Large-Scale Dynamic Predictive Regressions*

Daniele Bianchi[†] Kenichiro McAlinn[‡]

Abstract

We propose and evaluate a large-scale dynamic predictive strategy for forecasting and economic decision making in a data-rich environment. Under this framework, clusters of predictors generate different predictive densities that are later synthesized within an implied time-varying latent factor model. We test our procedure by predicting both the inflation rate and the equity premium across different industries in the U.S., based on a large set of macroeconomic and financial variables. The main results show that our framework generates both statistically and economically significant out-of-sample outperformance compared to a variety of sparse and dense regression-based models while maintaining critical economic interpretability.

Keywords: Dynamic Forecasting, Predictive Regressions, Data-Rich Models, Forecast Combination, Macroeconomic Forecasting, Returns Predictability.

JEL codes: C11, C53, D83, E37, G11, G12, G17

*We thank Roberto Casarin, Marco Del Negro, Dimitris Korobilis, Andrew Patton, Davide Pettenuzzo, Mike West (discussant), and participants at the NBER-NSF Seminar on Bayesian Inference in Econometrics and Statistics at Stanford, the Barcelona GSE Summer Forum in Time Series Econometrics and Applications for Macroeconomics and Finance, the 10th ECB Workshop on Forecasting Techniques, the world meeting of the International Society of Bayesian Analysis in Edinburgh for their helpful comments and suggestions.

[†]The University of Warwick, Warwick Business School, Coventry, UK. Daniele.Bianchi@wbs.ac.uk

[‡]The University of Chicago, Booth School of Business, Chicago, IL, USA. kenichiro.mcalinn@chicagobooth.edu

1 Introduction

The increasing availability of large datasets, both in terms of the number of variables and the number of observations, combined with the recent advancements in the field of econometrics, statistics, and machine learning, have spurred the interest in predictive models with many explanatory variables, both in finance and economics.[1] As not all predictors are necessarily relevant, decision makers often pre-select the most important candidate explanatory variables by appealing to economic theories, existing empirical literature, and their own heuristic arguments. Nevertheless, a decision maker is often still left with tens– if not hundreds– of sensible predictors that may possibly provide useful information about the future behavior of quantities of interest. However, the out-of-sample performance of standard techniques, such as ordinary least squares, maximum likelihood, or Bayesian inference with uninformative priors tends to deteriorate as the dimensionality of the data increases, which is the well known curse of dimensionality.[2]

Confronted with a large set of predictors, two main classes of models became popular, even standard, within the regression framework. *Sparse* modeling focus on the selection of a sub-set of variables with the highest predictive power out of a large set of predictors, and discard those with the least relevance. LASSO-type regularizations are by far the most used in both research and practice. Regularized models take a large number of predictors and introduce penalization to discipline the model space. Similarly, in the Bayesian literature, a prominent example is the spike-and-slab prior proposed by George and McCulloch (1993), which introduced variable selection through a data-augmentation approach. A second class of models fall under the heading of *dense* modeling; this is based on the assumption that, a priori, all variables could bring useful information for

[1]See, e.g., Elliott and Timmermann (2004), Timmermann (2004), Bai and Ng (2010), Rapach, Strauss, and Zhou (2010), Billio, Casarin, Ravazzolo, and van Dijk (2013), Manzan (2015), Pettenuzzo and Ravazzolo (2016), Harvey, Liu, and Zhu (2016), Giannone, Lenza, and Primiceri (2017), and McAlinn and West (2017), just to cite a few.

[2]Even with a moderate number of predictors the empirical investigation of all possible model combinations could rapidly become infeasible. For instance, for a moderate size linear regression with $p = 30$ regressors, investigating the whole set of possible features combinations would require estimating $2^{30} \approx 1.07$ billion regression models.

prediction, although the impact of some of these might be small. As a result, the statistical features of a large set of predictors are assumed to be captured by a much smaller set of common latent components, which could be either static or dynamic. Factor analysis is a clear example of dense statistical modeling, which is highly popular in applied macroeconomics (see, e.g., Stock and Watson 2002 and De Mol, Giannone, and Reichlin 2008 and the references therein).

Both these approaches entail either an implicit or explicit reduction of the model space. The intention is to arbitrarily lower model complexity to balance bias and variance, in order to potentially minimize predictive losses. For instance, in LASSO-type shrinkage estimators, increasing the tuning parameter (i.e. increasing shrinkage) leads to a higher bias, thus using cross-validation aims to balance the bias-variance tradeoff by adjusting the tuning parameter. Similarly, in factor models, the optimal number of latent common components is chosen by using information criteria to reduce the variance by reducing the model dimensionality at the cost of increasing the bias (see, e.g., Bai and Ng 2002). In addition, for economic and financial decision making, in particular, these dimension reduction techniques always lead to a decrease in consistent interpretability, something that might be critical for policy makers, analysts, and investors.

In this paper, we propose a novel class of data-rich predictive synthesis techniques and contribute to the literature on predictive modeling and decision making with large datasets. We take a significantly different approach towards the bias-variance tradeoff by breaking a large dimensional problem into a set of small dimensional ones. More specifically, we retain all of the information available and *decouple* a large predictive regression model into a set of smaller regressions constructed by clustering the set of regressors into J different groups, each one containing fewer regressors than the whole, according to their economic meaning or some quantitative clustering. Rather than assuming a priori the existence of a sparse structure or few latent common components, we retain all of the information by estimating J different predictive densities– separately and sequentially– one for each group of predictors, and *recouple* them dynamically to generate aggregate predictive densities for the quantity of interest. By decoupling a large predictive regression model into smaller, less complex regressions, we keep the aggregate model variance low while sequentially learning and correcting for the misspecification bias that characterize each group. As this is the case, the recoupling step benefits from biased models, as long as the bias has a signal that can be learned. This flips the bias-variance tradeoff around, exploiting the weakness of low complexity models to an advantage in the recoupling step, therefore improving the out-of-sample predictive performance.

Our methodology differs from existing model combination schemes by utilizing the theoretical foundations and recent developments in dynamic density forecast with multiple models (see, e.g., McAlinn and West 2017). That is, the decoupled models are effectively treated as separate latent states that are learned and calibrated using the Bayes theorem in an otherwise typical dynamic linear modeling setup. Under this framework, the inter-dependencies between the group-specific predictive densities, as well as biases within each group, can be sequentially learned and corrected; information that is critical, though lost in typical model combination techniques. Along this line, Clemen (1989), Makridakis (1989), Diebold and Lopez (1996), and Stock and Watson (2004) pointed out that individual forecasting models are likely to be subject to misspecification bias of unknown form. Even in a stationary world, the data generating process is likely to be far more complex than assumed by the best forecasting model and it is unlikely that the same set of regressors dominates all others at all points in time. As a result, sequentially learning the aggregate bias and exploiting the latent inter-dependencies among group-specific predictions can be viewed as a way to robustify the aggregate prediction against model misspecification and measurement errors underlying the individual forecasts.

Unlike sparse modeling, we do not assume a priori that there is sparsity in the set of predictors. As a matter of fact, using standard LASSO-type shrinkage will implicitly impose a dogmatic prior that only a small subset of regressors is useful for predictions and the rest is noise, i.e., sparsity is pre-assumed. Yet, there is no guarantee that the Lasso estimator is smooth and asymptotically consistent to the true sparsity pattern in the presence of highly correlated predictors and model instability; two conditions that are often encountered in empirical applications (see, e.g., Meinshausen, Yu et al. 2009).

We implement the proposed the methodology, which we call decouple-recouple synthesis (DRS), and explore both its econometric underpinnings and economic gains on both a macroeconomic and a finance application. More specifically, in the first application we test the performance of our decouple-recouple approach to forecast the one- and three-month ahead annual inflation rate in the U.S. over the period 1986/1 to 2015/12, a context of topical interest (see, e.g., Cogley and Sargent 2005, Primiceri 2005, Stock and Watson 2007, Koop and Korobilis 2010, and Nakajima and West 2013, among others). The set of monthly macroeconomic predictors consists of an updated version of the Stock and Watson macroeconomic panel available at the Federal Reserve Bank of St.Louis. Details on the construction of the dataset can be found in McCracken and Ng (2016). The second application relates to forecasting monthly year-on-year total excess returns across different industries in the U.S. from 1970/1 to 2015/12, based on a large set of both industry-specific and aggregate predictors. The predictors have been chosen from previous academic studies and existing economic theory (see, e.g., Goyal and Welch 2008 and Rapach et al. 2010).

We compare forecasts against a set of mainstream model combination techniques such as a standard Bayesian model averaging (BMA), in which the forecast densities are mixed with respect to sequentially updated model probabilities (see, e.g., Harrison and Stevens 1976, Sect 12.2 West and Harrison 1997 and Pettenuzzo and Ravazzolo 2016), as well as against simpler, equal-weighted averages of the model-specific forecast densities using linear pools, i.e., arithmetic means of forecast densities, with some theoretical underpinnings (see, e.g., West 1984 and Diebold and Shin 2017). While some of these strategies might seem overly simplistic, they have been shown to dominate some more complex aggregation strategies in some contexts (Genre, Kenny, Meyler, and Timmermann, 2013). In addition, we compare the forecasts from our setting with a state-of-the-art LASSO-type regularization, PCA based latent factor modeling (see, e.g., Stock and Watson 2002 and McCracken and Ng 2016), as well as the simple historical average (HA), as suggested by Campbell and Thompson (2007) and Goyal and Welch (2008). Finally, we compare our decouple-recouple predictive strategy against the marginal predictive densities computed from the group-specific set of predictors taken separately.

Forecasting accuracy is assessed in a statistical sense based on two different out-of-sample performance metrics. We report as a main performance metric the Log Predictive Density Ratio (LPDR), at forecast horizon k and across time indices t. In addition, although our main focus is on density forecasts, we also report the Root Mean Squared Forecast Error (RMSFE), which captures the forecast optimality for a mean squared utility. Irrespective of the performance evaluation metric, our decouple-recouple model synthesis scheme emerges as the best for forecasting the yearly total excess returns across different industries. The differences in the LPDRs are stark and clearly shows a performance gap in favor of DRS.

As far as the out-of-sample economic performance is concerned, we run a battery of tests based on a power-utility representative investor with moderate risk aversion. The comparison is conducted for the unconstrained as well as short-sales constrained investor at monthly horizons, for the entire sample. We find that our DRS strategy results in a higher CER (relative to an investor that uses the historical mean as forecast) of more than 150 basis points per year, on average across sectors. Consistent with the predictive accuracy results, we generally find that the DRS strategy produces higher CER improvements than the competing specifications, both with and without short-sales portfolio constraints. In addition, we show that DRS allows to reach a higher CER also on a "per-period" basis, which suggests that there are economically important gains for a power utility investor.

2 Decouple-Recouple Predictive Strategy

A decision maker $\mathcal{D}$ is interested in predicting some quantity y, in order to make some informed decision based on a large set of predictors, which are all considered relevant to $\mathcal{D}$, but with varying degree. In the context of macroeconomics, for example, this might be a policy maker interested in forecasting inflation using multiple macroeconomic indicators, that a policy maker can or cannot control. Similar interests are also relevant in finance, with, for example, portfolio managers tasked with implementing optimal portfolio allocations on the basis of expected future returns on risky assets. A canonical and relevant approach is to consider a basic linear regression;

$$y_t = \boldsymbol{\beta}' \boldsymbol{z}_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \nu_t), \tag{1}$$

where $\boldsymbol{z}_t$ is a $p-$dimensional vector of predictors, $\boldsymbol{\beta}$ is the $p-$dimensional vector of betas, and ϵ_t is some observation noise, which is assumed here to be Gaussian to fix ideas.

In many practically important applications, the dimension of predictors relevant to make an informed decision is large, possibly too large to directly fit something as simple as an ordinary linear regression. As a matter of fact, at least a priori, all of these predictors could provide relevant information for $\mathcal{D}$. Under this setting, regularization or shrinkage would not be consistent with $\mathcal{D}$'s decision making process, as she has no dogmatic priors on the size of the model space. Similarly, dimension reduction techniques such as principal component analysis and factor models, e.g., Stock and Watson (2002) and Bernanke, Boivin, and Eliasz (2005), while using all of the predictors available, reduces them to a small preset number of latent factors that are hard to interpret or control, in the sense of decision making.

Our decouple-recouple strategy[3] exploits the fact that the potentially large $p-$dimensional vector of predictors can be partitioned into smaller groups $j = 1{:}J$, modifying Eq. (1) to

$$y_t = \boldsymbol{\beta}_1' \boldsymbol{z}_{t-1,1} + ... + \boldsymbol{\beta}_j' \boldsymbol{z}_{t-1,j} + ... + \boldsymbol{\beta}_J' \boldsymbol{z}_{t-1,J} + \epsilon_t, \quad \epsilon_t \sim N(0, \nu_t). \tag{2}$$

These groups can be partitioned based on some qualitative categories (e.g. group of predictors related to the same economic phenomenon), or by some quantitative measure (e.g. clustering based on similarities, correlation, etc.), though the dimension of each partitioned group should be relatively small in order to obtain

[3]We note that the term "decouple/recouple" stems from emerging developments in multivariate analysis and graphical models, where a large cross-section of data are decoupled into univariate models and recoupled via a post-process recovery of the dependence structure (see Gruber and West 2016 and the recent developments in Gruber and West 2017; Chen, K., Banks, Haslinger, Thomas, and West 2017). While previous research focuses on making complex multivariate models scalable, our approach does not directly recover some specific portion of a model (full models are available but not useful), instead aims to improve forecasts and understand the underlying structure through the subgroups.

sensible estimates. The first step of our model combination strategy is to *decouple* Eq. (2) into J smaller predictive models, such as,

$$y_t = \boldsymbol{\beta}_j' \boldsymbol{z}_{t-1,j} + \epsilon_{tj}, \quad \epsilon_{tj} \sim N(0, \nu_{tj}), \tag{3}$$

for all $j = 1{:}J$, producing forecast distributions $p(y_{t+k}|\mathcal{A}_j)$, where $\mathcal{A}_j$ denotes each group of predictors and k denotes the forecast horizon, $1 \leq k$. Since Eq. (3) is a linear projection of data from each group of explanatory variables, we can consider, without loss of generality, that $p(y_{t+k}|\mathcal{A}_j)$ is reflecting the group-specific information regarding the future behavior of the quantity of interest. In the second step, we *recouple* the densities $p(y_{t+k}|\mathcal{A}_j)$ for $j = 1{:}J$ in order to obtain a forecast distribution $p(y_{t+k}|\mathcal{A})$ reflecting and incorporating all of the information that arises from each group of predictors. In the most simple setting, $p(y_{t+k}|\mathcal{A}_j)$ can be recoupled via linear pooling (see, e.g., Geweke and Amisano 2011);

$$p\left(y_{t+k}|\mathcal{A}\right) = \sum_{j=1}^{J} w_j p(y_{t+k}|\mathcal{A}_j), \tag{4}$$

where weights $w_{1:J}$ are often estimated based on past observations and predictive performances (e.g. using $w_{1:J}$ proportional to the marginal likelihood). However, while this linear combination structure is conceptually and practically appealing, it does not capture the fact that we expect and understand that each $p(y_{t+k}|\mathcal{A}_j)$ to be biased and dependent with each other (i.e., groups of predictors could be highly correlated). Arguably, each group-specific prediction $p(y_{t+k}|\mathcal{A}_j)$ is misspecified unless one of them is the data generating process, which is something that we can hardly expect in economics or finance. In this respect, Geweke and Amisano (2012) formally show that even when none of the constituent models are true, linear pooling and BMA assign positive weights to several models.

The dependence between $p(y_{t+k}|\mathcal{A}_j)$ and $p(y_{t+k}|\mathcal{A}_q)$, for $j \neq q$, is also a crucial aspect of model combination. The optimal combination of weights should be chosen to minimize the expected loss of the *combined* forecast, which, by definition, reflects both the forecasting accuracy of each sub-model and the correlation across forecasts. For instance, it is evident that the marginal predictive power of macroeconomic variables related to the labor market is somewhat correlated with the explanatory power of output and income. In addition, correlations across predictive densities are arguably latent and dynamic. For instance, the spillover effects interest rates, market liquidity, and aggregate financial variables possibly changed before and after the great financial crisis of 2008/2009. Thus, an effective combination scheme must be able to sequentially learn and recover the latent inter-dependencies between the groups/sub-models.

2.1 Time-Varying Predictive Synthesis

The baseline assumption is that a decision maker $\mathcal{D}$ aims to incorporate information from J individual predictive models labeled $\mathcal{A}_j$, $(j = 1{:}J)$. The predictive density from each group of predictors is considered to be a latent state, such that $p(y_t|\mathcal{A}_j)$ represents a distinct prior on state $j = 1, ..., J$. That is, each $\mathcal{A}_j$ provides their own prior distribution about what they believe the outcome in the form of a predictive distribution $h_{tj}(x_{tj}) = p(y_t|\mathcal{A}_j)$; the collection of which defines the information set $\mathcal{H}_t = \{h_{t1}(x_{t1}), \dots, h_{tJ}(x_{tJ})\}$. The difference between this approach and more general latent factor models, such as PCA, is that we allow to anchor each latent state, using priors $p(y_t|\mathcal{A}_j)$ at each time t, to a group that $\mathcal{D}$ specifies. These latent states are then calibrated and learned using Bayesian updating.

A formal prior-posterior updating scheme posits that, for a given prior $p(y_t)$, and (prior) information set $\mathcal{H}_t$ provided by $\mathcal{A}_{1:J}$, we can update using the Bayes theorem to obtain a posterior $p(y_t|\mathcal{H}_t)$. Due to the complexity of $\mathcal{H}_t$– a set of J density functions with cross-sectional time-varying dependencies as well as individual biases– the aggregate predictive density might be difficult to define. We build on the work of McAlinn and West 2017 (linking to past literature on Bayesian pooling of expert opinion analysis by West and Crosse (1992) and West (1992), which extend the basic theorem of Genest and Schervish (1985)), that show that, under a specific consistency condition, $\mathcal{D}$'s the time-varying posterior density takes the form

$$p(y_t|\mathbf{\Phi}_t, \mathcal{H}_t) = \int \alpha_t(y_t|\boldsymbol{x}_t, \mathbf{\Phi}_t) \prod_{j=1:J} h_{tj}(x_{tj}) dx_{tj} \tag{5}$$

where $\boldsymbol{x}_t = x_{t,1:J}$ is a $J-$dimensional latent state vector at time t, $\alpha_t(y_t|\boldsymbol{x}_t, \boldsymbol{\Phi}_t)$ is a conditional density function, which reflects how the decision maker believes these latent states $\boldsymbol{x}_t$ to be synthesized, and $\boldsymbol{\Phi}_t$ represents some time-varying parameters learned and calibrated over $\tau = 1, \ldots, t$. It is important to note that the theory does not specify the form of $\alpha_t(y_t|\boldsymbol{x}_t, \boldsymbol{\Phi}_t)$. In fact, McAlinn and West (2017) show that many forecast combination methods, from linear combinations (including BMA) to more recently developed density pooling methods (e.g. Aastveit, Gerdrup, Jore, and Thorsrud, 2014; Kapetanios, Mitchell, Price, and Fawcett, 2015; Pettenuzzo and Ravazzolo, 2016), are special cases of Eq.(5).

