbol{\Sigma}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_t) \quad (\text{A.1})$$

where \mathbf{B}_t is the $p \times q$ matrix of time-varying regression coefficients and \mathbf{v}_t is a $q \times 1$ error vector, independent over time. The regression coefficients \mathbf{B}_t vary over time according to pq random walk processes,

$$\text{vec}(\mathbf{B}_t) = \text{vec}(\mathbf{B}_{t-1}) + \boldsymbol{\omega}_t \quad \boldsymbol{\omega}_t | \boldsymbol{\Sigma}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_t \otimes \mathbf{W}_t) \quad (\text{A.2})$$

where $\boldsymbol{\omega}_t$ is a $pq \times 1$ vector of zero-mean normally distributed error terms. The initial conditions are given by

$$\begin{aligned} \text{vec}(\mathbf{B}_0) | \boldsymbol{\Sigma}_0, \mathcal{D}_0 &\sim \mathcal{N}(\text{vec}(\mathbf{M}_0), \boldsymbol{\Sigma}_0 \otimes \mathbf{C}_0) \\ \boldsymbol{\Sigma}_0 | \mathcal{D}_0 &\sim \mathcal{IW}(n_0, \mathbf{S}_0) \end{aligned} \quad (\text{A.3})$$

A.2 Evolution

To begin the W-DLM forward filter, start at $t = 1$ such that the posterior distribution in step 1 below is the initial state of the filter based on $\mathcal{D}_{t-1} = \mathcal{D}_0$, the training dataset. After the three steps of the filter below are fulfilled for time $t = 1$, repeat the steps for $t = 2, \dots, T$, where T is the last time period in the data.

1. Evolve Posterior of time $t - 1$ to Prior of time t

Given the posterior distribution of the parameters at time $t - 1$

$$\begin{aligned} \text{vec}(\mathbf{B}_{t-1}) | \boldsymbol{\Sigma}_{t-1}, \mathcal{D}_{t-1} &\sim \mathcal{N}(\text{vec}(\mathbf{M}_{t-1}), \boldsymbol{\Sigma}_{t-1} \otimes \mathbf{C}_{t-1}) \\ \boldsymbol{\Sigma}_{t-1} | \mathcal{D}_{t-1} &\sim \mathcal{IW}(n_{t-1}, \mathbf{S}_{t-1}) \end{aligned} \quad (\text{A.4})$$

which we abbreviate as

$$\mathbf{B}_{t-1}, \mathbf{\Sigma}_{t-1} | \mathcal{D}_{t-1} \sim \mathcal{NIW}(\mathbf{M}_{t-1}, \mathbf{C}_{t-1}, n_{t-1}, \mathbf{S}_{t-1}). \quad (\text{A.5})$$

we evolve forward to create a prior for time t

$$\mathbf{B}_t, \mathbf{\Sigma}_t | \mathcal{D}_{t-1} \sim \mathcal{NIW}(\mathbf{M}_{t-1}, \hat{\mathbf{C}}_t, \hat{n}_t, \mathbf{S}_{t-1}). \quad (\text{A.6})$$

where, due to our choice of \mathbf{W}_t in (8), $\hat{\mathbf{C}}_t = \frac{1}{\delta_\beta} \mathbf{C}_{t-1}$ and $\hat{n}_t = \delta_v n_{t-1}$, for chosen values of $\delta_\beta, \delta_v \in (0, 1]$.