This general framework implies that $\boldsymbol{x}_t$ is a realization of the inherent dynamic latent factors at time t and synthesis is achieved by recoupling these separate latent predictive densities through the time-varying conditional distribution $\alpha_t(y_t|\boldsymbol{x}_t, \boldsymbol{\Phi}_t)$. Though the theory does not specify $\alpha_t(y_t|\boldsymbol{x}_t, \boldsymbol{\Phi}_t)$, a natural choice is to impose linear dynamics (see, e.g., McAlinn and West, 2017), such that,

$$\alpha_t(y_t|\boldsymbol{x}_t, \boldsymbol{\Phi}_t) = N(y_t|\boldsymbol{F}_t'\boldsymbol{\theta}_t, v_t), \tag{6}$$

where $\boldsymbol{F}_t = (1, \boldsymbol{x}_t')'$ and $\boldsymbol{\theta}_t = (\theta_{t0}, \theta_{t1}, ..., \theta_{tJ})'$ represents a $(J+1)-$vector of time-varying synthesis coefficients. Observation noise is reflected in the innovation variance term v_t, and the time-varying parameters $\boldsymbol{\Phi}_t$ is defined as $\boldsymbol{\Phi}_t = (\boldsymbol{\theta}_t, v_t)$.

The evolution of these parameters needs to be specified to complete the model specification. We follow existing literature in dynamic linear models and assume that both $\boldsymbol{\theta}_t$ and v_t evolve as a random walk to allow for stochastic changes over time as is tradition in the Bayesian time series literature (see West and Harrison 1997; Prado and West 2010). Thus, we consider

$$y_t = \boldsymbol{F}_t'\boldsymbol{\theta}_t + \nu_t, \quad \nu_t \sim N(0, v_t), \tag{7a}$$

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(0, v_t\boldsymbol{W}_t), \tag{7b}$$

where $v_t\boldsymbol{W}_t$ represents the innovations covariance for the dynamics of $\boldsymbol{\theta}_t$ and v_t the residuals variance in predicting y_t, which is based on past information and the set of models' predictive densities. The residual ν_t and the evolution *innovation* $\boldsymbol{\omega}_s$ are independent over time and mutually independent for all t, s. The dynamics of $\boldsymbol{W}_t$ is imposed by a standard, single discount factor specification as in West and Harrison (1997) (Ch.6.3) and Prado and West (2010) (Ch.4.3). The residual variance v_t follows a beta-gamma random-walk volatility model such that $v_t = v_{t-1}\delta/\gamma_t$, where $\delta \in (0, 1]$ is a discount parameter, and $\gamma_t \sim Beta\,(\delta n_t/2, (1-\delta)\,n_t/2)$ are innovations independent over time and independent of $v_s, \boldsymbol{\omega}_r$ for all t, s, r, with $n_t = \delta n_{t-1} + 1$, the degrees of freedom.

Figure 1 visually summarizes the main difference between our approach and a standard forecast combination scheme. Unlike existing model ensemble techniques, we do not assume the forecasts to be independent, and sequentially recalibrate $h_{tj}(x_{tj}) = p(y_t|\mathcal{A}_j)$ as latent states, which are then effectively transferred onto the time varying parameters $\boldsymbol{\Phi}_t = (\boldsymbol{\theta}_t, v_t)$. These parameters are then used to compute the posterior forecast distribution.

2.2 Estimation Strategy

Estimation for the decouple step is straightforward and depends on the model assumptions for each group-specific model. For instance, for a typical dynamic linear regressions, we can compute each $h_{tj}(x_{tj}) = p(y_t|\mathcal{A}_j)$ using conjugate Bayesian updating. As for the recouple step, some discussion is needed. In particular, the joint posterior distribution of the latent states and the structural parameters is not available in closed form. We implement a Markov Chain Monte Carlo (MCMC) approach using an efficient Gibbs sampling scheme. In our framework, the latent states are represented by the predictive densities of the models, $\mathcal{A}_j, j = 1, ..., J$, and the synthesis parameters, $\boldsymbol{\Phi}_t$. As a result, posterior estimates provide insights into the nature of the biases and inter-dependencies of those latent states.

More precisely, the MCMC algorithm involves a sequence of standard steps in a customized two-component block Gibbs sampler: the first component simulates from the conditional posterior distribution of the latent states given the data and the second component simulates the synthesis parameters. The first step is the "calibration" step, whereby we learn the biases and inter-dependencies of the agent forecasts (latent states). In the second step, we "combine" the models' predictions by effectively mapping the biases and inter-dependencies of the latent states, $h_{tj}(x_{tj})$, onto the parameters $\boldsymbol{\Phi}_t$ in a dynamic manner.

The second step involves a standard implementation of the FFBS algorithm central to MCMC in all conditionally normal dynamic linear models (Frühwirth-Schnatter 1994; West and Harrison 1997, Sect 15.2; Prado and West 2010, Sect 4.5). In our sequential learning and forecasting context, the full MCMC analysis is redone at each time point as time evolves and new data are observed. Standing at time T, the historical information $\{y_{1:T}, \mathcal{H}_{1:T}\}$ is available and initial prior $\boldsymbol{\theta}_0 \sim N(\boldsymbol{m}_0, \boldsymbol{C}_0 v_0/s_0)$ and $1/v_0 \sim G(n_0/2, n_0 s_0/2)$, and discount factors (β, δ) are specified. At each iteration of the sampler we sequentially cycle through the above steps.

Finally, posterior predictive distributions of quantities of interest are computed as mixtures of the model-dependent marginal predictive densities synthesized by $\alpha_t(y_t|\boldsymbol{x}_t, \boldsymbol{\Phi}_t)$. Integration over the model space is performed using our MCMC scheme, which provides consistent estimates of the latent states and parameters. A more detailed description of the algorithm and how forecasts are generated can be found in Appendix A.

2.3 Simulation Study

To test and exemplify our proposed method in a controlled setting, we conduct a simple simulation study that emulates conditions observed in economic data; namely that all variables are correlated and that there are omitted variables, with the true data generating process being unattainable. To do this, we simulate data by the following data generating process:

$$y = -2z_1 + 3z_2 + 5z_3 + \epsilon, \quad \epsilon \sim N(0, 0.01), \tag{8a}$$

$$z_1 = \frac{1}{3}z_3 + \nu_1, \quad \nu_1 \sim N\left(0, \frac{2}{3}\right), \quad z_2 = \frac{1}{5}z_3 + \nu_2, \quad \nu_2 \sim N\left(0, \frac{4}{5}\right), \tag{8b}$$

$$z_3 = \nu_3, \quad \nu_3 \sim N(0, 0.01), \tag{8c}$$

where only $\{y, z_1, z_2\}$ are observed and z_3 is omitted. Firstly, all covariates are correlated. Secondly, since the key variable z_3 is not observed, we have a serious omitted variable that drives all the data observed. Because of this, all models that can be constructed will be misspecified. Additionally, because z_3 drives everything else, there is significant bias in all models generated.

We consider forecasting 500 simulated data points and compare the eight different strategies that are also considered in the empirical application. Notably, the individual models are subset of all possible models with either $\{z_1\}$, $\{z_2\}$, or $\{z_1, z_2\}$ as regressors in a linear regression. We test a simplified version of our proposed "decouple-recouple" predictive strategy, where the synthesis function is a simple linear regression with non-informative priors (Jeffreys' prior). This yields a simpler setup for DRS in order to specifically consider and compare the strengths of our strategy.

Testing predictive performance by measuring the Root Mean Squared one-step ahead Forecast Error (RMSE) for different number of samples, we find that DRS outperforms all other methods and strategies by at least 2%, which, although small, is substantial and consistent across different data lengths. The results indicate the strengths and superiority of DRS in a controlled setting that emulates the conditions encountered in real economic data. Full descriptions and results can be found in Appendix B.

3 Research Design

In a realistic setting, the data generating process is not necessarily time invariant and effects of variables change over time with shifts and shocks. To cope with this, we introduce dynamics into the decoupled predictive densities to fully exploit the flexibility of our predictive strategy. Specifically, for the decouple step we use a dynamic linear model (DLM: West and Harrison, 1997; Prado and West, 2010), for each group, $j = 1{:}J$,

$$y_t = \boldsymbol{\beta}'_{tj}\boldsymbol{z}_{t-1,j} + \epsilon_{tj}, \quad \epsilon_{tj} \sim N(0, \nu_{tj}), \tag{9a}$$

$$\boldsymbol{\beta}_{tj} = \boldsymbol{\beta}_{t-1,j} + \boldsymbol{u}_{tj}, \quad \boldsymbol{u}_{tj} \sim N(0, \nu_{tj}\boldsymbol{U}_{tj}), \tag{9b}$$

where the coefficients follow a random walk and the observation variance evolves with discount stochastic volatility. Priors for each decoupled predictive regression are assumed fairly uninformative, such as $\boldsymbol{\beta}_{0j}|v_{0j} \sim N(\boldsymbol{m}_{0j}, (v_{0j}/s_{0j})\boldsymbol{I})$

with $\boldsymbol{m}_{0j} = \mathbf{0}'$ and $1/v_{0j} \sim G(n_{0j}/2, n_{0j}s_{0j}/2)$ with $n_{0j} = 10, s_0 = 0.01$. For the recouple step, we follow the synthesis function in Eq. (6), with the following priors: $\boldsymbol{\theta}_{0n}|v_{0n} \sim N(\boldsymbol{m}_{0n}, (v_{0n}/s_{0n})\boldsymbol{I})$ with $\boldsymbol{m}_0 = (0, \mathbf{1}'/J)'$ and $1/v_{0n} \sim G(n_{0n}/2, n_{0n}s_{0n}/2)$ with $n_{0n} = 10, s_{0n} = 0.01$. The discount factors are $(\beta, \delta) = (0.95, 0.99)$. The dynamic specification in Eq. (9) is attractive due to its parsimony, ease to compute, and the smoothness it induces to the parameters.[4]

3.1 Competing Predictive Strategies

For both studies, we compare our framework against a variety of competing predictive strategies. First, we compare the aggregate predictive density from DRS against the predictive densities from each group-specific predictive regressions calculated from Eq.(9a)-(9b). That is, we test the benefits of the recoupling step and the calibration of the aggregate model prediction upon learning the latent biases and inter-dependencies.

Second, we compare our DRS strategy against a LASSO shrinkage regression, where the coefficients in Eq.(1) are estimated in an expanding window fashion from a penalized least-squares regression, i.e.,

$$\hat{\boldsymbol{\beta}}_{LASSO} = \arg\min_{\boldsymbol{\beta}} \parallel \boldsymbol{y} - \boldsymbol{\beta z} \parallel_2^2 + \lambda \sum_{i=1}^{n} \mid \beta_i \mid$$

where the shrinkage parameter λ is calibrated by leave-one-out cross-validation, that is the model is trained and the shrinkage parameter is selected based on the quasi-out-of-sample prediction accuracy. Although such an approach is computationally expensive, it provides an accurate out-of-sample calibration of the shrinkage parameter (see, e.g., Shao 1993).

A third competing predictive strategy relates to dynamic factor modeling where factors are latent and extracted from the set of predictors. More precisely, the factor model relates each y_t to an underlying vector of $q < n$ of random

[4]See, e.g., Jostova and Philipov (2005), Nardari and Scruggs (2007), Adrian and Franzoni (2009), Pastor and Stambaugh (2009), Binsbergen, Jules, and Koijen (2010), Dangl and Halling (2012), Pastor and Stambaugh (2012), and Bianchi, Guidolin, and Ravazzolo (2017b), among others.

variables $\boldsymbol{f}_t$, the latent common factors, via

$$\begin{aligned} y_t &= \boldsymbol{\beta}'\boldsymbol{f}_t + \epsilon_t, \qquad \epsilon_t \sim N(0,\nu_t), \\ \boldsymbol{z}_t &= \boldsymbol{\gamma}\boldsymbol{f}_t + u_t, \qquad u_t \sim N(0,\tau), \end{aligned}$$

where (i) the factors $\boldsymbol{f}_t$ are independent with $\boldsymbol{f}_t \sim N(0, I_q)$, (ii) the ϵ_t are independent and normally distributed with a discount-factor volatility dynamics, (iii) $u_t \perp \boldsymbol{f}_s \forall s,t$, and (iv) $\boldsymbol{\gamma}$ is the $n \times q$ matrix of factor loadings. We recursively estimate the factor model by using an expanding window where the optimal number of factors is selected using the Bayesian information criterion (BIC). Also, we assume that the factor coefficients on the latent factors are time-varying and follow a dynamic linear model consistent with the dynamic specification in Eq.(9). More precisely, at each time t we replace $\boldsymbol{z}_{tj}$ with $\boldsymbol{f}_t$ in Eq. (9a) and the slope parameters have a random walk dynamics as in Eq. (9b). We note that for both the LASSO regression and factor model, we have tested and compared the expanding window to the moving window strategy, and found that the expanding window strategy to perform better overall in the applications considered in this paper.

The fourth competing strategy is dynamic Bayesian Model Averaging (BMA), in which the forecast densities are mixed with respect to sequentially updated model probabilities whereby the weights are restricted to be inside the unit circle and the sum of the model weights is restricted to be equal to one (e.g. Harrison and Stevens, 1976; West and Harrison, 1997, Sect 12.2), i.e.,

$$p\left(y_{t+k}|\mathcal{A}\right) = \sum_{j=1}^{J} w_j p(y_{t+k}|\mathcal{A}_j), \qquad \sum_{j=1}^{J} w_{jt} = 1, \qquad w_{jt} \geq 0$$

where the restrictions on the weights w_{it} are necessary and sufficient to assure that $p\left(y_{t+k}|\mathcal{A}\right)$ is a density function for all values of the weights and all arguments of the group-specific predictive regressions (see, e.g., Geweke and Amisano 2011). As often in the BMA literature, the weights $w_{jt}, j = 1, ..., J$, are chosen based on the posterior model probabilities, i.e., $w_j = p(\mathcal{A}_j|y_{1:t})$, where

$$p(\mathcal{A}_j|y_{1:t}) = \frac{p(y_t|\mathcal{A}_j)p\left(\mathcal{A}_j|y_{1:t-1}\right)}{\sum_{j=1}^{J} p(y_t|\mathcal{A}_j)p\left(\mathcal{A}_j|y_{1:t-1}\right)},$$

Choice of weights in any forecast combination is widely regarded as a difficult and important question. Existing literature shows that, despite being theoretically suboptimal, an equal weighting scheme generates a substantial outperformance with respect to optimal weights based on log-score or in-sample calibration (see, e.g., Timmermann 2004, Smith and Wallis 2009, and Diebold and Shin 2017). For this reason, a fifth competing predictive strategy we used is linear pooling of predictive densities with equal weights, that is each sub-model has the same weight in the aggregate forecast, i.e., $w_j = 1/J$.

Both the BMA and the equal-weight linear combination allow us to compare the benefit of the predictive density calibration that is featured in the recoupling step underlying our DRS strategy. Finally, we also compare DRS against the prediction from the historical average for the financial application.

3.2 Out-of-Sample Performance Measures

Following standard practice in the forecasting literature, we evaluate the quality of our predictive strategy against competing models based on both point and density forecasts. In particular, we first compare predictive strategies based on the Root Mean Squared Error (RMSE), i.e.,

$$RMSE_s = \left(\frac{1}{T-\tau-1} \sum_{t=\tau}^{T-1} \left(y_{t+1} - E\left[y_{t+1}|y_{1:t}, \mathcal{M}_s\right]\right)^2 \right)^{1/2}$$

where $T-\tau-1$ represents the out-of-sample period, $E\left[y_{t+1}|y_{1:t}, \mathcal{M}_s\right]$ the one-step ahead point forecast conditional on information up to time t from the predictive strategy $\mathcal{M}_s$, and y_{t+1} is the realized returns.

Although informative, performance measures based on point forecasts only give a partial assessment. Ideally, one also wants to compare the predictive densities across strategies. As a matter of fact, performance measures based on predictive densities weigh and compare dispersion of forecast densities along with location, and elaborate on raw RMSE measures; comparing both measurements, i.e., point and density forecasts, gives a broader understanding of the predictive abilities of the different strategies. That is, performance measures based on the predictive density provide an assessment of a model ability to explain not only the expected value, i.e., the equity premium, but also the overall distribution of excess returns, naturally penalizing the size/complexity of different models. We compare predictive strategies based on the log predictive density ratios (LPDR); at horizon k and across time indices t, i.e.,

$$\mathrm{LPDR}_t = \sum_{i=1}^{t} \log\{p(y_{i+k}|y_{1:i}, \mathcal{M}_s)/p(y_{i+k}|y_{1:i}, \mathcal{M}_0)\}, \tag{10}$$

where $p(y_{t+k}|y_{1:t}, \mathcal{M}_s)$ is the predictive density computed at time t for the horizon $t+k$ under the model or model combination/aggregation strategy indexed by $\mathcal{M}_s$, compared against our forecasting framework labeled by $\mathcal{M}_0$. As used by several authors recently (e.g. Nakajima and West, 2013; Aastveit, Ravazzolo, and Van Dijk, 2016), LPDR measures provide a direct statistical assessment of relative accuracy at multiple horizons that extend traditional 1-step focused Bayes' factors.

We also evaluate the economic significance within the context of the finance application by considering the optimal portfolio choice of a representative investor with moderate risk aversion. An advantage of our Bayesian setting is that we are not reduced to considering only mean-variance utility, but can use more general constant relative risk aversion preferences (see, e.g., Pettenuzzo, Timmermann, and Valkanov 2014). In particular, we construct a two asset portfolio with a risk-free asset (r_t^f) and a risky asset (y_t; industry returns) for each t, by assuming the existence of a representative investor that needs to solve the optimal asset allocation problem

$$\omega_\tau^\star = \arg\max_{w_\tau} E\left[U\left(\omega_\tau, y_{\tau+1}\right)|\mathcal{H}_\tau\right], \tag{11}$$

with $\mathcal{H}_\tau$ indicating all information available up to time τ, and $\tau = 1, ..., t$. The investor is assumed to have power utility

$$U\left(\omega_{\tau}, y_{\tau+1}\right) = \frac{\left[(1-\omega_{\tau})\exp\left(r_{\tau}^{f}\right) + \omega_{\tau}\exp\left(r_{\tau}^{f} + y_{\tau+1}\right)\right]^{1-\gamma}}{1-\gamma}, \tag{12}$$

here, γ is the investor's coefficient of relative risk aversion. The time τ subscript reflects the fact that the investor chooses the optimal portfolio allocation conditional on her available information set at that time. Taking expectations with respect to the predictive density in Eq. (5), we can rewrite the optimal portfolio allocation as

$$\omega_{\tau}^{\star} = \arg\max_{\omega_{\tau}} \int U\left(\omega_{\tau}, y_{\tau+1}\right) p(y_{\tau+1}|\mathcal{H}_{\tau}) dy_{\tau+1}, \tag{13}$$

As far as DRS is concerned, the integral in Eq. (13) can be approximated using the draws from the predictive density in Eq. (5). The sequence of portfolio weights $\omega_{\tau}^{\star}, \tau = 1, ..., t$ is used to compute the investor's realized utility for each model-combination scheme. Let $\hat{W}_{\tau+1}$ represent the realized wealth at time $\tau+1$ as a function of the investment decision, we have

$$\hat{W}_{\tau+1} = \left[(1-\omega_{\tau}^{\star})\exp\left(r_{\tau}^{f}\right) + \omega_{\tau}^{\star}\exp\left(r_{\tau}^{f} + y_{\tau+1}\right)\right], \tag{14}$$

The certainty equivalent return (CER) for a given model is defined as the annualized value that equates the average realized utility. We follow Pettenuzzo et al. (2014) and compare the the average realized utility of DRS $\hat{U}_{\tau}$ to the average realized utility of the model based on the alternative predicting scheme i, over the forecast evaluation sample:

$$CER_i = \left[\frac{\sum_{\tau=1}^{t}\hat{U}_{\tau,i}}{\sum_{\tau=1}^{t}\hat{U}_{\tau}}\right]^{\frac{1}{1-\gamma}} - 1, \tag{15}$$

with the subscript i indicating a given model combination scheme, $\hat{U}_{\tau,i} = \hat{W}_{\tau,i}^{1-\gamma}/(1-\gamma)$, and $\hat{W}_{\tau,i}$ the wealth generated by the competing model i at time τ according to Eq. (14). A negative CER_i shows that model i generates a lower (certainty equivalent) return than our predictive strategy.

4 Empirical Results

4.1 Forecasting Aggregate Inflation in the U.S.

The first application concerns monthly forecasting of annual inflation in the U.S., a context of topical interest (Cogley and Sargent, 2005; Primiceri, 2005; Koop, Leon-Gonzalez, and Strachan, 2009; Nakajima and West, 2013). We consider a balanced panel of $N = 128$ monthly macroeconomic and financial variables over the period 1986:01 to 2015:12. A detailed description of how variables are collected and constructed is provided in McCracken and Ng (2016). These variables are classified into eight main categories depending on their economic meaning: Output and Income, Labor Market, Consumption and Orders, Orders and Inventories, Money and Credit, Interest Rate and Exchange Rates, Prices, and Stock Market.

The empirical application is conducted as shown in Figure 2; first, the decoupled models are analyzed in parallel over 1986:01-1993:06 as a training period, simply estimating the DLM in Eq. (9) to the end of that period to estimate the forecasts from each subgroup. This continues over 1993:07-2015:12, but with the calibration of recouple strategies, which, at each quarter t during this period, is run with the MCMC-based DRS analysis using data from 1993:07 up to time t. We discard the forecast results from 1993:07-2000:12 as training data and compare predictive performance from 2001:01-2015:12. The time frame includes key periods that tests the robustness of the framework, such as the inflating and bursting of the dot.com bubble, the building up of the Iraq war, the 9/11 terrorist attacks, the sub-prime mortgage crisis and the subsequent great recession of 2008–2009. We consider a 1-, 3-, and 12-step step ahead forecasts, in order to reflect interests and demand in practice.

Panel A of Table 1 shows results aggregated over the testing sample. Our decouple-recouple strategy improves the one-step ahead out-of-sample forecasting accuracy relative to the group-specific models, LASSO, PCA, equal-weight averaging, and BMA. The RMSE of DRS is about half of the one obtained by LASSO-type shrinkage, a quarter compared to that of PCA, and significantly lower than equal-weight linear pooling and Bayesian model averaging. In general, our decouple-recouple strategy exhibits improvements of 4% up to over 250% in comparison to the competing predictive strategies considered. For each group-specific model, we note that the Labor Market achieve similarly good point forecasts, which suggests that the labor market and price levels might be intertwined and dominate the aggregate predictive density. Also, past prices alone provide a good performance, consistent with the conventional wisdom that a simple AR(1) model often represent a tough benchmark to beat. Output and Income, Orders and Inventories, and Money and Credit, also perform well, with Output and Income outperforming Labor Market in terms of density forecasts.

Similarly, Panel B and panel C of Table 1 both show that DRS for the 3- and 12-step ahead forecasts reflect a critical benefit of using our model combination scheme for multi-step ahead evaluation. As a whole, the results are relatively similar to that of the 1-step ahead forecasts, with DRS outperforming all other methods, though the order of performance is different for each horizon. Interestingly, the LASSO sensibly deteriorates as the forecasting horizon increases when it comes to predicting the overall ahead distribution of future inflation. Similarly, both the equal weight and BMA show a significant -50% in terms of density forecast accuracy. It is fair to notice though that the LASSO predictive strategy is the only one that does not explicitly consider time varying volatility of inflation, which is a significant limitation of the methodology, even though stochastic volatility is something that has been shown to substantially affect inflation forecasting (see, e.g., Clark 2011 and Chan 2017, among others). In terms of equal-weight pooling and BMA, we observe that BMA does outperform equal weight, though this is because the BMA weights degenerated quickly to Orders and Inventories, which highlights the problematic nature of BMA, as it acts more as a model selection device rather than a forecasting calibration procedure.

Appendix C shows the recursive one-step ahead out-of-sample performance of DRS in terms of predictive density. The results make clear that the out-of-sample performance of DRS with respect to the benchmarking model combination/shrinkage schemes tend to steadily increase throughout the sample.

Delving further into the dynamics of our decouple-recouple model combination scheme, Figure 3 highlights the first critical component of the recoupling step, namely learning the latent inter-dependencies among and between the subgroups. For the sake of interpretability Figure 3 reports a rescaled version of the J-dimensional vector of posterior estimates $\hat{\boldsymbol{\theta}}_t = \left(\hat{\theta}_{1t}, \ldots, \hat{\theta}_{Jt}\right)'$ by using a logistic transformation, i.e.,

$$\tilde{\theta}_{jt} = \frac{\exp\left(\hat{\theta}_{jt}\right)}{\sum_{j=1}^{J} \exp\left(\hat{\theta}_{jt}\right)}, \quad j = 1, \ldots, J \qquad \text{s.t.} \quad \tilde{\theta}_{jt} \in (0,1), \quad \sum_{j=1}^{J} \tilde{\theta}_{jt} = 1. \quad (16)$$

That is, each posterior estimates are rescaled to be inside the unit simplex, and sum to one across groups of predictors. This allows to give a clearer interpretability of the relative importance of these latent interdependencies through time. The left panel shows the results for the one-step ahead forecast; we note that prior to the dot.com bubble, Money and Credit, Output and Income, and Order and Inventories have the largest weight although they quickly reduce their weight throughout the rest of the testing period.

One large trend in coefficients is with Labor Market, Prices, and Orders and Inventories. After the dot.com crash, we see a large increase in weight assigned to Labor Market, making it the group with the highest impact on the predictive density for most of the period. A similar pattern also emerges with Interest and Exchange Rates at the early stages of the great financial crisis, though to a lesser extent. Yet, Labor Market does not always represent the group with the largest weight towards the end of the sample. In the aftermath of the the dot.com crash the marginal weight of Prices trends significantly upwards, crossing Labor Market around the sub-prime mortgage crisis, making it by far the highest weighted group and the end of the test period.

Compared to the results from the one-step ahead forecasts, the right panel of Figure 3 shows that there are specific differences in the dynamics of the latent inter-dependencies when forecasting inflation on a longer horizon. More specifically, we note a significant decrease in importance of Labor Market before and after the great recession, and a marked increase of the relative importance of Prices after the great financial crisis, with Labor Market still quite significant towards the end of the sample. This is a stark contrast to the results of the 1-step ahead forecasts and reflects an interesting dynamic shift in importance of each subgroup that highlights the flexible specification of DRS for multi-step ahead modeling.

Since the parameters of the recoupling step are considered to be latent states, the conditional intercept of the recoupling scheme can be interpreted as the aggregate bias, namely a free-roaming component, which is not directly pinned down by any group of predictors. Specifically, the time variation in the conditional intercept can be thought of as a reflection of unanticipated (by the group-specific models, and as an extension, the group indicators) economic shocks, which then affect inflation forecasts with some lag.

Figure 4, the intercept in the synthesis model, clearly shows a sign switch in the aftermath of the short recession in the early 2000s and the financial crisis of 2008–2009. In addition, we note some specific differences between the predictive bias for the one-step ahead (solid light-blue line) and the three-step ahead (dashed light-blue line) forecasts. These differences are key to understand the long-term dynamics of inflation. For one, compared to the one-step ahead conditional intercept, the conditional intercept of the longer-run forecast is clearly amplified. This is quite intuitive, as we expect forecast performance to deteriorate as the forecast horizon moves further away, and thus more reliant on the free-roaming component of the latent states. Second, the bias of both forecasts substantially change in the aftermath of both the mild recession in the US in the early 2000s and the great financial crisis. The lag here should not look suspicious as the persistent time variation of both the sub-model predictive densities and the recoupling step imply some stickiness in the bias adjustment.

Further results, including retrospective analysis of the latent interdependencies, can be found in Appendix .

4.2 Forecasting the Equity Premium for Different U.S. Industries

We consider a large set of predictors to forecast monthly year-on-year excess returns in the U.S. across different industries from 1970:01 to 2015:12. The choice of the predictors is guided by previous academic studies and existing economic theory with the goal of ensuring the comparability of our results with these studies (see, e.g., Lewellen 2004, Avramov 2004, Goyal and Welch 2008, Rapach et al. 2010, and Dangl and Halling 2012, among others). We collect monthly data on more than 63 pre-calculated financial ratios for all U.S. companies which can be classified in eight main categories: Valuation, Profitability, Capitalization, Financial Soundness, Solvency, Liquidity, Efficiency Ratios, and Other. Both returns and predictors are aggregated at the industry level by constructing value-weighted returns in excess of the risk-free rate and value-weighted averages of the single-firm predictors. Industry aggregation is based on the four-digit SIC codes of the existing firm at each time t. We use the ten industry classification codes obtained from Kenneth French's website.

Together with industry-specific predictors, we use additional 14 aggregate explanatory variables, which are divided into two additional categories; aggregate financials and macroeconomic variables. In particular, following Goyal and Welch (2008) and Rapach et al. (2010), the aggregate financial predictors consist of the monthly realized volatility of the value-weighted market portfolio (svar), the ratio of 12-month moving sums of net issues divided by the total end-of-year market capitalization (ntis), the default yield spread (dfy) calculated as the difference between BAA and AAA-rated corporate bond yields, and the term spread (tms) calculated as the difference between the long term yield on government bonds and the Treasury-bill. Additionally, we consider the traded liquidity factor (liq) of Pástor and Stambaugh (2003), and the year-on-year growth rate of the amount of loans and leases in Bank credit for all commercial banks.

For the aggregate macroeconomic predictors, we utilize the inflation rate (infl), measured as the monthly growth rate of the CPI All Urban Consumers index, the real interest rate (rit) measured as the return on the treasury bill minus inflation rate, the year-on-year growth rate of the initial claims for unemployment (icu), the year-on-year growth rate of the new private housing units authorized by building permits (house), the year-on-year growth of aggregate industrial production (ip), the year-on-year growth of the manufacturers' new orders (mno), the M2 monetary aggregate growth (M2), and the year-on-year growth of the consumer confidence index (conf) based on a survey of 5,000 US households.

The empirical application is conducted similar to the forecasting of U.S. inflation (see Figure 2). More precisely, first, the decoupled models are analyzed in parallel over 1970:01-1992:09 as a training period, simply estimating the DLM in Eq. (9) to the end of that period to estimate the forecasts from each group of predictors. This continues over 1993:07-2015:12, but with the calibration of recouple strategies, which, at each time t during this period, is run with the MCMC-based DRS analysis using data from 1993:07 up to time t. We discard the forecast results from 1993:07-2000:12 as training data and compare predictive performance from 2001:01-2015:12. The time frame includes key periods, such as the early 2000s– marked by the passing of the Gramm-Leach-Bliley act, the inflating and bursting of the dot.com bubble, the ensuing financial scandals such as Enron and Worldcom and the 9/11 attacks– and the great financial crisis of 2008–2009, which has been previously led by the burst of the sub-prime mortgage crisis (see, e.g., Bianchi, Guidolin, and Ravazzolo 2017a). Arguably, these periods exhibit sharp changes in financial markets, and more generally might lead to substantial biases and time variation in the latent inter-dependencies among relevant predictors.

Panel A of Table 2 shows that our decouple-recouple strategy improves the out-of-sample forecasting accuracy relative to the group-specific models, LASSO, PCA, equal-weight averaging, and BMA. Consistent with previous literature, the recursively computed equal-weighted linear-pooling is a challenging benchmark to beat by a large margin (see, e.g., Diebold and Shin 2017). The performance gap between Equal Weight and DRS is not as significant compared to others across industries. The out-of-sample performance of the LASSO and PCA are worse than other competing model combination schemes as well as the historical average (HA). These results hold for all the ten industries under investigation.

The outperformance of DRS is quite luminous related to the log predictive density ratios. In fact, as seen in Panel B of Table 2, none of the alternative specifications come close to DRS when it comes to predicting one-step ahead. With the only partial exception of the Energy sector, DRS strongly outperforms both the competing model combination/shrinkage schemes and the group-specific predictive densities.

Two comments are in order. First, while both the equal-weight linear pooling and the sequential BMA tend to outperform the group-specific predictive regressions, the LASSO strongly underperforms when it comes to predicting the density of future excess returns. This result is consistent with the recent evidence in Diebold and Shin (2017). They show that simple average combination schemes are highly competitive with respect to standard LASSO shrinkage algorithm. In particular, they show that good out-of-sample performances are hard to achieve in real-time forecasting exercise, due to the intrinsic difficulty of small-sample real-time cross-validation of the LASSO tuning parameter.

Delving further into the dynamics of our DRS predictions, Figure 5 shows the posterior mean estimates of the latent interdependencies among predictive densities which have been rescaled by using a logistic transformation as in Eq.(16). For the ease of exposition we report the results for a handful of industries, namely Consumer Durable, Consumer Non-Durable, Manufacturing and Other. The posterior estimates for the other industries are available upon request.

Although the interpretation of the dynamics of the latent inter-dependencies is not always clean, some interesting picture emerge. First, there is a substantial time variation in the inter-dependencies among predictive densities. In particular, abrupt changes in the relative effects of groups of predictors can be identified around the great financial crisis, especially for the Manufacturing and Other industries. This is likely not due to idiosyncratic volatility effects, as we explicitly take into account time varying volatility for the unexpected returns for each of the group-specific regressions (see Eq. 9). Second, the "weight" of aggregate financials on the aggregate predictive density tend to increase over time for all industries with a rather stable upward trend. Third, the fact that we impose a random-walk dynamics to the latent interdependencies does not prevent the predictive synthesis to be stable over time. Indeed, the posterior estimates of $\mathbf{\Phi}_t$ for Consumer Non-Durables are rather stable throughout the evaluation sample. Fourth, the role of Value and Financial Soundness is highly significant in predictive stock returns, with substantial fluctuations and differences around the great financial crisis of 2008–2009. Financial Soundness indicators involve variables such as cash flow over total debt, short-term debt over total debt, current liabilities over total liabilities, long-term debt over book equity, and long-term debt over total liabilities, among others. These variables arguably capture a company's risk level in the medium-to-long term as evaluated in relation to the company's debt level, and therefore collectively capture the ability of a company to manage its outstanding debt effectively to keep its operations. Quite understandably, the interplay between debt (especially medium term debt) and market value increasingly affect risk premia, and therefore the predicted value of future excess returns in a significant manner.

The time variation in the latent inter-dependencies is reflected in the aggregate dynamic bias, which is sequentially corrected within our decouple-recouple dynamic predictive framework. Figure 6 shows the dynamics of the calibrated bias across different industries. The figure makes clear that there is a substantial change in the aggregate bias in the aftermath of both the dot.com bubble and the great financial crisis. Which is to say, the aggregate predictive density that is synthesized from each class of predictors is significantly recalibrated around periods of market turmoil. Finally, one comment is in order. It should be clear that our goal in this paper is not to over-throw other results from the empirical finance literature with respect to the correlation among predictors and/or the misspecification of others modeling frameworks, but to deal with two crucial aspects of in dynamic forecasting of the equity premium: (1) capture the dynamic interplay between different, economically motivated, predictive densities, and (2) sequentially learn and correct for eventual models' misspecification.

4.2.1 Economic Significance. We now investigate the economic gains obtained by using our DRS strategy as opposed to one of the competing predictive strategies. In particular, we take the perspective often used in returns predictability studies of a representative investor with power utility and moderate risk aversion, i.e., $\gamma = 5$ (see, e.g., Barberis 2000, Johannes, Korteweg, and Polson 2014, Pettenuzzo et al. 2014, and Pettenuzzo and Ravazzolo 2016). Panel A of Table 3 shows the results for portfolios with unconstrained weights, which means short sales are allowed to maximize the portfolio returns. In particular, we report the CER of a competing strategy relative to the benchmark DRS as obtained from Eq.(15).

The economic performance of our decouple-recouple strategy is rather stark in contrast to both group-specific forecasts and the competing dimension reduction and forecasts combination schemes. The realized CER from DRS is substantially larger than any of the other model specifications across different industries. Not surprisingly, given that the statistical accuracy of a simple recursive historical mean model is not remarkable, the HA model leads to a very low CER. The results show that there is substantial economic evidence of returns predictability: a representative investor using our predictive strategy could have earned consistently positive utility gains across different U.S. industries relative to an investor using the historical mean. Interestingly, the equally-weighted linear pooling and Bayesian model averaging turn out to be both strong competitors, although still generate lower CERs.

Panel B of Table 3 shows that the performance gap in favor of DRS is confirmed under the restriction that the portfolio weights have to be positive, i.e., long-only strategy. Our predictive strategy generates a larger performance than BMA and equal-weight linear pooling. Notably, both the performance of other benchmark strategies such as the LASSO and dynamic PCA substantially improve by imposing no-short sales constraints.