2. Forecast response variable at time t

As shown in equations (10)–(12), the predictive distribution of \mathbf{r}_t , based on time $t-1$ data, is given by

$$\mathbf{r}_t | \delta_\beta, \delta_v, \mathcal{D}_{t-1} \sim \mathcal{T}_{\hat{n}_t} \left(\mathbf{M}'_{t-1} \mathbf{x}_{t-1}, \quad \mathbf{S}_{t-1} (1 + \mathbf{x}'_{t-1} \hat{\mathbf{C}}_t \mathbf{x}_{t-1}) \right). \quad (\text{A.7})$$

with mean and covariance matrix given by

$$\mathbb{E}[\mathbf{r}_t | \delta_\beta, \delta_v, \mathcal{D}_{t-1}] = \mathbf{M}'_{t-1} \mathbf{x}_{t-1} \quad (\text{A.8})$$

$$\text{cov}[\mathbf{r}_t | \delta_\beta, \delta_v, \mathcal{D}_{t-1}] = \frac{\hat{n}_t}{\hat{n}_t - 2} \mathbf{S}_{t-1} (1 + \mathbf{x}'_{t-1} \hat{\mathbf{C}}_t \mathbf{x}_{t-1}). \quad (\text{A.9})$$

3. Update prior for time t into posterior for time t based on forecast error

After observing \mathbf{r}_t compute time t posterior distribution for \mathbf{B}_t and $\mathbf{\Sigma}_t$:

$$\mathbf{B}_t, \mathbf{\Sigma}_t | \mathcal{D}_t \sim \mathcal{NIW}(\mathbf{M}_t, \mathbf{C}_t, n_t, \mathbf{S}_t) \quad (\text{A.10})$$

In particular, we have that

$$\text{Posterior mean matrix} \quad \mathbf{M}_t = \mathbf{M}_{t-1} + \mathbf{a}_t \mathbf{e}'_t \quad (\text{A.11})$$

$$\text{Posterior covariance matrix factor} \quad \mathbf{C}_t = \hat{\mathbf{C}}_t - q_t \mathbf{a}_t \mathbf{a}'_t \quad (\text{A.12})$$

$$\text{Posterior degrees of freedom} \quad n_t = \hat{n}_t + 1 \quad (\text{A.13})$$

$$\text{Posterior residual covariance estimate} \quad \mathbf{S}_t = n_t^{-1} (\hat{n}_t \mathbf{S}_{t-1} + q_t^{-1} \mathbf{e}_t \mathbf{e}'_t). \quad (\text{A.14})$$

where

$$\text{1-step ahead forecast error} \quad \mathbf{e}_t = \mathbf{r}_t - \mathbf{M}'_{t-1} \mathbf{x}_{t-1} \quad (\text{A.15})$$

$$\text{1-step ahead coefficient variance factor} \quad q_t = 1 + \mathbf{x}'_{t-1} \hat{\mathbf{C}}_t \mathbf{x}_{t-1} \quad (\text{A.16})$$

$$\text{Adaptive coefficient vector} \quad \mathbf{a}_t = q_t^{-1} \hat{\mathbf{C}}_t \mathbf{x}_{t-1} \quad (\text{A.17})$$

Appendix B The Simultaneous Graphical DLM

In this Appendix, we describe our implementation of the SG-DLM of [Gruber and West \(2016\)](#).

B.1 Basic Equations

For convenience, we reproduce here the key equations of the SG-DLM. For $j = 1, \dots, q$, we write

$$r_{jt} = \mathbf{x}'_{j,t-1} \boldsymbol{\beta}_{jt} + \mathbf{r}'_{<j,t} \boldsymbol{\gamma}_{<j,t} + \nu_{jt} \quad \nu_{jt} \sim N(0, \sigma_{jt}^2) \quad (\text{B.1})$$

where

$$\begin{pmatrix} \boldsymbol{\beta}_{jt} \\ \boldsymbol{\gamma}_{<j,t} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}_{j,t-1} \\ \boldsymbol{\gamma}_{<j,t-1} \end{pmatrix} + \boldsymbol{\omega}_{jt} \quad \boldsymbol{\omega}_{jt} \sim N(\mathbf{0}, \mathbf{W}_{jt}) \quad (\text{B.2})$$

and the initial conditions are given by:

$$\begin{aligned} \begin{pmatrix} \boldsymbol{\beta}_{j0} \\ \boldsymbol{\gamma}_{<j,0} \end{pmatrix} \Big| \sigma_{j0}^2, \mathcal{D}_0 &\sim \mathcal{N} \left(\mathbf{m}_{j0}, \frac{\sigma_{j0}^2}{s_{j0}} \mathbf{C}_{j0} \right) \\ \sigma_{j0}^{-2} | \mathcal{D}_0 &\sim \mathcal{G} \left(\frac{n_{j0}}{2}, \frac{n_{j0} s_{j0}}{2} \right) \end{aligned} \quad (\text{B.3})$$