In addition to the full sample evaluation above, we also study how the different models perform in real time. Specifically, we first calculate the $CER_{i\tau}$ at each time τ as

$$CER_{i\tau} = \left[\frac{\hat{U}_{\tau,i}}{\hat{U}_{\tau}}\right]^{\frac{1}{1-\gamma}} - 1, \tag{17}$$

Similarly to Eq (15), a negative $CER_{i\tau}$ can be interpreted as evidence that model i generates a lower (certainty equivalent) return at time τ than our DRS strategy. Panel A of Table 4 shows the average, annualized, single-period CER for an unconstrained investor. The results show that the out-of-sample performance is robustly in favor of the DRS model-combination scheme. As for the whole-sample results reported in Table 3, the equal-weighted linear pooling turns out to be a challenging benchmark to beat. Yet, DRS generates constantly higher average CERs throughout the sample.

Panel B shows the results for a short-sales constrained investor. Although the gap between DRS and the competing forecast combination schemes is substantially reduced, DRS robustly generates higher performances in the order of 10 to 40 basis points, depending on the industry and the competing strategy. As a whole, Tables 3-4 suggest that by sequentially learning latent interdependencies and biases improve the out-of-sample economic performance within the context of a typical portfolio allocation example.

5 Conclusion

In this paper, we propose a framework for predictive modeling when the decision maker is confronted with a large number of predictors. Our new approach retains all of the information available by first decoupling a large predictive model into a set of smaller predictive regressions, which are constructed by similarity among classes of predictors, then recoupling them by treating each of the subgroup of predictors as latent states; latent states, which are learned and calibrated via Bayesian updating, to understand the latent inter-dependencies and biases. These inter-dependencies and biases are then effectively mapped onto a latent dynamic factor model, in order to provide the decision maker with a dynamically updated forecast of the quantity of interest.

This is a drastically different approach from the literature where there were mainly two strands of development; shrinking the set of active regressors by imposing regularization and sparsity, e.g., LASSO and ridge regression, or assuming a small set of factors can summarize the whole information in an unsupervised manner, e.g., PCA and factor models.

We implement and evaluate the proposed methodology on both a macroeconomic and a finance application. We compare forecasts from our framework against a variety of standard sparse and dense modeling benchmarks used in finance and macroeconomics within a linear regression context. Irrespective of the performance evaluation metric, our decouple-recouple model synthesis scheme emerges as the best for forecasting both the annual inflation rate for the U.S. economy as well as the equity premium for different industries in the U.S.

References

Aastveit, K. A., K. R. Gerdrup, A. S. Jore, and L. A. Thorsrud. 2014. Nowcasting GDP in real time: A density combination approach. *Journal of Business & Economic Statistics* 32:48–68.

Aastveit, K. A., F. Ravazzolo, and H. K. Van Dijk. 2016. Combined density nowcasting in an uncertain economic environment. *Journal of Business & Economic Statistics* pp. 1–42.

Adrian, T., and F. Franzoni. 2009. Learning about beta: Time-varying factor loadings, expected returns, and the conditional CAPM. *Journal of Empirical Finance* pp. 537–556.

Avramov, D. 2004. Stock return predictability and asset pricing models. *Review of Financial Studies* 17:699–738.

Bai, J., and S. Ng. 2002. Determining the number of factors in approximate factor models. *Econometrica* 70:191–221.

Bai, J., and S. Ng. 2010. Instrumental variable estimation in a data rich environment. *Econometric Theory* 26:1577–1606.

Barberis, N. 2000. Investing for the long-run when returns are predictable. *The Journal of Finance* 55:225–264.

Bernanke, B. S., J. Boivin, and P. Eliasz. 2005. Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach. *The Quarterly journal of economics* 120:387–422.

Bianchi, D., M. Guidolin, and F. Ravazzolo. 2017a. Dissecting the 2007–2009 real estate market bust: Systematic pricing correction or just a housing fad? *Journal of Financial Econometrics* 16:34–62.

Bianchi, D., M. Guidolin, and F. Ravazzolo. 2017b. Macroeconomic factors strike back: A Bayesian change-point model of time-varying risk exposures and premia in the US cross-section. *Journal of Business & Economic Statistics* 35:110–129.

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

Binsbergen, V., H. Jules, and R. S. Koijen. 2010. Predictive regressions: A present-value approach. *The Journal of Finance* 65:1439–1471.

Campbell, J. Y., and S. B. Thompson. 2007. Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies* 21:1509–1531.

Chan, J. C. 2017. The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling. *Journal of Business & Economic Statistics* 35:17–28.

Chen, X., K., D. Banks, R. Haslinger, J. Thomas, and M. West. 2017. Scalable Bayesian modeling, monitoring and analysis of dynamic network flow data. *Journal of the American Statistical Association* Forthcoming.

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

Clemen, R. T. 1989. Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting* 5:559–583.

Cogley, T., and T. J. Sargent. 2005. Drifts and volatilities: Monetary policies and outcomes in the post WWII U.S. *Review of Economic Dynamics* 8:262–302.

Dangl, T., and M. Halling. 2012. Predictive regressions with time-varying coefficients. *Journal of Financial Economics* 106:157–181.

De Mol, C., D. Giannone, and L. Reichlin. 2008. Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics* 146:318–328.

Diebold, F. X., and J. A. Lopez. 1996. Forecast evaluation and combination. *Handbook of Statistics* 14:241–268.

Diebold, F. X., and M. Shin. 2017. Beating the simple average: Egalitarian LASSO for combining economic forecasts .

Elliott, G., and A. Timmermann. 2004. Optimal forecast combinations under general loss functions and forecast error distributions. *Journal of Econometrics* 122:47–79.

Frühwirth-Schnatter, S. 1994. Data augmentation and dynamic linear models. *Journal of Time Series Analysis* 15:183–202.

Genest, C., and M. J. Schervish. 1985. Modelling expert judgements for Bayesian updating. *Annals of Statistics* 13:1198–1212.

Genre, V., G. Kenny, A. Meyler, and A. Timmermann. 2013. Combining expert forecasts: Can anything beat the simple average? *International Journal of Forecasting* 29:108–121.

George, E. I., and R. E. McCulloch. 1993. Variable selection via Gibbs sampling. *Journal of the American Statistical Association* 88:881–889.

Geweke, J., and G. G. Amisano. 2012. Prediction with misspecified models. *The American Economic Review* 102:482–486.

Geweke, J. F., and G. G. Amisano. 2011. Optimal prediction pools. *Journal of Econometrics* 164:130–141.

Giannone, D., M. Lenza, and G. Primiceri. 2017. Economic predictions with big data: The illusion of sparsity. *Working Paper* .

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

Gruber, L. F., and M. West. 2016. GPU-accelerated Bayesian learning in simultaneous graphical dynamic linear models. *Bayesian Analysis* 11:125–149.

Gruber, L. F., and M. West. 2017. Bayesian forecasting and scalable multivariate volatility analysis using simultaneous graphical dynamic linear models. *Econometrics and Statistics* (published online March 12). ArXiv:1606.08291.

Harrison, P. J., and C. F. Stevens. 1976. Bayesian forecasting. *Journal of the Royal Statistical Society (Series B: Methodological)* 38:205–247.

Harvey, C. R., Y. Liu, and H. Zhu. 2016. ... and the cross-section of expected returns. *The Review of Financial Studies* 29:5–68.

Johannes, M., A. Korteweg, and N. Polson. 2014. Sequential learning, predictability, and optimal portfolio returns. *The Journal of Finance* 69:611–644.

Jostova, G., and A. Philipov. 2005. Bayesian Analysis of Stochastic Betas. *Journal of Financial and Quantitative Analysis* 40:747–778.

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

Koop, G., and D. Korobilis. 2010. Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics* 3:267–358.

Koop, G., R. Leon-Gonzalez, and R. W. Strachan. 2009. On the evolution of the monetary policy transmission mechanism. *Journal of Economic Dynamics and Control* 33:997–1017.

Lewellen, J. 2004. Predicting returns with financial ratios. *Journal of Financial Economics* 74:209–235.

Makridakis, S. 1989. Why combining works? *International Journal of Forecasting* 5:601–603.

Manzan, S. 2015. Forecasting the distribution of economic variables in a data-rich environment. *Journal of Business & Economic Statistics* 33:144–164.

McAlinn, K., and M. West. 2017. Dynamic Bayesian predictive synthesis in time series forecasting. *Journal of Econometrics* Forthcoming.

McCracken, M. W., and S. Ng. 2016. FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics* 34:574–589.

Meinshausen, N., B. Yu, et al. 2009. Lasso-type recovery of sparse representations for high-dimensional data. *The Annals of Statistics* 37:246–270.

Nakajima, J., and M. West. 2013. Bayesian analysis of latent threshold dynamic models. *Journal of Business & Economic Statistics* 31:151–164.

Nardari, F., and J. Scruggs. 2007. Bayesian analysis of linear factor models with latent factors, multivariate stochastic volatility, and APT pricing restrictions. *Journal of Financial and Quantitative Analysis* 42:857–891.

Pastor, L., and F. Stambaugh, R. 2009. Predictive systems: Living with imperfect predictors. *The Journal of Finance* pp. 1583–1628.

Pastor, L., and F. Stambaugh, R. 2012. Are stocks really less volatile in the long-run? *The Journal of Finance* pp. 431–477.

Pástor, L., and R. F. Stambaugh. 2003. Liquidity risk and expected stock returns. *Journal of Political Economy* 111:642–685.

Pettenuzzo, D., and F. Ravazzolo. 2016. Optimal portfolio choice under decision-based model combinations. *Journal of Applied Econometrics* 31:1312–1332.

Pettenuzzo, D., A. Timmermann, and R. Valkanov. 2014. Forecasting stock returns under economic constraints. *Journal of Financial Economics* 114:517–553.

Prado, R., and M. West. 2010. *Time Series: Modelling, Computation & Inference*. Chapman & Hall/CRC Press.

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

Rapach, D., J. Strauss, and G. Zhou. 2010. Out-of-sample equity prediction: Combination forecasts and links to the real economy. *The Review of Financial Studies* 23:822–862.

Shao, J. 1993. Linear model selection by cross-validation. *Journal of the American Statistical Association* 88:486–494.

Smith, J., and K. F. Wallis. 2009. A simple explanation of the forecast combination puzzle. *Oxford Bulletin of Economics and Statistics* 71:331–355.

Stock, J. H., and M. W. Watson. 2002. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97:1167–1179.

Stock, J. H., and M. W. Watson. 2004. Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting* 23:405–430.

Stock, J. H., and M. W. Watson. 2007. Why has US inflation become harder to forecast? *Journal of Money, Credit and Banking* 39:3–33.

Timmermann, A. 2004. Forecast combinations. In G. Elliott, C. W. J. Granger, and A. Timmermann (eds.), *Handbook of Economic Forecasting*, vol. 1, chap. 4, pp. 135–196. North Holland.

West, M. 1984. Bayesian aggregation. *Journal of the Royal Statistical Society (Series A: General)* 147:600–607.

West, M. 1992. Modelling agent forecast distributions. *Journal of the Royal Statistical Society (Series B: Methodological)* 54:553–567.

West, M., and J. Crosse. 1992. Modelling of probabilistic agent opinion. *Journal of the Royal Statistical Society (Series B: Methodological)* 54:285–299.

West, M., and P. J. Harrison. 1997. *Bayesian Forecasting & Dynamic Models*. 2nd ed. Springer Verlag.

Table 1. Out-of-sample forecast performance: Forecasting inflation.

This table reports the out-of-sample comparison of our decouple-recouple framework against each individual model, LASSO, PCA, equal weight average of models, and BMA for inflation forecasting. Performance comparison is based on the Root Mean Squared Error (RMSE), and the Log Predictive Density Ratio (LPDR) as in Eq. (10). The testing period is 2001/1-2015/12, monthly.

Panel A: Forecasting 1-Step Ahead Inflation

	Group-Specific Models								LASSO	PCA	EW	BMA	DRS
	Output & Income	Labor Market	Consump.	Orders & Invent.	Money & Credit	Int. Rate & Ex. Rates	Prices	Stock Market					
RMSE	0.2488	0.2247	0.7339	0.2721	0.2624	0.4258	0.2223	0.5027	0.3348	0.9329	0.2945	0.2721	0.2051
(%)	-7.35%	-7.37%	-122.06%	-8.73%	-15.75%	-40.56%	-6.83%	-59.59%	-63.24%	-354.85%	-43.59%	-32.68%	-
LPDR	-40.48	-42.05	-233.09	-59.15	-56.34	-134.18	-20.00	-171.21	-3785.15	-285.41	-88.81	-60.40	-

Panel B: Forecasting 3-Step Ahead Inflation

	Group-Specific Models								LASSO	PCA	EW	BMA	DRS
	Output & Income	Labor Market	Consump.	Orders & Invent.	Money & Credit	Int. Rate & Ex. Rates	Prices	Stock Market					
RMSE	0.3594	0.3595	0.7435	0.3640	0.3875	0.4706	0.3577	0.5343	0.3991	0.9223	0.3777	0.3640	0.3348
(%)	-21.32%	-9.57%	-257.86%	-32.68%	-27.95%	-107.66%	-8.39%	-145.14%	-19.21%	-175.45%	-12.87%	-8.73%	-
LPDR	-78.65	-225.75	-156.59	-61.96	-122.27	-77.76	-101.55	-101.82	-3804.35	-203.12	-41.00	-78.54	-

Panel C: Forecasting 12-Step Ahead Inflation

	Group-Specific Models								LASSO	PCA	EW	BMA	DRS
	Output & Income	Labor Market	Consump.	Orders & Invent.	Money & Credit	Int. Rate & Ex. Rates	Prices	Stock Market					
RMSE	0.5685	0.5969	0.7672	0.5679	0.6280	0.6307	0.6289	0.6075	0.6373	0.8815	0.5653	0.6280	0.4187
(%)	-35.78%	-42.57%	-83.26%	-35.64%	-50.00%	-50.64%	-50.21%	-45.10%	-52.23%	-110.56%	-35.01%	-50.00%	-
LPDR	-325.32	-941.58	-238.38	-363.58	-531.34	-248.71	-632.56	-181.60	-4112.97	-201.17	-649.90	-264.78	-

Table 2. Out-of-sample forecast performance: Forecasting Stock Industry Returns.

This table reports the out-of-sample comparison of our predictive framework against standard model combination methodologies, across ten different industries. Performance comparison is based on the Root Mean Squared Error (RMSE), and the Log Predictive Density Ratio (LPDR) as in Eq. (10). We report the results obtained for each of the group-specific predictors, the results obtained by simply taking the historical average of the stock returns (HA), and the results from a set of competing model combination/shrinkage schemes, e.g., LASSO, PCA, Equal Weight, and Bayesian Model Averaging (BMA). The sample period is 01:1970-12:2015, monthly.

Panel A: Root Mean Squared Error

Industry		Group-Specific Models										LASSO	EW	BMA	PCA	DRS
	HA	Value	Profit	Capital	Soundness	Solvency	Liquidity	Efficiency	Other	Aggregate Fin	Macro					
Durbl	0.308	0.193	0.208	0.260	0.228	0.225	0.238	0.292	0.267	0.248	0.205	0.245	0.169	0.193	0.366	0.154
NoDurbl	0.129	0.092	0.108	0.115	0.097	0.109	0.113	0.108	0.111	0.100	0.103	0.119	0.076	0.090	0.121	0.072
Manuf	0.197	0.130	0.140	0.175	0.165	0.162	0.182	0.137	0.180	0.152	0.150	0.140	0.110	0.125	0.174	0.103
Energy	0.214	0.123	0.151	0.176	0.155	0.142	0.166	0.181	0.197	0.142	0.170	0.188	0.121	0.123	0.173	0.105
HiTech	0.289	0.139	0.231	0.206	0.184	0.212	0.185	0.192	0.238	0.231	0.220	0.197	0.143	0.139	0.223	0.121
Health	0.171	0.104	0.099	0.127	0.097	0.109	0.118	0.101	0.111	0.108	0.098	0.140	0.083	0.094	0.115	0.080
Shops	0.232	0.180	0.159	0.182	0.131	0.157	0.163	0.157	0.194	0.175	0.148	0.171	0.114	0.131	0.235	0.108
Telecomm	0.157	0.096	0.101	0.128	0.107	0.119	0.118	0.120	0.119	0.102	0.098	0.124	0.082	0.096	0.222	0.072
Utils	0.261	0.142	0.139	0.172	0.134	0.159	0.179	0.129	0.196	0.172	0.144	0.178	0.106	0.121	0.190	0.098
Other	0.173	0.112	0.110	0.132	0.117	0.143	0.136	0.114	0.142	0.128	0.123	0.146	0.091	0.097	0.143	0.083

Panel B: Log-Predictive Density Ratio

Industry		Group-Specific Models										LASSO	EW	BMA	PCA	DRS
	HA	Value	Profit	Capital	Soundness	Solvency	Liquidity	Efficiency	Other	Aggregate Fin	Macro					
Durbl	-44.55	-45.83	-51.75	-107.39	-69.84	-82.57	-91.23	-124.08	-114.72	-64.51	-59.86	-91.93	-146.55	-341.35	-135.55	-
NoDurbl	-109.50	-44.08	-60.24	-91.38	-55.51	-81.21	-90.07	-73.37	-87.24	-45.40	-67.43	-228.77	-173.04	-334.41	-92.57	-
Manuf	-36.34	-35.87	-53.78	-108.57	-53.15	-95.02	-113.86	-66.02	-123.30	-51.07	-72.63	-167.35	-85.53	-232.15	-113.60	-
Energy	-56.17	-26.65	-52.81	-96.36	-42.06	-52.94	-86.32	-80.02	-118.86	-45.25	-84.58	-150.12	-113.92	-281.18	-89.40	-
HiTech	-179.58	-17.83	-75.16	-110.60	-69.41	-100.94	-82.43	-87.47	-134.93	-110.40	-106.30	-138.79	-257.24	-848.69	-124.59	-
Health	-124.31	-46.58	-27.30	-77.57	-26.73	-53.59	-65.62	-41.45	-63.09	-49.76	-32.44	-200.14	-275.66	-437.22	-60.41	-
Shops	-81.20	-53.20	-56.88	-99.89	-27.24	-75.51	-90.22	-75.54	-117.14	-65.67	-61.57	-161.13	-166.28	-468.83	-108.89	-
Telecomm	-112.92	-42.74	-60.49	-99.38	-69.44	-83.03	-85.85	-88.10	-82.07	-51.55	-51.25	-215.60	-174.92	-275.45	-100.66	-
Utils	-165.95	-64.61	-49.29	-98.94	-45.90	-85.61	-107.23	-47.20	-124.94	-74.01	-65.39	-162.67	-203.81	-410.01	-123.01	-
Other	-115.31	-50.11	-29.38	-86.38	-37.23	-99.04	-99.83	-46.18	-102.96	-71.07	-73.01	-197.83	-128.33	-314.82	-94.06	-

Table 3. Out-of-sample economic performance for stock industry returns: Certainty equivalent returns

This table reports the out-of-sample comparison of our predictive framework against standard model combination methodologies, across ten different industries. Performance comparison is based on the Certainty Equivalent (CER), and its modification whereby short sales are not allowed. We report the results obtained for each of the group-specific predictors, the results obtained by simply taking the historical average of the stock returns (HA), and the results from a set of competing model combination/shrinkage schemes, e.g., LASSO, PCA, Equal Weight, and Bayesian Model Averaging (BMA). The sample period is 01:1970-12:2015, monthly.