B.2 Evolution

To begin the SG-DLM forward filter, begin with $t = 1$ such that the posterior distribution in step 1 below is the initial state of the filter based on $\mathcal{D}_{t-1} = \mathcal{D}_0$, the training dataset. After the three steps of the filter are fulfilled for time $t = 1$, repeat the steps for $t = 2, \dots, T$, where T is the last time period in the data.

1. Evolve Posterior of time $t - 1$ to Prior of time t

Given the Posterior at time $t - 1$

$$\begin{pmatrix} \boldsymbol{\beta}_{j,t-1} \\ \boldsymbol{\gamma}_{<j,t-1} \end{pmatrix} \Big| \sigma_{j,t-1}^2, \mathcal{D}_{t-1} \sim \mathcal{N} \left(\mathbf{m}_{j,t-1}, \frac{\sigma_{j,t-1}^2}{s_{j,t-1}} \mathbf{C}_{j,t-1} \right) \quad (\text{B.4})$$

$$\sigma_{j,t-1}^{-2} | \mathcal{D}_{t-1} \sim \mathcal{G} \left(\frac{n_{j,t-1}}{2}, \frac{n_{j,t-1} s_{j,t-1}}{2} \right), \quad (\text{B.5})$$

which we abbreviate as

$$\begin{pmatrix} \boldsymbol{\beta}_{j,t-1} \\ \boldsymbol{\gamma}_{<j,t-1} \end{pmatrix}, \sigma_{j,t-1}^2 \Big| \mathcal{D}_{t-1} \sim \mathcal{NG}(\mathbf{m}_{j,t-1}, \mathbf{C}_{j,t-1}, n_{j,t-1}, s_{j,t-1}), \quad (\text{B.6})$$

we evolve it to create a prior for time t

$$\begin{pmatrix} \boldsymbol{\beta}_{j,t} \\ \boldsymbol{\gamma}_{<j,t} \end{pmatrix}, \sigma_{j,t}^2 \Big| \mathcal{D}_{t-1} \sim \mathcal{NG}(\mathbf{m}_{j,t-1}, \hat{\mathbf{C}}_{jt}, \hat{n}_{jt}, s_{j,t-1}) \quad (\text{B.7})$$

where, due to our choice of \mathbf{W}_t in (25), $\hat{\mathbf{C}}_{jt} = \begin{bmatrix} \mathbf{C}_{\beta\beta_{j,t-1}}/\delta_{\beta j} & \mathbf{C}_{\beta\gamma_{j,t-1}} \\ \mathbf{C}_{\gamma\beta_{j,t-1}} & \mathbf{C}_{\gamma\gamma_{j,t-1}}/\delta_{\gamma j} \end{bmatrix}$ and $\hat{n}_{jt} = \delta_{vj}n_{j,t-1}$, where $\mathbf{C}_{\beta\beta_{j,t-1}}$ and $\mathbf{C}_{\gamma\gamma_{j,t-1}}$ are, respectively, the covariance matrix factors for $\boldsymbol{\beta}_{j,t-1}$ and $\boldsymbol{\gamma}_{j,t-1}$.