Panel A: Certainty Equivalent

Industry		Group-Specific Models										LASSO	EW	BMA	PCA
	HA	Value	Profit	Capital	Soundness	Solvency	Liquidity	Efficiency	Other	Aggregate Fin	Macro				
Durbl	-0.084	-0.047	-0.032	-0.973	-0.075	-0.063	-0.306	-0.863	-0.342	-0.067	-0.082	-0.721	-0.021	-0.047	-0.996
NoDurbl	-0.195	-0.294	-0.114	-0.608	-0.097	-0.211	-0.163	-0.800	-0.211	-0.119	-0.149	-0.548	-0.125	-0.261	-0.719
Manuf	-0.126	-0.881	-0.101	-0.141	-0.465	-0.206	-0.099	-0.765	-0.101	-0.005	-0.210	-0.544	-0.019	-0.840	-0.983
Energy	-0.141	-0.074	-0.118	-0.092	-0.055	-0.099	-0.102	-0.915	-0.180	-0.050	-0.201	-0.990	-0.032	-0.005	-0.651
HiTech	-0.149	-0.029	-0.079	-0.165	-0.077	-0.135	-0.119	-0.082	-0.257	-0.152	-0.113	-0.203	-0.070	-0.029	-0.361
Health	-0.133	-0.023	-0.024	-0.112	-0.044	-0.044	-0.142	-0.024	-0.072	-0.044	-0.030	-0.163	-0.038	-0.017	-0.989
Shops	-0.150	-0.045	-0.070	-0.414	-0.039	-0.496	-0.770	-0.044	-0.107	-0.059	-0.154	-0.498	-0.059	-0.040	-0.746
Telecomm	-0.174	-0.067	-0.097	-0.208	-0.104	-0.106	-0.104	-0.426	-0.077	-0.078	-0.050	-0.864	-0.081	-0.067	-0.749
Utils	-0.216	-0.104	-0.078	-0.462	-0.072	-0.122	-0.185	-0.160	-0.160	-0.103	-0.098	-0.474	-0.109	-0.075	-0.439
Other	-0.096	-0.009	0.031	-0.038	0.022	-0.105	-0.089	0.038	-0.042	-0.019	-0.044	-0.657	0.016	0.055	-0.433

Panel B: Certainty Equivalent (no-short sales)

Industry		Group-Specific Models										LASSO	EW	BMA	PCA
	HA	Value	Profit	Capital	Soundness	Solvency	Liquidity	Efficiency	Other	Aggregate Fin	Macro				
Durbl	-0.045	-0.007	-0.003	-0.024	-0.010	-0.021	-0.011	-0.012	-0.014	-0.005	-0.015	-0.018	-0.005	-0.007	-0.021
NoDurbl	-0.031	-0.005	-0.012	-0.016	-0.010	-0.010	-0.020	-0.013	-0.013	-0.007	-0.016	-0.032	-0.006	-0.004	-0.010
Manuf	-0.068	-0.006	-0.010	-0.047	0.000	-0.028	-0.032	-0.005	-0.027	-0.007	-0.009	-0.017	-0.006	-0.006	-0.017
Energy	-0.067	-0.001	-0.016	-0.013	-0.003	-0.011	-0.019	-0.011	-0.037	-0.008	-0.013	-0.032	-0.004	0.003	-0.018
HiTech	-0.068	-0.004	-0.014	-0.023	-0.013	-0.036	-0.021	-0.031	-0.053	-0.046	-0.037	-0.042	-0.016	-0.004	-0.029
Health	-0.041	-0.003	-0.005	-0.021	-0.003	-0.009	-0.007	-0.006	-0.010	-0.008	-0.005	-0.027	-0.005	-0.004	-0.005
Shops	-0.058	-0.010	-0.013	-0.064	-0.005	-0.014	-0.023	-0.007	-0.021	-0.015	-0.009	-0.024	-0.009	-0.004	-0.025
Telecomm	-0.042	-0.009	-0.006	-0.016	-0.009	-0.006	-0.009	-0.011	-0.003	-0.010	-0.011	-0.023	-0.005	-0.009	-0.013
Utils	-0.070	-0.011	-0.011	-0.019	-0.010	-0.014	-0.032	-0.006	-0.029	-0.010	-0.010	-0.021	-0.011	-0.006	-0.027
Other	-0.055	-0.009	-0.006	-0.017	-0.009	-0.018	-0.017	-0.005	-0.023	-0.015	-0.007	-0.037	-0.007	-0.005	-0.015

Table 4. Out-of-sample economic performance for stock industry returns: Average single-period certainty equivalent returns

This table reports the out-of-sample comparison of our predictive framework against standard model combination methodologies, across ten different industries. Performance comparison is based on the single-period Certainty Equivalent (CER) (see Eq. (17)), and its modification whereby short sales are not allowed. We report the results obtained for each of the group-specific predictors, the results obtained by simply taking the historical average of the stock returns (HA), and the results from a set of competing model combination/shrinkage schemes, e.g., LASSO, PCA, Equal Weight, and Bayesian Model Averaging (BMA). The sample period is 01:1970-12:2015, monthly.

Panel A: Average Single-Period Certainty Equivalent

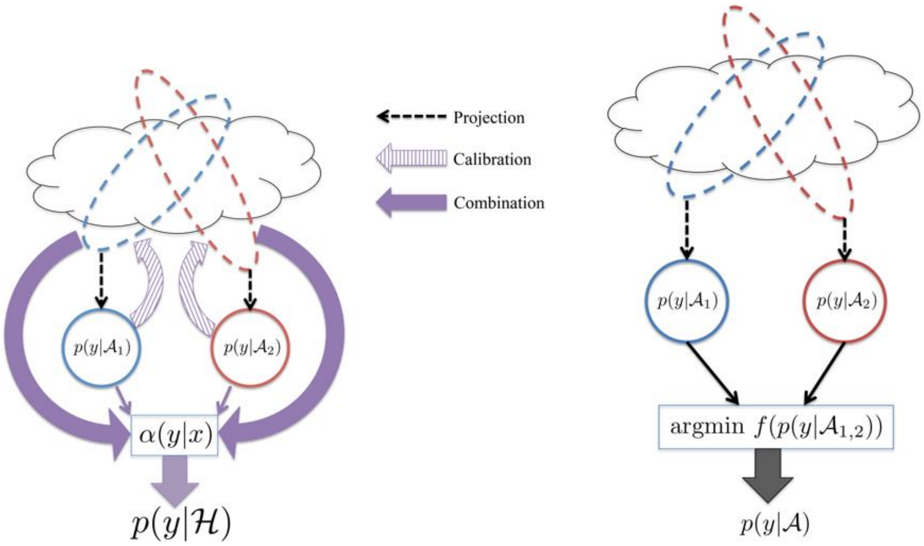
Industry		Group-Specific Models										LASSO	EW	BMA	PCA
	HA	Value	Profit	Capital	Soundness	Solvency	Liquidity	Efficiency	Other	Aggregate Fin	Macro				
Durbl	-0.373	-0.092	-0.116	-0.300	-0.097	-0.200	-0.227	-0.280	-0.282	-0.145	-0.161	-0.070	-0.258	-0.092	-0.304
NoDurbl	-0.487	-0.266	-0.305	-0.430	-0.265	-0.367	-0.415	-0.365	-0.416	-0.288	-0.323	-0.271	-0.393	-0.266	-0.370
Manuf	-0.361	-0.070	-0.112	-0.255	-0.087	-0.229	-0.256	-0.164	-0.312	-0.133	-0.166	-0.048	-0.247	-0.085	-0.240
Energy	-0.444	-0.066	-0.116	-0.329	-0.067	-0.160	-0.241	-0.229	-0.380	-0.190	-0.216	-0.216	-0.283	-0.073	-0.205
HiTech	-0.550	-0.135	-0.363	-0.411	-0.286	-0.396	-0.396	-0.396	-0.497	-0.447	-0.426	-0.252	-0.439	-0.135	-0.430
Health	-0.461	-0.208	-0.192	-0.343	-0.164	-0.256	-0.306	-0.216	-0.308	-0.238	-0.175	-0.223	-0.293	-0.169	-0.282
Shops	-0.453	-0.158	-0.164	-0.348	-0.068	-0.222	-0.297	-0.192	-0.381	-0.256	-0.206	-0.141	-0.320	-0.077	-0.270
Telecomm	-0.614	-0.324	-0.388	-0.515	-0.378	-0.407	-0.472	-0.468	-0.434	-0.381	-0.358	-0.394	-0.475	-0.324	-0.478
Utils	-0.642	-0.319	-0.268	-0.476	-0.262	-0.415	-0.501	-0.304	-0.543	-0.426	-0.361	-0.268	-0.476	-0.278	-0.480
Other	-0.524	-0.204	-0.167	-0.347	-0.180	-0.397	-0.387	-0.222	-0.424	-0.349	-0.275	-0.168	-0.372	-0.157	-0.275

Panel B: Average Single-Period Certainty Equivalent (no-short sales)

Industry		Group-Specific Models										LASSO	EW	BMA	PCA
	HA	Value	Profit	Capital	Soundness	Solvency	Liquidity	Efficiency	Other	Aggregate Fin	Macro				
Durbl	-0.089	-0.009	-0.009	-0.040	-0.010	-0.026	-0.026	-0.031	-0.036	-0.011	-0.017	-0.027	-0.025	-0.009	-0.051
NoDubl	-0.022	-0.005	-0.013	-0.017	-0.008	-0.011	-0.021	-0.012	-0.014	-0.007	-0.016	-0.025	-0.007	-0.004	-0.012
Manuf	-0.041	-0.006	-0.010	-0.028	-0.001	-0.022	-0.028	-0.007	-0.030	-0.007	-0.010	-0.019	-0.011	-0.006	-0.023
Energy	-0.054	-0.003	-0.011	-0.014	-0.008	-0.010	-0.017	-0.014	-0.035	-0.009	-0.011	-0.033	-0.008	-0.002	-0.020
HiTech	-0.086	-0.003	-0.017	-0.032	-0.014	-0.027	-0.025	-0.024	-0.050	-0.035	-0.024	-0.043	-0.025	-0.003	-0.043
Health	-0.038	-0.003	-0.006	-0.021	-0.002	-0.009	-0.007	-0.005	-0.010	-0.007	-0.005	-0.027	-0.006	-0.004	-0.006
Shops	-0.065	-0.009	-0.012	-0.035	-0.004	-0.015	-0.027	-0.008	-0.030	-0.018	-0.010	-0.013	-0.003	-0.031	
Telecomm	-0.037	-0.008	-0.006	-0.016	-0.009	-0.006	-0.009	-0.013	-0.003	-0.010	-0.011	-0.021	-0.006	-0.008	-0.014
Utils	-0.087	-0.011	-0.009	-0.022	-0.008	-0.015	-0.040	-0.006	-0.033	-0.012	-0.012	-0.021	-0.015	-0.006	-0.032
Other	-0.059	-0.009	-0.006	-0.016	-0.010	-0.020	-0.017	-0.005	-0.020	-0.015	-0.007	-0.036	-0.008	-0.005	-0.013

Figure 1. Outline of the Methodology

This figure visually presents our strategy compared to standard combination strategies. Here, the cloud above is considered the data generating process, dotted ovals are the data generating processes of the agents' forecasts, the dotted lines are the agents' projections, and the solid circles are the observed agents' forecasts. In our predictive synthesis framework (left panel), the agent-specific predictive densities are calibrated based on latent inter-dependencies and biases (where the overlapping areas of dotted ovals are inter-dependencies and areas off the cloud are biases) and are combined using the synthesis function. Opposed to this, a standard model combination scheme (right panel) ignores the latent inter-dependencies and biases and minimizes a function of the observed agents' forecasts.



(a) Our Framework (b) Standard Combination

Figure 2. Timeline of the Inflation Forecasting Exercise

This figure visually presents the timeline of the inflation forecasting exercise by separating the train sample, the train and combine and the evaluation sample.

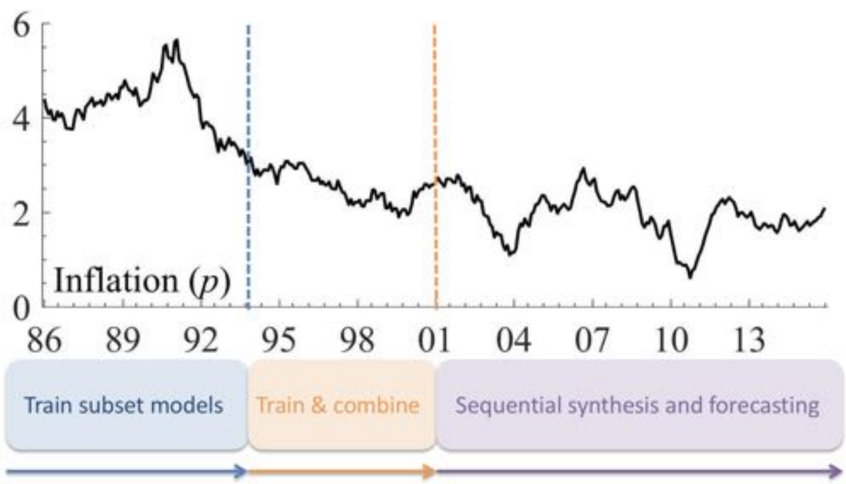
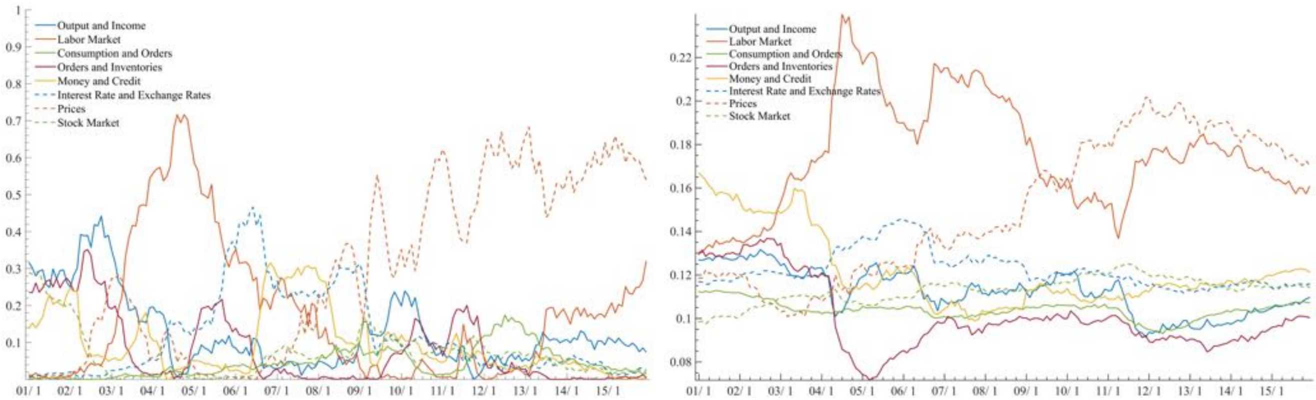


Figure 3. Posterior Means of Rescaled Latent Inter-Dependencies for the U.S. Inflation Forecasting

This figure shows the latent interdependencies across groups of predictive densities– measured through the predictive coefficients– used in the recoupling step for both the one- and three-month ahead forecasting exercise. For the sake of interpretability we report the rescaled coefficients which are normalized by using a logistic transformation.



(a) 1-step ahead (b) 3-step ahead

Figure 4. Out-of-Sample Dynamic Predictive Bias for U.S. Inflation Forecasting

This figure shows the dynamics of the out-of-sample predictive bias obtained as the time-varying intercept from the recoupling step of the DRS strategy. The sample evaluation period is 01:2001 to 12:2015.

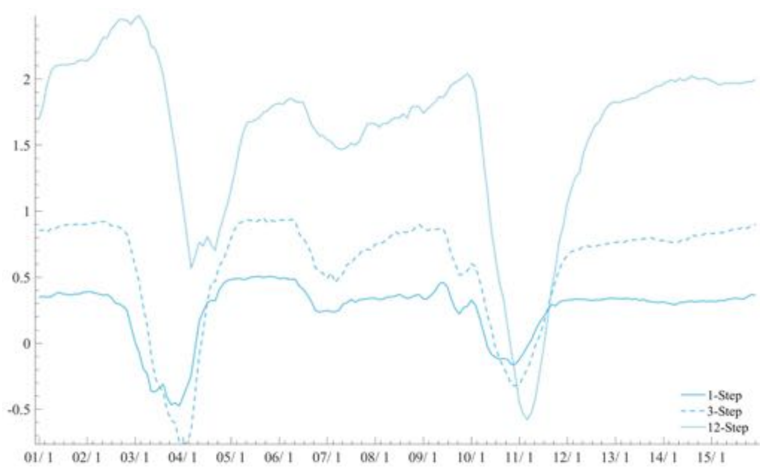
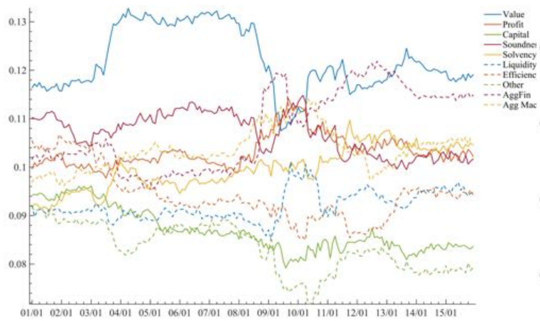
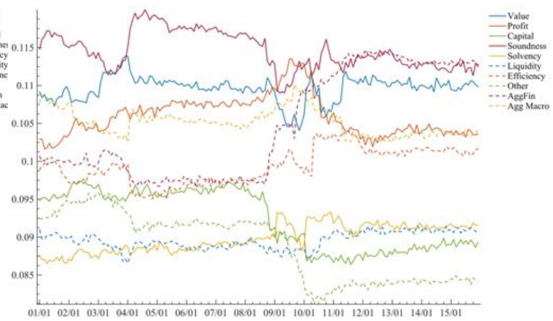


Figure 5. Posterior Means of Rescaled Latent Inter-Dependencies for the U.S. Industry Equity Premium

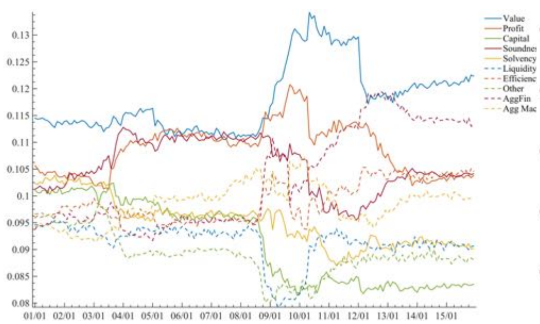
This figure shows the one-step ahead latent interdependencies across groups of predictive densities– measured through the predictive coefficients– used in the recoupling step. For the ease of exposition we report the results for four representative industries, namely, Consumer Durables, Consumer non-Durables, Manufacturing, Shops, Utils and Other. Industry aggregation is based on the four-digit SIC codes of the existing firm at each time t following the industry classification from Kenneth French's website. The sample period is 01:1970-12:2015, monthly.



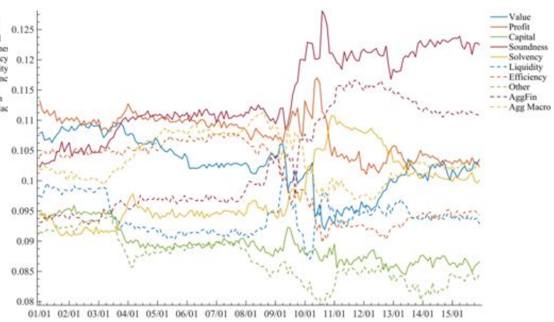
(a) Consumer Durable



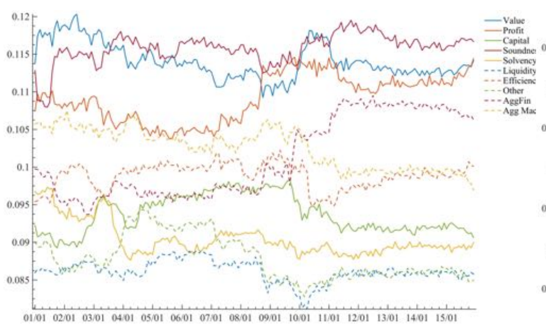
(b) Cons. Non-Durable



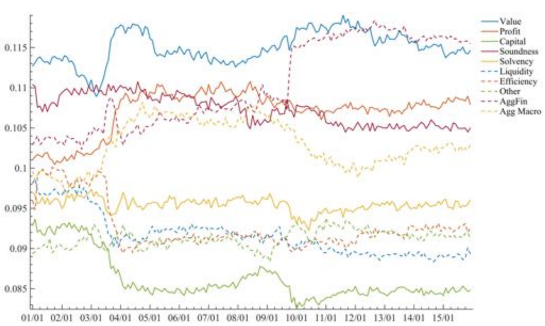
(c) Manufacturing



(d) Other



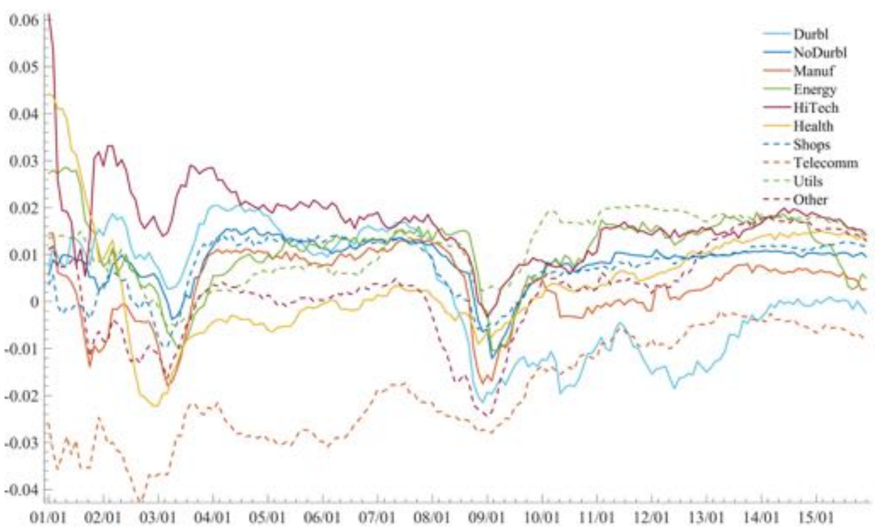
(e) Utils



(f) Shops

Figure 6. Out-of-Sample Dynamic Predictive Bias for the U.S. Industry Equity Premium

This figure shows the dynamics of the out-of-sample predictive bias obtained as the time-varying intercept from the recoupling step of the DRS strategy. The figure reports the results across all industries. The sample period is 01:2001-12:2015, monthly. The objective function is the one-step ahead density forecast of stock excess returns across different industries. Industry classification is based on 4-digit SIC codes.



Appendix for Online Publication : Large-Scale Dynamic Predictive Regressions

Outline

This Appendix provides additional details regarding our methodology, the estimation strategy, some test based on a simulated dataset, as well as some additional out-of-sample empirical results. Note that all notations and model definitions are similar to those in the main article.

A MCMC Algorithm

In this section we provide details of the Markov Chain Monte Carlo (MCMC) algorithm implemented to estimate the BPS recouple step. This involves a sequence of standard steps in a customized two-component block Gibbs sampler: the first component learns and simulates from the joint posterior predictive densities of the subgroup models; this the "learning" step. The second step samples the predictive synthesis parameters, that is we "synthesize" the models' predictions in the first step to obtain a single predictive density using the information provided by the subgroup models. The latter involves the FFBS algorithm central to MCMC in all conditionally normal DLMs (Frühwirth-Schnatter 1994; West and Harrison 1997, Sect 15.2; Prado and West 2010, Sect 4.5).

In our sequential learning and forecasting context, the full MCMC analysis is performed in an extending window manner, re-analyzing the data set as time and data accumilates. We detail MCMC steps for a specific time t here, based on all data up until that time point.

A.1 Initialization:

First, initialize by setting $\boldsymbol{F}_t = (1, x_{t1}, ..., x_{tJ})'$ for each $t = 1{:}T$ at some chosen initial values of the latent states. Initial values can be chosen arbitrarily, though following McAlinn and West (2017) we recommend sampling from the priors, i.e., from the forecast distributions, $x_{tj} \sim h_{tj}(x_{tj})$ independently for all

$t = 1{:}T$ and $j = 1{:}J$.

Following initialization, the MCMC iterates repeatedly to resample two coupled sets of conditional posteriors to generate the draws from the target posterior $p(\boldsymbol{x}_{1:T}, \boldsymbol{\Phi}_{1:T}|y_{1:T}, \mathcal{H}_{1:T})$. These two conditional posteriors and algorithmic details of their simulation are as follows.

A.2 Sampling the synthesis parameters $\boldsymbol{\Phi}_{1:T}$

Conditional on any values of the latent agent states, we have a conditionally normal DLM with known predictors. The conjugate DLM form,

$$\begin{aligned} y_t &= \boldsymbol{F}_t'\boldsymbol{\theta}_t + \nu_t, \quad \nu_t \sim N(0, v_t), \\ \boldsymbol{\theta}_t &= \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(0, v_t\boldsymbol{W}_t), \end{aligned}$$

has known elements $\boldsymbol{F}_t, \boldsymbol{W}_t$ and specified initial prior at $t = 0$. The implied conditional posterior for $\boldsymbol{\Phi}_{1:T}$ then does not depend on $\mathcal{H}_{1:T}$, reducing to $p(\boldsymbol{\Phi}_{1:T}|\boldsymbol{x}_{1:T}, y_{1:T})$. Standard Forward-Filtering Backward-Sampling algorithm can be applied to efficiently sample these parameters, modified to incorporate the discount stochastic volatility components for v_t (e.g. Frühwirth-Schnatter 1994; West and Harrison 1997, Sect 15.2; Prado and West 2010, Sect 4.5).

A.2.1 Forward filtering:

One step filtering updates are computed, in sequence, as follows:

1. *Time $t-1$ posterior:*

$$\begin{aligned} \boldsymbol{\theta}_{t-1}|v_{t-1}, \boldsymbol{x}_{1:t-1}, y_{1:t-1} &\sim N(\boldsymbol{m}_{t-1}, \boldsymbol{C}_{t-1}v_{t-1}/s_{t-1}), \\ v_{t-1}^{-1}|\boldsymbol{x}_{1:t-1}, y_{1:t-1} &\sim G(n_{t-1}/2, n_{t-1}s_{t-1}/2), \end{aligned}$$

with point estimates $\boldsymbol{m}_{t-1}$ of $\boldsymbol{\theta}_{t-1}$ and s_{t-1} of v_{t-1}.

2. *Update to time t prior:*

$$\boldsymbol{\theta}_t | v_t, \boldsymbol{x}_{1:t-1}, y_{1:t-1} \sim N(\boldsymbol{m}_{t-1}, \boldsymbol{R}_t v_t / s_{t-1}) \quad \text{with} \quad \boldsymbol{R}_t = \boldsymbol{C}_{t-1}/\delta,$$
$$v_t^{-1} | \boldsymbol{x}_{1:t-1}, y_{1:t-1} \sim G(\beta n_{t-1}/2, \beta n_{t-1} s_{t-1}/2),$$

with (unchanged) point estimates $\boldsymbol{m}_{t-1}$ of $\boldsymbol{\theta}_t$ and s_{t-1} of v_t, but with increased uncertainty relative to the time $t-1$ posteriors, where the level of increased uncertainty is defined by the discount factors.

3. *1-step predictive distribution:* $y_t | \boldsymbol{x}_{1:t}, y_{1:t-1} \sim T_{\beta n_{t-1}}(f_t, q_t)$ where

$$f_t = \boldsymbol{F}_t' \boldsymbol{m}_{t-1} \quad \text{and} \quad q_t = \boldsymbol{F}_t' \boldsymbol{R}_t \boldsymbol{F}_t + s_{t-1}.$$

4. *Filtering update to time t posterior:*

$$\boldsymbol{\theta}_t | v_t, \boldsymbol{x}_{1:t}, y_{1:t} \sim N(\boldsymbol{m}_t, \boldsymbol{C}_t v_t / s_t),$$
$$v_t^{-1} | \boldsymbol{x}_{1:t}, y_{1:t} \sim G(n_t/2, n_t s_t/2),$$

with defining parameters as follows:

i. For $\boldsymbol{\theta}_t | v_t : \boldsymbol{m}_t = \boldsymbol{m}_{t-1} + \boldsymbol{A}_t e_t$ and $\boldsymbol{C}_t = r_t(\boldsymbol{R}_t - q_t \boldsymbol{A}_t \boldsymbol{A}_t')$,

ii. For $v_t : n_t = \beta n_{t-1} + 1$ and $s_t = r_t s_{t-1}$,

based on 1-step forecast error $e_t = y_t - f_t$, the state adaptive coefficient vector (a.k.a. "Kalman gain") $\boldsymbol{A}_t = \boldsymbol{R}_t \boldsymbol{F}_t / q_t$, and volatility estimate ratio $r_t = (\beta n_{t-1} + e_t^2/q_t)/n_t$.

A.2.2 Backward sampling:

Having run the forward filtering analysis up to time T, the backward sampling proceeds as follows.

a. *At time T:* Simulate $\boldsymbol{\Phi}_T = (\boldsymbol{\theta}_T, v_T)$ from the final normal/inverse gamma posterior $p(\boldsymbol{\Phi}_T | \boldsymbol{x}_{1:T}, y_{1:T})$ as follows. First, draw v_T^{-1} from $G(n_T/2, n_T s_T/2)$, and then draw $\boldsymbol{\theta}_T$ from $N(\boldsymbol{m}_T, \boldsymbol{C}_T v_T / s_T)$.

b. *Recurse back over times* $t = T-1, T-2, \ldots, 0$: At time t, sample $\mathbf{\Phi}_t = (\boldsymbol{\theta}_t, v_t)$ as follows:

i. Simulate the volatility v_t via $v_t^{-1} = \beta v_{t+1}^{-1} + \gamma_t$ where γ_t is an independent draw from $\gamma_t \sim G((1-\beta)n_t/2, n_t s_t/2)$,

ii. Simulate the state $\boldsymbol{\theta}_t$ from the conditional normal posterior $p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, v_t, \boldsymbol{x}_{1:T}, y_{1:T})$ with mean vector $\boldsymbol{m}_t + \delta(\boldsymbol{\theta}_{t+1} - \boldsymbol{m}_t)$ and variance matrix $\boldsymbol{C}_t(1-\delta)(v_t/s_t)$.

A.3 Sampling the latent states $\boldsymbol{x}_{1:T}$

Conditional on the sampled values from the first step, the MCMC iterate completes with resampling of the posterior joint latent states from $p(\boldsymbol{x}_{1:t}|\mathbf{\Phi}_{1:t}, y_{1:t}, \mathcal{H}_{1:t})$. We note that $\boldsymbol{x}_t$ are conditionally independent over time t in this conditional distribution, with time t conditionals

$$p(\boldsymbol{x}_t|\mathbf{\Phi}_t, y_t, \mathcal{H}_t) \propto N(y_t|\boldsymbol{F}_t'\boldsymbol{\theta}_t, v_t) \prod_{j=1:J} h_{tj}(x_{tj}) \quad \text{where} \quad \boldsymbol{F}_t = (1, x_{t1}, x_{t2}, ..., x_{tJ})'. \tag{A.1}$$

Since $h_{tj}(x_{tj})$ has a density of $T_{n_{tj}}(h_{tj}, H_{tj})$, we can express this as a scale mixture of Normal, $N(h_{tj}, H_{tj})$, with $\boldsymbol{H}_t = \text{diag}(H_{t1}/\phi_{t1}, H_{t2}/\phi_{t2}, ..., H_{tJ}/\phi_{tJ})$, where ϕ_{tj} are independent over t, j with gamma distributions, $\phi_{tj} \sim G(n_{tj}/2, n_{tj}/2)$.

The posterior distribution for each $\boldsymbol{x}_t$ is then sampled, given ϕ_{tj}, from

$$p(\boldsymbol{x}_t|\mathbf{\Phi}_t, y_t, \mathcal{H}_t) = N(\boldsymbol{h}_t + \boldsymbol{b}_t c_t, \boldsymbol{H}_t - \boldsymbol{b}_t \boldsymbol{b}_t' g_t) \tag{A.2}$$

where $c_t = y_t - \theta_{t0} - \boldsymbol{h}_t'\boldsymbol{\theta}_{t,1:J}$, $g_t = v_t + \boldsymbol{\theta}_{t,1:J}'\boldsymbol{q}_t\boldsymbol{\theta}_{t,1:J}$, and $\boldsymbol{b}_t = \boldsymbol{q}_t\boldsymbol{\theta}_{t,1:J}/g_t$. Here, given the previous values of ϕ_{tj}, we have $\boldsymbol{H}_t = \text{diag}(H_{t1}/\phi_{t1}, H_{t2}/\phi_{t2}, ..., H_{tJ}/\phi_{tJ})$ Then, conditional on these new samples of $\boldsymbol{x}_t$, updated samples of the latent scales are drawn from the implied set of conditional gamma posteriors $\phi_{tj}|x_{tj} \sim G((n_{tj}+1)/2, (n_{tj}+d_{tj})/2)$ where $d_{tj} = (x_{tj} - h_{tj})^2/H_{tj}$, independently for each t, j. This is easily computed and then sampled independently for each 1:T to provide resimulated agent states over 1:T.

A.4 Forecasting

In terms of forecasting, at time t, we generate predictive distributions of the object of interest as follows: (i) For each sampled $\mathbf{\Phi}_t$ from the posterior MCMC above, draw v_{t+1} from its stochastic dynamics, and then $\boldsymbol{\theta}_{t+1}$ conditional on $\boldsymbol{\theta}_t, v_{t+1}$ from Eq.(7b)– this gives a draw $\mathbf{\Phi}_{t+1} = \{\boldsymbol{\theta}_{t+1}, v_{t+1}\}$ from $p(\mathbf{\Phi}_{t+1}|y_{1:t}, \mathcal{H}_{1:t})$; (ii) draw $\boldsymbol{x}_{t+1}$ via independent sampling from $h_{t+1,j}(x_{t+1,j})$, $(j = 1{:}J)$; (iii) conditional on the parameters and latent states draw y_{t+1} from Eq.(7a). Repeating, this generates a random sample from the 1-step ahead synthesized forecast distribution for time $t + 1$.

Forecasting over multiple horizons is often of equal or greater importance than 1-step ahead forecasting. However, forecasting over longer horizons is typically more difficult than over shorter horizons, since predictors that are effective in the short term might not be effective in the long term. Our modeling framework provides a natural and flexible procedure to recouple subgroups over multiple horizons.

In general, there are two ways to forecast over multiple horizons, through traditional DLM updating or through customized synthesis. The former, direct approach follows traditional DLM updating and forecasting via simulation as for 1-step ahead, where the synthesis parameters are simulated forward from time t to $t+k$. The latter, customized synthesis involves a trivial modification, in which the model at time $t-1$ for predicting y_t is modified so that the k-step ahead forecast densities made at time $t-k$, i.e., $h_{t-k,j}(x_{tj})$ replace $h_{tj}(x_{tj})$. While the former is theoretically correct, it does not address how effective predictors (and therefore subgroups) can drastically change over time as it relies wholly on the model as fitted, even though one might be mainly interested in forecasting several steps ahead. McAlinn and West (2017) find that, compared to the direct approach, the customized synthesis approach significantly improves multi-step ahead forecasts, since the dynamic model parameters, $\{\boldsymbol{\theta}_t, v_t\}$, are now explicitly geared to the k-step horizon.

B Simulation Study

We consider a simple– yet relevant– simulation study to illustrate and highlight our proposed methodology and its implications for real data applications. This simulation study allows to isolate the gains coming from the combination and re-calibration steps as opposed to the inherent dynamics of the synthesis function, since the data generating process impose stationarity.

To construct a meaningful simulation study, the data generating process must contain certain characteristics that represent conditions often observed empirically. The first characteristic is that all covariates need to be correlated, since most covariates in financial applications are– to a varying degree– correlated. Intuitively, this is a characteristic that is coherent with observation, though not always taken into account or explicitly considered. In terms of dimension reduction techniques, LASSO-type shrinkage methods fail with inconsistent model selection when covariates are highly correlated (Zhao and Yu, 2006). On the other hand, PCA methods perform well when the correlation is high, due to its ability to extract the underlying latent correlation structure, though underperforms when the correlation is mild and change over time.

The second characteristic is that there are omitted variables and the true data generating process is unattainable, i.e., all models are wrong. This is indeed a critical feature, as we cannot realistically expect any model to be fully specified in economic or financial studies. Additionally, the omitted variable might be the key component in understanding the data process. For example, if we are interested in modeling/forecasting the economy, we might consider a latent variable, such as the economic activity, that, while realizes itself through observed variables, e.g., unemployment, is not observed. Thus, a critical component of a modeling technique would necessarily have to account for the biases induced by the omitted variables. These two characteristics build the main components of our simulation study.