2. Forecast response variable at time t

We calculate the moments of forecast returns one asset at a time, according to the order of dependence. Similarly-derived forecasting steps for this type of model can be found Zhao et al. (2016), Appendix B. As it does not depend on other assets' returns, the forecast for the first asset is given by

$$r_{1t}|\boldsymbol{\delta}_j, \mathcal{D}_{t-1} \sim \mathcal{T}_{\hat{n}_{1t}} \left(\mathbf{x}'_{1,t-1} \mathbf{m}_{1,t-1}, \mathbf{x}'_{1,t-1} \hat{\mathbf{C}}_{1t} \mathbf{x}_{1,t-1} + s_{1,t-1} \right) \quad (\text{B.8})$$

with mean and variance that are equal to:

$$\mathbb{E}[r_{1t}|\boldsymbol{\delta}_j, \mathcal{D}_{t-1}] = \mathbf{x}'_{1,t-1} \mathbf{m}_{1,t-1} \quad (\text{B.9})$$

$$\text{Var}[r_{1t}|\boldsymbol{\delta}_j, \mathcal{D}_{t-1}] = \frac{\hat{n}_{1t}}{\hat{n}_{1t} - 2} (\mathbf{x}'_{1,t-1} \hat{\mathbf{C}}_{1t} \mathbf{x}_{1,t-1} + s_{1,t-1}) \quad (\text{B.10})$$

where $\boldsymbol{\delta}_j = (\delta_{\beta j}, \delta_{\gamma j}, \delta_{vj})$. Now, all other assets' forecast moments can be found sequentially ($j = 2, \dots, q$). Similarly, their conditional distributions follow Student's t-distribution, with predictive moments given by

$$\mathbb{E}[r_{jt}|\boldsymbol{\delta}_j, \mathcal{D}_{t-1}] = \mathbf{x}'_{j,t-1} \mathbf{m}_{\beta_{j,t-1}} + \mathbb{E}[\mathbf{r}_{<j,t}|\boldsymbol{\delta}_j, \mathcal{D}_{t-1}]' \mathbf{m}_{\gamma_{<j,t-1}}, \quad (\text{B.11})$$

$$\begin{aligned} \text{Var}[r_{jt}|\boldsymbol{\delta}_j, \mathcal{D}_{t-1}] &= \frac{\hat{n}_{jt}}{\hat{n}_{jt} - 2} \left\{ \text{tr} \left(\hat{\mathbf{C}}_{\gamma_{<j,t}} \text{Cov}[\mathbf{r}_{<j,t}|\boldsymbol{\delta}_j, \mathcal{D}_{t-1}] \right) + c_{jt} + s_{j,t-1} \right\} \\ &\quad + \mathbf{m}'_{\gamma_{<j,t-1}} \text{Cov}[\mathbf{r}_{<j,t}|\boldsymbol{\delta}_j, \mathcal{D}_{t-1}] \mathbf{m}_{\gamma_{<j,t-1}} \end{aligned} \quad (\text{B.12})$$

and

$$\text{Cov}[r_{jt}, \mathbf{r}_{<j,t}|\boldsymbol{\delta}_j, \mathcal{D}_{t-1}] = \mathbf{m}'_{\gamma_{<j,t-1}} \text{Cov}[\mathbf{r}_{<j,t}|\boldsymbol{\delta}_j, \mathcal{D}_{t-1}] \quad (\text{B.13})$$

where $\mathbb{E}[\mathbf{r}_{<j,t}|\boldsymbol{\delta}_j, \mathcal{D}_{t-1}]$ and $\text{Cov}[\mathbf{r}_{<j,t}|\boldsymbol{\delta}_j, \mathcal{D}_{t-1}]$ are known, $\text{tr}()$ stands for the trace of a matrix, and

$$c_{jt} = \left(\mathbb{E}[\mathbf{r}_{<j,t}|\boldsymbol{\delta}_j, \mathcal{D}_{t-1}] \right)' \hat{\mathbf{C}}_{jt} \left(\mathbb{E}[\mathbf{r}_{<j,t}|\boldsymbol{\delta}_j, \mathcal{D}_{t-1}] \right). \quad (\text{B.14})$$