We simulate data by the following data generating process:

$$y = -2z_1 + 3z_2 + 5z_3 + \epsilon, \quad \epsilon \sim N(0, 0.01), \tag{B.3a}$$

$$z_1 = \frac{1}{3}z_3 + \nu_1, \quad \nu_1 \sim N\left(0, \frac{2}{3}\right), \tag{B.3b}$$

$$z_2 = \frac{1}{5}z_3 + \nu_2, \quad \nu_2 \sim N\left(0, \frac{4}{5}\right), \tag{B.3c}$$

$$z_3 = \nu_3, \quad \nu_3 \sim N(0, 0.01), \tag{B.3d}$$

where only $\{y, z_1, z_2\}$ are observed and z_3 is omitted. We note that, due to $\{z_1, z_2\}$ being generated from z_3, they are both correlated, though not to an extreme degree to be unrealistic. Since $\{z_1, z_2\}$ are the only two variables observed, we satisfy the aforementioned first characteristic. Secondly, since $\{y, z_1, z_2\}$ are all generated by z_3, and z_3 is not observed, we have a serious omitted variable that drives all the data observed. Because of this, all models that can be constructed will be misspecified (possible models are z_1 or z_2 only, or both $\{z_1, z_2\}$). Additionally, because z_3 drives everything else, there is significant bias in all models generated (i.e. models have high bias and small variance).

We generate $N = 510$ samples, use the first ten to fit the initial model, and forecast 500 data points. We consider eight different strategies that are also considered in the empirical application. A more detailed description of these models will be provided in Section 3 below. The first three models are subset of the possible models with either $\{z_1\}$, $\{z_2\}$, or $\{z_1, z_2\}$ are considered as regressors and the models are estimated using ordinary least squares. We also consider a penalized LASSO-type regression and a PCA regression, where in the first step the latent principle component factors are extracted and used as covariates in a linear regression.

Further, we construct two model combination strategies combining two models generated from linear regressions with only $\{z_1\}$ or $\{z_2\}$, i.e., $p(y|\mathcal{A}_j) = \hat{\beta}z_j + \epsilon_j$ for $j = 1, 2$, where each $\hat{\beta}_j$ is the ordinary least squares estimate. The first model combination scheme is a simple average of the two models, also known as equal weight averaging. It is important to note that, since we only have two covariates, the equal weight averaging is equivalent to the complete subset regression of Elliott et al. (2013). We also consider Bayesian model averaging (BMA), where the weights are determined by the marginal likelihood of the predictive density.

Finally, we compare the seven competing strategies against a simplified, namely time invariant, version of our proposed "decouple-recouple" predictive strategy. Here, the latent states are, as with the two forecast combination schemes, the forecasts from the two linear regressions with $\{z_1\}$ or $\{z_2\}$, but the synthesis function is time invariant instead of the dynamic specification. This yields a simpler setup for DRS by removing the dynamics from the equation and following suit with the model and strategies compared. Here, the synthesis parameters are estimated using a simple Bayesian linear regression with non-informative priors (Jeffreys' prior).

We test the predictive performance by measuring the Root Mean Squared one-step ahead Forecast Error (RMSE) for the first $n = 10, 50, 100, 250, 500$, as well as for the last $l = 400, 300, 200, 100$ data points to emulate a extending window analysis. Table B.1 shows the results from the simulation study, with Panel A being the result of the first n samples and Panel B being the result of the last l samples. Looking at Panel A, we see that, with very small samples, DRS significantly improves over the other methods with an improvement of approximately 60%.

As the sample increases, we see the improvements of DRS shrink, finally settling around 1%. Overall, the gains are small, but is clearly persistent, showing how DRS is able to improve forecasts by learning biases and inter-dependencies and incorporating the information to improve forecasts. Comparatively, we note that LASSO does the worst of the models and strategies considered, while PCA does the best, which is what we expect, since z_1 and z_2 are substantially correlated. Equal weight averaging and BMA also fail and the RMSE does not improve on both models, and in fact its predictive performance is roughly the average of the two models. The full model, interestingly, does worse than the model combination strategies, suggesting that model combination is a legitimate strategy when the covariates are correlated and variables are omitted.

Panel B emulates a setting where a researcher decides to use the first number of samples as a learning period and focuses on sampling the last l in an extending window fashion, a setting familiar in time series analysis. Here, the results are more pronounced, with DRS improving over the other methods by nearly 2% for all l considered. Overall, the simulation study validates the predictive properties of our predictive strategy in a controlled setting; where the study is set up to emulate data often observed in economics and finance, albeit simplified.

C Further Empirical Results

This Section reports further empirical results for both applications. In particular, we report the out-of-sample recursive Log Predictive Density Ratios (LPDR) as calculated by Eq.(9) in the main text. As far as the application on forecasting the equity premium across industries is concerned, we also report a recursive measure of CER which complement the full sample estimates reported in Table 3 and 4 in the main text.

C.1 Forecasting the Aggregate U.S. Inflation

Delving further into the dynamics of the LDPR, Figure C.1 shows the one-step ahead out-of-sample performance of DRS in terms of predictive density. The figure makes clear that the out-of-sample performance of DRS with respect to the benchmarking model combination/shrinkage schemes tend to steadily increase throughout the sample. Interestingly, the LASSO sensibly deteriorates when it comes to predicting the overall one-step ahead distribution of future inflation. Similarly, both the equal weight and BMA show a significant -50% in terms of density forecast accuracy. Consistent with the results in Table (1) in the main text, both Labor Market and Prices on their own outperform the competing combination/shrinkage schemes, except for DRS. Output and Income, Orders and Inventories, and Money and Credit, also perform well, with Output and Income outperforming Labor Market in terms of density forecasts.

On the other hand, we note that Consumption, Interest Rate and Exchange Rates, and the Stock Market, perform the worst compared to the rest by a large margin. LASSO fails poorly in this exercise due to the persistence of the data, and erratic, inconsistent regularization the LASSO estimator imposes. Also, it is fair to notice that the LASSO predictive strategy is the only one that does not explicitly consider time varying volatility of inflation, which is a significant limitation of the methodology, even though stochastic volatility is something that has been shown to substantially affect inflation forecasting (see, e.g., Clark 2011 and Chan 2017, among others). In terms of equal-weight pooling and BMA, we observe that BMA does outperform equal weight, though this is because the BMA weights degenerated quickly to Orders and Inventories, which highlights the problematic nature of BMA, as it acts more as a model selection device rather than a forecasting calibration procedure.

C.2 Retrospective Analysis of Aggregate U.S. Inflation

While the main scope of the paper is on forecasting and basic interpretability from the synthesis weights, using BPS within the DRS framework allows for further analysis into the biases and inter-dependencies of the subgroups: a topic covered here. Note that all analyses in this section is retrospective, that is, the results are given using all of the data in the period examined (i.e. the results are not forward-looking, but looking back from the end of the analysis).

We first analyze the posterior latent correlation between the subgroups by simply taking the posterior MCMC samples and computing the correlation. Here, we report three snapshots within the time period examined; 12:2003 (Figure C.2), a period before the crisis, 12:2008 (Figure C.3), during the great financial crisis, and 12:2014 (Figure C.4), after the crisis. The three periods represent starkly different economic conditions that exemplify how BPS captures the time varying inter-dependency.

Looking at Figure C.2, a period of relative stability, we do not see strong levels of correlation between the subgroups, apart from some mild negative correlation between Labor Market, and Interest Rate and Exchange Rates and Prices, as well as some positive correlations. Moving to Figure C.3, during the crisis, the positive correlations between subgroups lessens overall and a strong negative correlation between Prices and Labor Market as well as Money and Credit appear, alongside positive correlation between Prices and Consumption and Orders and Inventories. The lessening of the positive correlations amongst subgroups suggests a dissipation of predictive information, meaning that dependence among predictability has been mostly lost during the crisis. This is expected during a crisis, as most models tend to deteriorate. On the other hand, the emergence of Prices, and its dependence with some of the other series is notable. This result echoes the dependence patterns seen in the forward looking synthesis weights in Figure 3, where the increase in the information provided from Prices coincides with the decrease in subgroups such as Labor Market. Finally, post-crisis (Figure C.4), we see another pattern emerge, where some positive correlation emerging from the Labor Market, while the strong correlations around Prices are still persistent. In contrast to previous periods, Stock Market looses almost all of its dependence with the other subgroups, which highlights the disjoint of the stock market to the overall economy after the crisis.

We further our retrospective analysis by considering the empirical R^2 between the latent subgroups. The empirical R^2 is defined as the variance of one subgroup explained by all of the other subgroups (Figure C.5) or by another subgroup (paired: Figure C.6). This measure provides an alternative view of dependencies from the correlation in Figure C.2-C.4, as it provides a broader dependence structure of a subgroup and another (group or individual) subgroup, as well as an easier exposition of the dependence over time.

The grouped empirical R^2 (Figure C.5) provides a metric to measure how different a subgroup is from the other groups. Thus, the higher the R^2, the more similar it is to the rest. A general pattern is that post crises (namely the dot.com bubble and the great financial crisis), there is an overall increase in R^2, which is expected as groups tend to "herd" together when uncertainty increases. A particularly interesting result is the connection between Interest Rate and Exchange Rates and Stock Market. In the post-dot.com bubble period, the patterns of R^2 between the two subgroups are disjoint, with the former following the other subgroups closely. This could be seen as a sign of a bubble, as the information provided by the stock market clearly cannot be explained by the other aspects of the economy (compared to the bond/currency market). However, before and after the great financial crisis, we see an almost identical trajectory between the two, indicating a shift in the relation of bonds/currency and stocks before and after the crisis, where the stock market is more "in-line" with the others.

To get a better picture of joint dependencies of subgroups, we use the paired empirical R^2 (Figure C.6). Here, instead of going over all of the combinations, we focus on two key subgroups; Labor Market and Prices. For Labor Market, Prices and Output and Income are the two dominant subgroups of dependency. The dependencies of the two interweave, with switches occurring around the two crises. Prices, on the other hand, is primarily dependent on the Labor Market, with Money and Credit creeping up at the end of the analysis.

C.3 Forecasting the Equity Premium for Different U.S. Industries

Figure C.7 shows the whole out-of-sample path of density forecasting accuracy across modeling specifications. For the ease of exposition, we report the results for Consumer Durable, Consumer Non-Durable, Manufacturing, Telecomm, HiTech, and Other industries. The results for the remaining industries are quantitatively similar and available upon request. Top-left panel shows the out-of-sample path for the Consumer Durable sector. The DRS compares favorably against alternative predictive strategies. Similar results appear in other sectors.

As a whole, Figure C.7 shows clear evidence of how the competing model combination/shrinkage schemes possibly fails to rapidly adapt to structural changes. Although the performance, pre-crisis, is good, it is notable that there is a large loss in predictive performance after the great recession in 2008/2009. DRS consistently shows a performance robust to shifts and shocks and stays in the best group of forecasts throughout the testing sample.

The out-of-sample performance of the LASSO sensibly deteriorates when it comes to predicting the overall one-step ahead distribution of excess returns. The equal-weight linear-pooling turns out to out-perform the competing combination schemes but DRS, as well as the group-specific predictive regressions. Arguably, the strong outperformance of DRS is due to its ability to quickly adjust to different market phases and structural changes in the latent inter-dependencies across groups of predictors. In addition, unlike others, the LASSO-type predictive strategy does not explicitly take into consideration stochastic volatility in the predictive regression, which possibly explains the substantial and persistent underperformance in the aftermath of the great financial crisis, a period of abrupt market fluctuations.

To parallel the results above, we also inspect the economic performance over time by reporting the cumulative sum of the CERs:

$$CCER_{it} = \sum_{\tau=1}^{t} \log \left(1 + CER_{i\tau}\right), \tag{C.4}$$

where CER_{it} is calculated as in Eq.(15) in the main text. Figure C.8 shows the out-of-sample cumulative CER for the Consumer durable, Consumer non-durable, Telecomm, Health, Shops and Other industrial sectors. Except for a few nuances, e.g., the pre-crisis period for Telecomm and Other, the DRS combination scheme constantly outperforms the other predictive strategies.

Interestingly, although initially generating a good CER, the LASSO failed to adjust to the abrupt underlying changes in the predictability of industry returns around the crisis. Despite the initial cumulative CER being slightly in favor of the LASSO vis-a-vis DRS, such good performance disappears around the great financial crisis and in the aftermath of the consequent aggregate financial turmoil. As a result, DRS generates a substantially higher cumulative CER by the end of the forecasting sample, showing much stronger real-time performance.

Results are virtually the same by considering an investor with short-sales constraints. Figure C.9 shows the out-of-sample cumulative CER for the Consumer durable, Consumer non-durable, Telecomm, Health, Shops and Other industrial sectors, but now imposing that the vector of portfolio weights should be positive and sum to one, that is, no-short sale constraints are imposed.

The picture that emerges is similar to the above. Except for a transitory period during the great financial crisis for the Health sector, the DRS strategy significantly outperforms all competing specifications. As before, by imposing no-short constraints the gap between DRS the competing specifications is substantially reduced, thought the gains are persistent.

Table B.1. Simulation Study: Out-of-Sample Forecasting Performance

This table reports the out-of-sample comparison of our decouple-recouple framework against each individual model, full model, LASSO, PCA, equal weight average of models, and BMA, for simulated data. Performance comparison is based on the Root Mean Squared Error (RMSE).

Panel A: Forecasting 1-Step Ahead Simulation Data (based on first n samples)

n	z_1	z_2	$\{z_1, z_2\}$	LASSO	PCA	EW	BMA	DRS
10	2.8768	2.8820	2.8830	2.7988	2.8613	2.8793	2.8793	1.7923
	-60.51%	-60.80%	-60.85%	-56.16%	-59.64%	-60.65%	-60.65%	-
50	2.8538	2.8618	2.8578	2.8557	2.8464	2.8577	2.8575	2.7568
	-3.52%	-3.81%	-3.66%	-3.58%	-3.25%	-3.66%	-3.65%	-
100	2.9091	2.9121	2.9114	2.8993	2.9020	2.9106	2.9105	2.8977
	-0.39%	-0.50%	-0.47%	-0.06%	-0.15%	-0.44%	-0.44%	-
250	2.8564	2.8583	2.8577	2.8606	2.8532	2.8573	2.8573	2.8475
	-0.31%	-0.38%	-0.36%	-0.46%	-0.20%	-0.35%	-0.34%	-
500	2.7506	2.7520	2.7516	2.7526	2.7494	2.7513	2.7513	2.7197
	-1.14%	-1.19%	-1.17%	-1.21%	-1.09%	-1.16%	-1.16%	-

Panel B: Forecasting 1-Step Ahead Simulation Data (based on last l samples)

l	z_1	z_2	$\{z_1, z_2\}$	LASSO	PCA	EW	BMA	DRS
400	2.6926	2.6934	2.6931	2.6973	2.6931	2.6930	2.6930	2.6573
	-1.33%	-1.36%	-1.35%	-1.51%	-1.35%	-1.34%	-1.34%	-
300	2.6269	2.6278	2.6272	2.6237	2.6281	2.6274	2.6273	2.5852
	-1.62%	-1.65%	-1.63%	-1.49%	-1.66%	-1.63%	-1.63%	-
200	2.6772	2.6779	2.6777	2.6797	2.6777	2.6776	2.6776	2.6183
	-2.25%	-2.27%	-2.27%	-2.34%	-2.27%	-2.26%	-2.26%	-
100	2.6186	2.6191	2.6188	2.6214	2.6182	2.6189	2.6189	2.5717
	-1.83%	-1.85%	-1.83%	-1.93%	-1.81%	-1.84%	-1.84%	-

Figure C.1. Out-of-sample LPDR for Forecasting U.S. Inflation

This figure shows the dynamics of the out-of-sample Log Predictive Density Ratio (LPDR) as in Eq.(9) obtained for each of the group-specific predictors, by taking the results from a set of competing model combination/shrinkage schemes, e.g., Equal Weight, and Bayesian Model Averaging (BMA). LASSO not included due to scaling. The sample period is 01:2001-12:2015, monthly. The objective function is the one-step ahead density forecast of annual inflation.

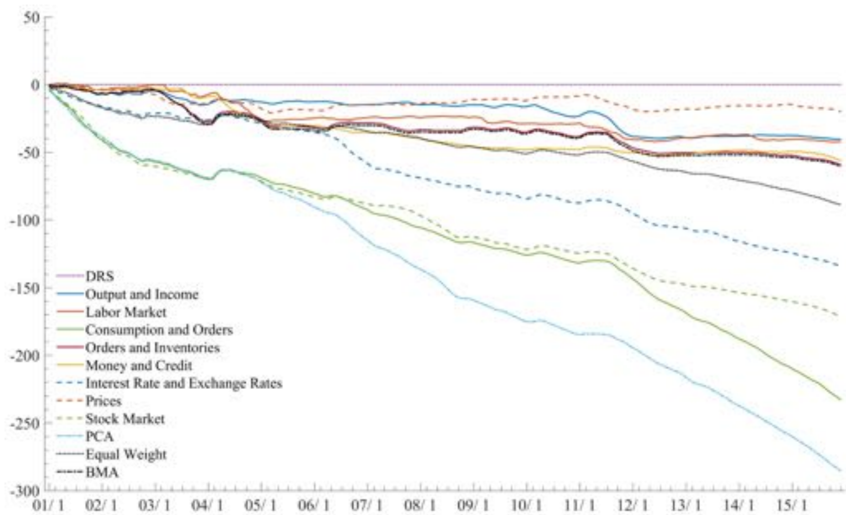


Figure C.2. US inflation rate forecasting: Retrospective posterior correlations of latent agent factors at 12:2003.

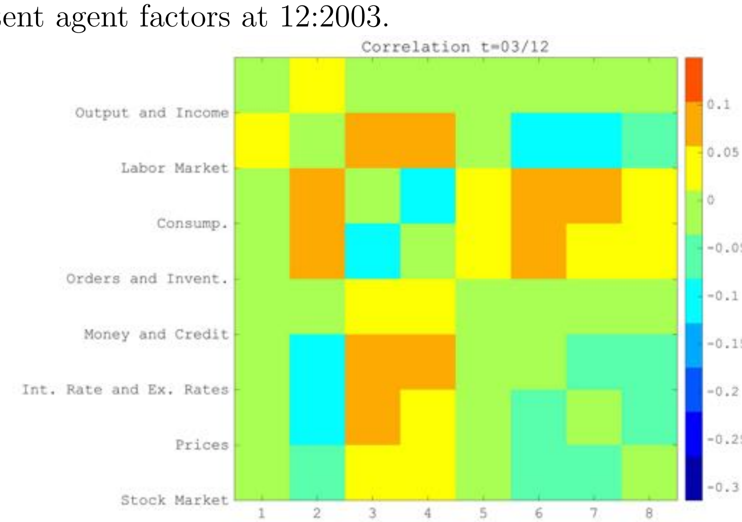


Figure C.3. US inflation rate forecasting: Retrospective posterior correlations of latent agent factors at 12:2008.

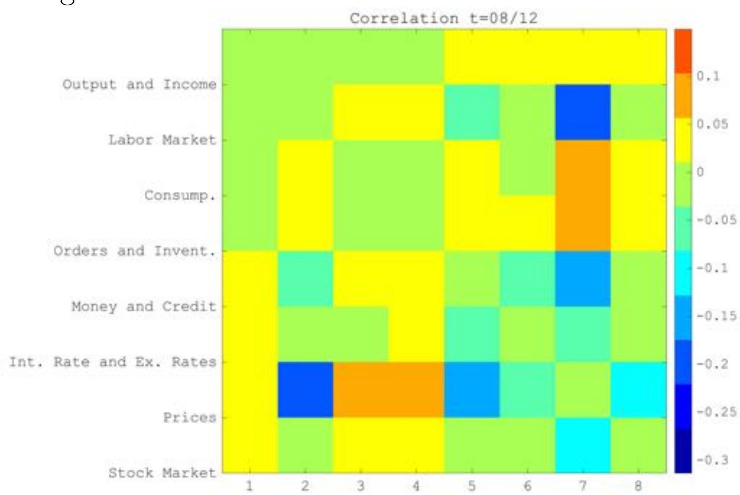


Figure C.4. US inflation rate forecasting: Retrospective posterior correlations of latent agent factors at 12:2014.

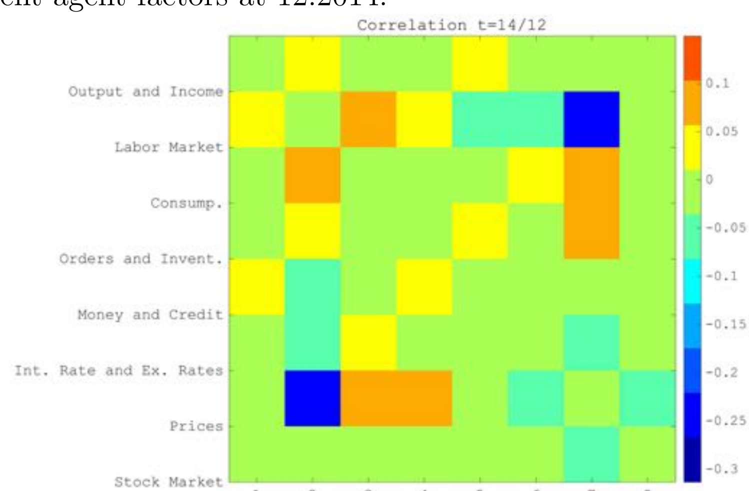


Figure C.5. US inflation rate forecasting: Retrospective latent dependencies

This figure shows the retrospective latent inter-dependencies across groups of predictive densities used in the recoupling step. The latent dependencies are measured using the MC-empirical R^2, i.e., variation explained of one model given the other models. These latent components are sequentially computed at each of the $t = 1{:}180$ months.

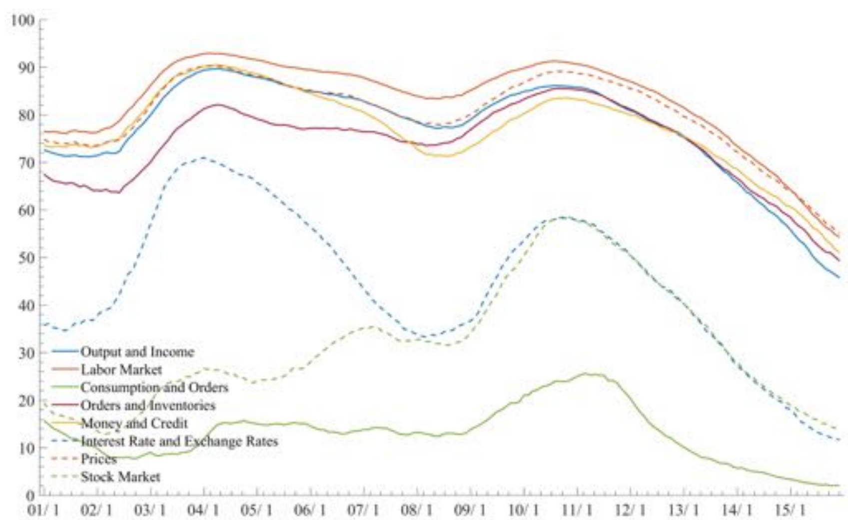
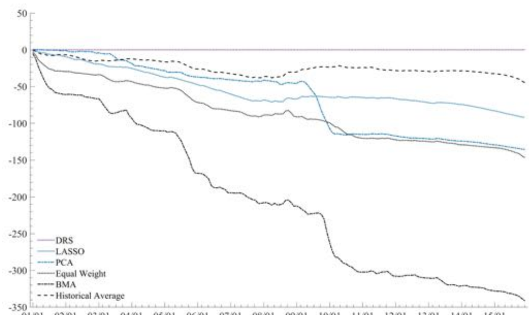


Figure C.6. US inflation rate forecasting: Retrospective latent dependencies (paired)

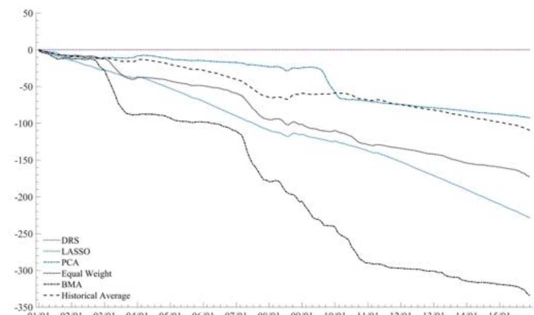
This figure shows the retrospective paired latent inter-dependencies across groups of predictive densities used in the recoupling step. The latent dependencies are measured using the paired MC-empirical R^2, i.e., variation explained of one model given another model, for Labor Market (top) and Prices (bottom). These latent components are sequentially computed at each of the $t = 1{:}180$ months.

Figure C.7. Out-of-sample LPDR for Forecasting the Equity Premium for Different Industries in the U.S.

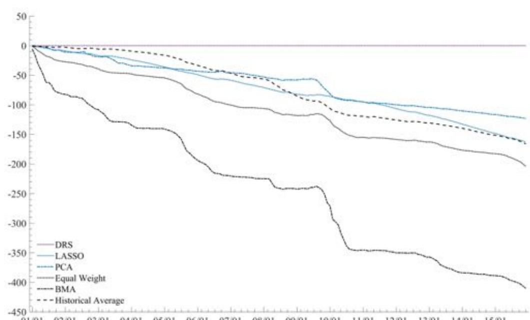
This figure shows the dynamics of the out-of-sample Log Predictive Density Ratio (LPDR) as in Eq.(7) obtained for each of the group-specific predictors, by taking the historical average of the stock returns (HA), and the results from a set of competing model combination/shrinkage schemes, e.g., LASSO, Equal Weight, and Bayesian Model Averaging (BMA). For the ease of exposition we report the results for four representative industries, namely, Consumer Durables, Consumer Non-Durables, Telecomm, Health, Shops, and Other. Industry aggregation is based on the four-digit SIC codes of the existing firm at each time t following the industry classification from Kenneth French's website.



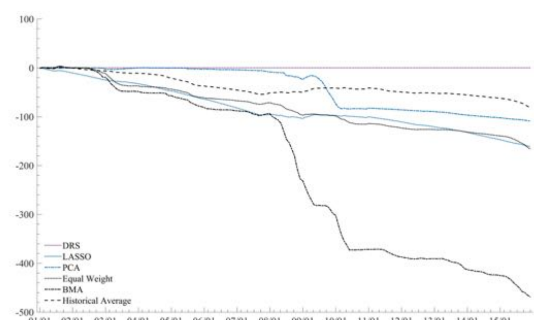
(a) Consumer Durable



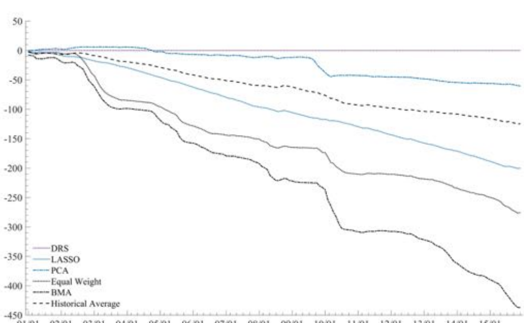
(b) Cons. Non-Durable



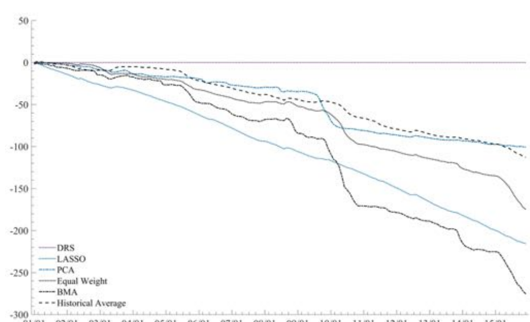
(c) Telecomm



(d) Other



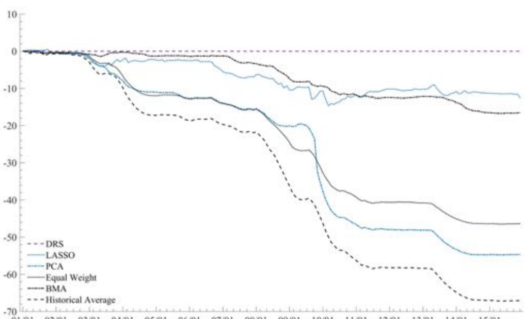
(e) Health



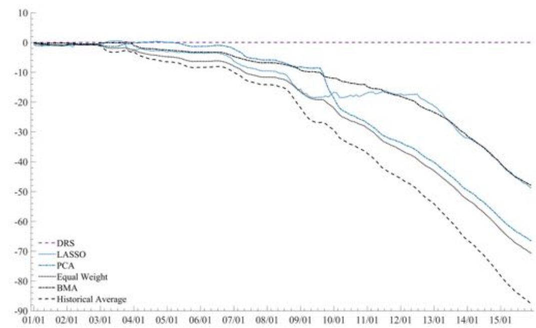
(f) Shops

Figure C.8. Out-of-Sample Cumulative CER without Constraints

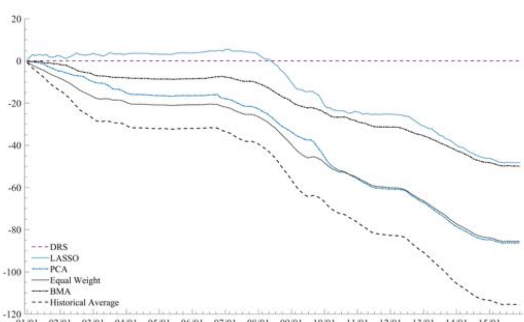
This figure shows the dynamics of the out-of-sample Cumulative Certainty Equivalent Return (CER) for an unconstrained as in Eq. (C.4) obtained for each of the group-specific predictors, by taking the historical average of the stock returns (HA), and the results from a set of competing model combination/shrinkage schemes, e.g., LASSO, Equal Weight, and Bayesian Model Averaging (BMA). For the ease of exposition we report the results for four representative industries, namely, Consumer Durables, Consumer Non-Durables, Telecomm, Health, Shops, and Other. Industry aggregation is based on the four-digit SIC codes of the existing firm at each time t following the industry classification from Kenneth French's website.



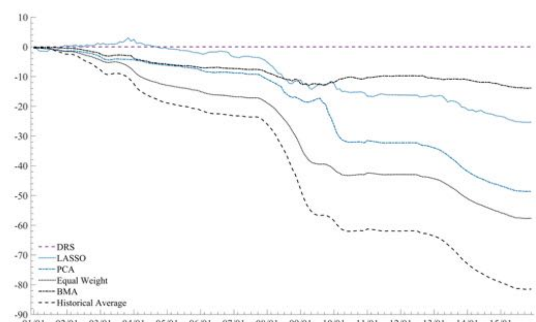
(a) Consumer Durable



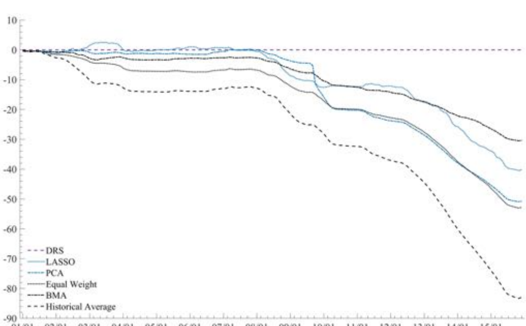
(b) Cons. Non-Durable



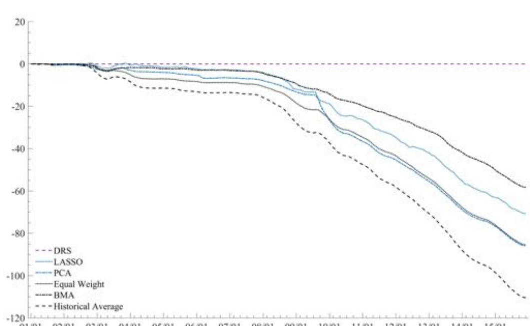
(c) Telecomm



(d) Other



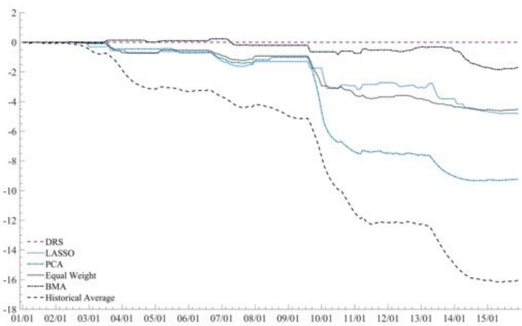
(e) Health



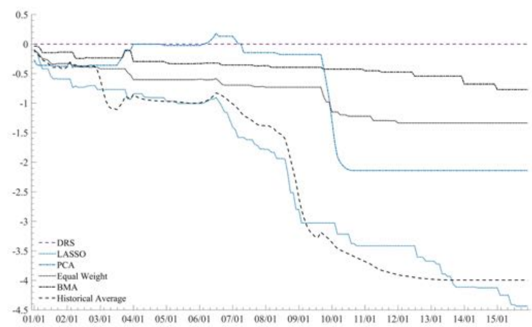
(f) Shops

Figure C.9. Out-of-sample Cumulative CER with Short-Sale Constraints

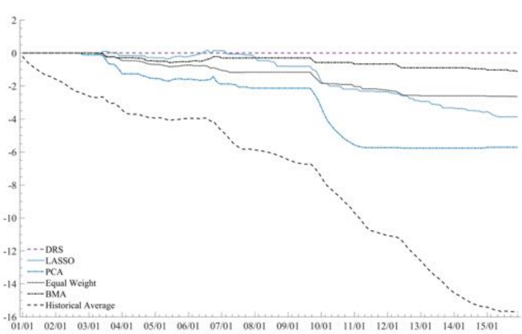
This figure shows the dynamics of the out-of-sample Cumulative Certainty Equivalent Return (CER) for a short-sale constrained investor as in Eq. (C.4) obtained for each of the group-specific predictors, by taking the historical average of the stock returns (HA), and the results from a set of competing model combination/shrinkage schemes, e.g., LASSO, Equal Weight, and Bayesian Model Averaging (BMA). For the ease of exposition we report the results for four representative industries, namely, Consumer Durables, Consumer Non-Durables, Telecomm, Health, Shops, and Other. Industry aggregation is based on the four-digit SIC codes of the existing firm at each time t following the industry classification from Kenneth French's website.



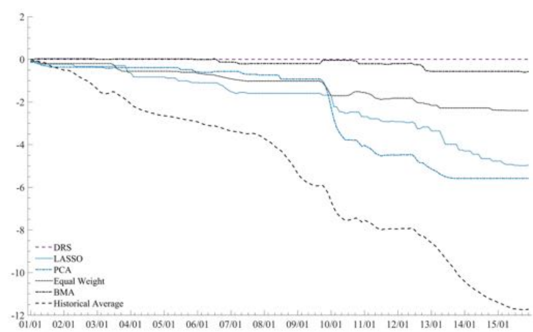
(a) Consumer Durable



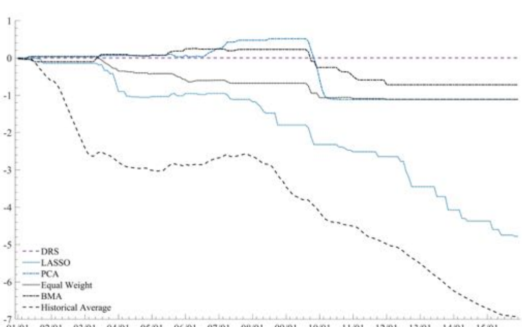
(b) Cons. Non-Durable



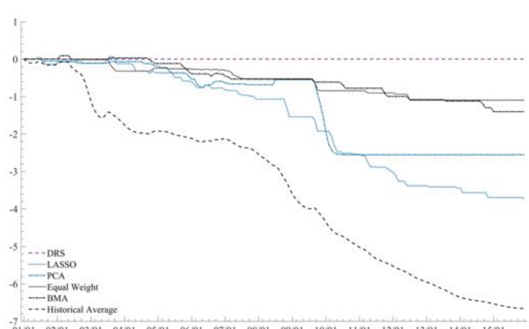
(c) Telecomm



(d) Other



(e) Health



(f) Shops

References

Chan, J. C. 2017. The stochastic volatility in mean model with time-varying parameters: An application to inflation modeling. *Journal of Business & Economic Statistics* 35:17–28.

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

Elliott, G., A. Gargano, and A. Timmermann. 2013. Complete subset regressions. *Journal of Econometrics* 177:357–373.

Frühwirth-Schnatter, S. 1994. Data augmentation and dynamic linear models. *Journal of Time Series Analysis* 15:183–202.

McAlinn, K., and M. West. 2017. Dynamic Bayesian predictive synthesis in time series forecasting. *Journal of Econometrics* Forthcoming.

Prado, R., and M. West. 2010. *Time Series: Modelling, Computation & Inference*. Chapman & Hall/CRC Press.

West, M., and P. J. Harrison. 1997. *Bayesian Forecasting & Dynamic Models*. 2nd ed. Springer Verlag.

Zhao, P., and B. Yu. 2006. On model selection consistency of Lasso. *Journal of Machine Learning Research* 7:2541–2563.