3. Update prior for time t into posterior for time t based on forecast error

After observing \mathbf{r}_t , time t posterior distribution for $\boldsymbol{\beta}_{j,t}$, $\boldsymbol{\gamma}_{<j,t}$, and $\sigma_{j,t}^2$ ($j = 1, \dots, q$) are given by

$$\begin{pmatrix} \boldsymbol{\beta}_{j,t} \\ \boldsymbol{\gamma}_{<j,t} \end{pmatrix}, \sigma_{j,t}^2 \Big| \mathcal{D}_t \sim \mathcal{NG}(\mathbf{m}_{j,t}, \mathbf{C}_{j,t}, n_{j,t}, s_{j,t}). \quad (\text{B.15})$$

In particular, we have that

$$\text{Posterior mean vector} \quad \mathbf{m}_{jt} = \mathbf{m}_{j,t-1} + \mathbf{a}_{jt}e_{jt} \quad (\text{B.16})$$

$$\text{Posterior covariance matrix factor} \quad \mathbf{C}_{jt} = (\hat{\mathbf{C}}_{jt} - \mathbf{a}_{jt}\mathbf{a}_{jt}'q_{jt})z_{jt} \quad (\text{B.17})$$

$$\text{Posterior degrees of freedom} \quad n_{jt} = \hat{n}_{jt} + 1 \quad (\text{B.18})$$

$$\text{Posterior residual variance estimate} \quad s_{jt} = z_{jt}s_{j,t-1}. \quad (\text{B.19})$$

where

$$\text{1-step ahead forecast error} \quad e_{jt} = r_{jt} - \begin{pmatrix} \mathbf{x}_{j,t-1} \\ \mathbf{r}_{<j,t} \end{pmatrix}' \mathbf{m}_{j,t-1} \quad (\text{B.20})$$

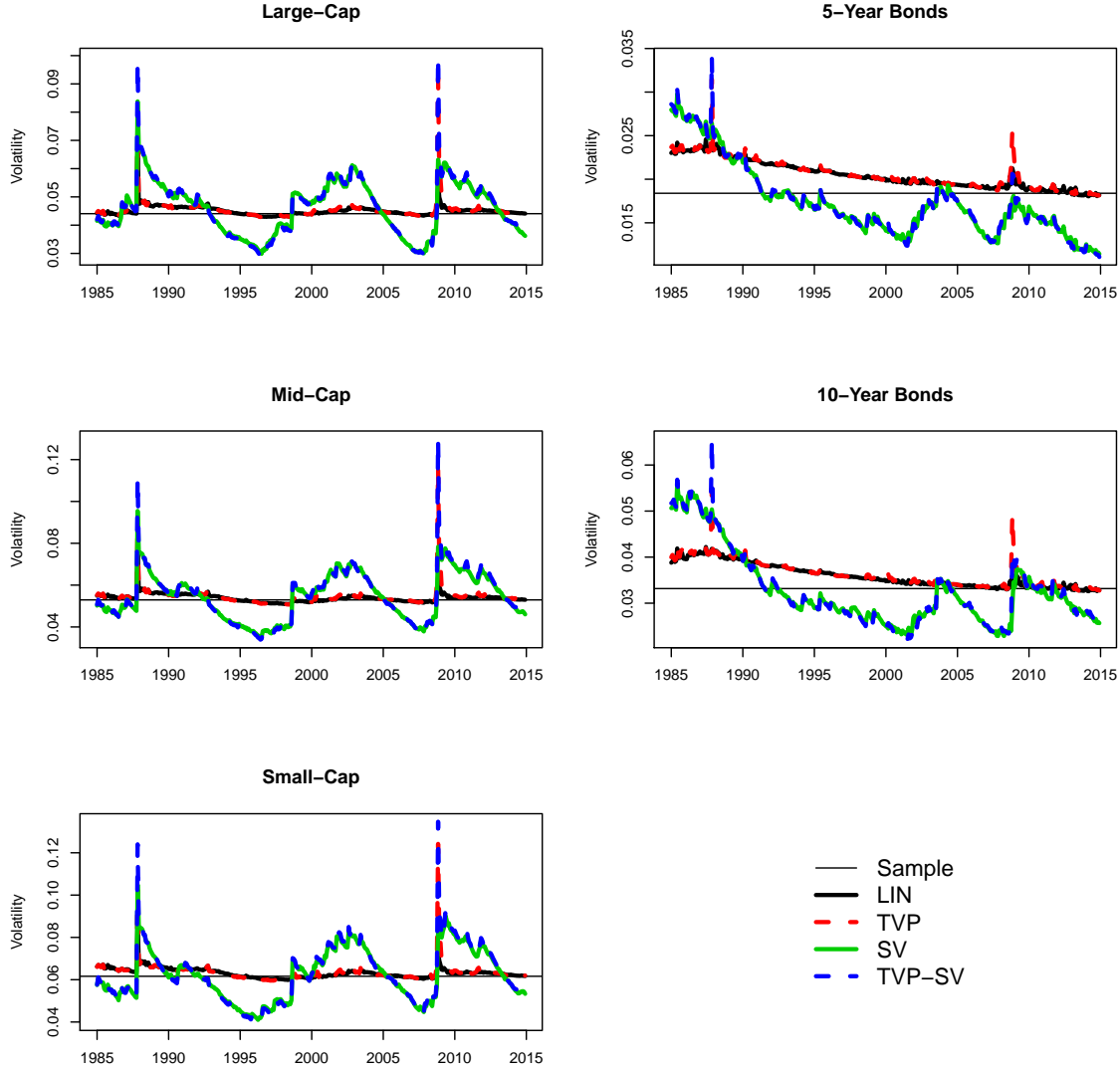
$$\text{1-step ahead forecast variance factor} \quad q_{jt} = s_{j,t-1} + \begin{pmatrix} \mathbf{x}_{j,t-1} \\ \mathbf{r}_{<j,t} \end{pmatrix}' \hat{\mathbf{C}}_{jt} \begin{pmatrix} \mathbf{x}_{j,t-1} \\ \mathbf{r}_{<j,t} \end{pmatrix} \quad (\text{B.21})$$

$$\text{Adaptive coefficient vector} \quad \mathbf{a}_{jt} = \hat{\mathbf{C}}_{jt} \begin{pmatrix} \mathbf{x}_{j,t-1} \\ \mathbf{r}_{<j,t} \end{pmatrix} / q_{jt} \quad (\text{B.22})$$

$$\text{Volatility update factor} \quad z_{jt} = (\hat{n}_{jt} + e_{jt}^2/q_{jt})/(\hat{n}_{jt} + 1) \quad (\text{B.23})$$

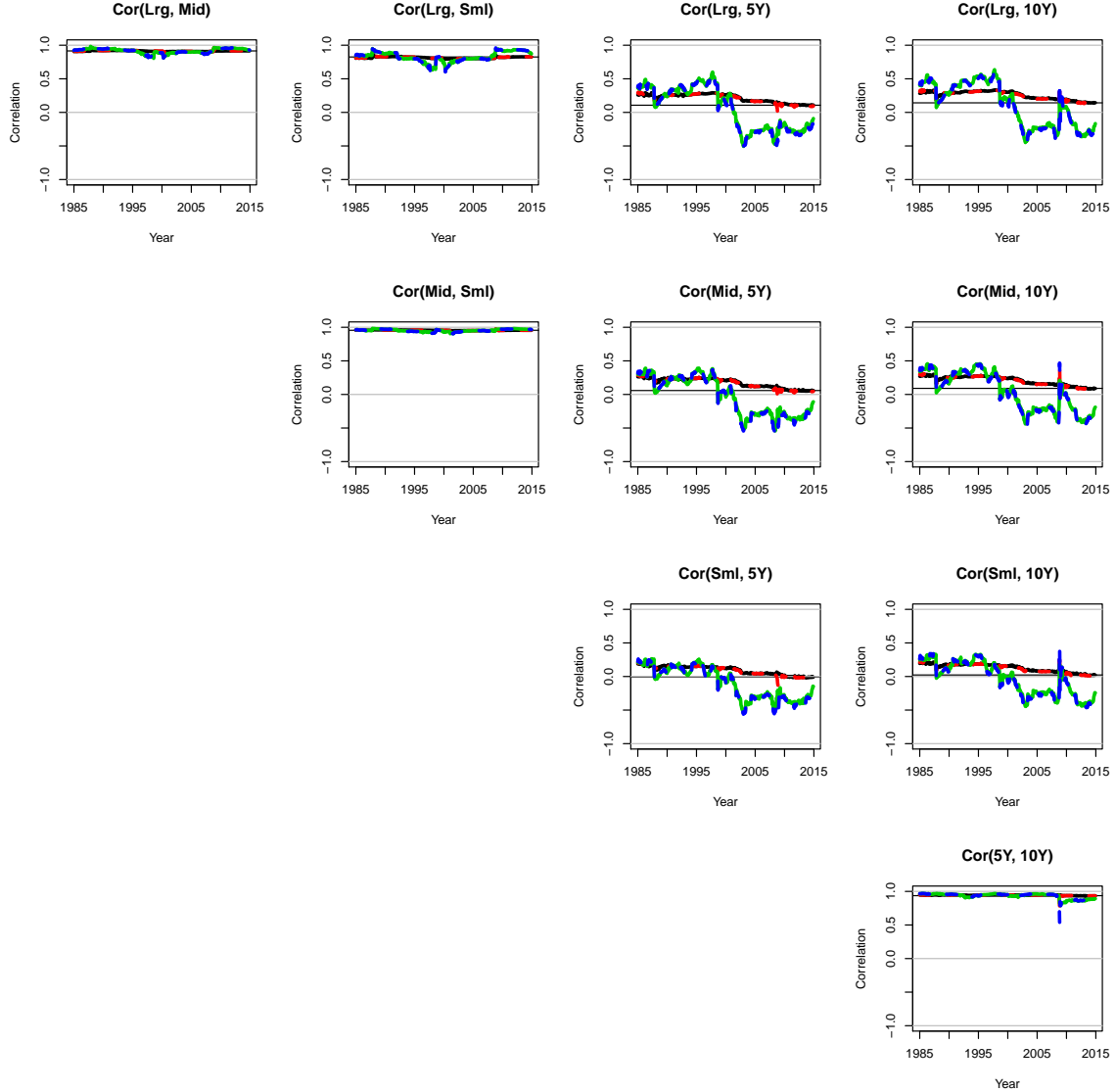
Appendix C Additional Results

Figure C.1. Time series of predicted volatilities for W-DLM models



The figure shows the time-series of predicted volatilities of expected excess returns for the four variants of the W-DLM score-based model combinations, namely LIN, TVP, SV, and TVP-SV. Each panel represents a different asset, as labeled. Note that the scales of the vertical axes are different for each asset in order to compare patterns of change over time, as opposed comparing the magnitude of volatilities across assets. The solid black line represents the LIN model; the dotted red line tracks the TVP model; the solid green line depicts the SV model, while the blue dotted line displays the TVP-SV model. In each panel we also display, as a reference, the level of the unconditional standard deviation of each asset, computed over the whole evaluation period, January 1985 – December 2014.

Figure C.2. Time series of predicted correlations for W-DLM models



The figure shows the time-series of predicted correlations of expected excess returns for the four variants of the W-DLM score-based model combinations, namely LIN, TVP, SV, and TVP-SV. Each panel represents a different pair of asset returns, as labeled. The solid black line represents the LIN model; the dotted red line tracks the TVP model; the solid green line depicts the SV model, while the blue dotted line displays the TVP-SV model. In each panel we also display, as a reference, the level of the unconditional correlation between each pair of asset returns, computed over the whole evaluation period, January 1985 – December 2014.