

Large-Scale Dynamic Predictive Regressions*

Daniele Bianchi[†]

Kenichiro McAlinn[‡]

First draft: December 2017.

This draft: June 9, 2018

Abstract

We propose a “decouple-recouple” dynamic predictive strategy and contribute to the literature on 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. As a result, the latent inter-dependencies across predictive densities are sequentially learned and the aggregate bias corrected. We test our procedure by predicting both the equity premium across different industries and the inflation rate in the U.S, based on a large set of financial ratios and macroeconomic variables. The main empirical results show that our framework generates both statistically and economically significant out-of-sample outperformance compared to a variety of sparse and dense modelling benchmarks, while maintaining interpretability on the relative importance of each class of predictors.

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

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

*We thank Roberto Casarin, Andrew Patton, Davide Pettenuzzo, and participants at the NBER-NSF Seminar on Bayesian Inference in Econometrics and Statistics at Stanford and Università Cá Foscari for their helpful comments and suggestions.

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

[‡]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 in a data-rich environment; 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 behaviour 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.

Confronted with a large set of predictors, two main classes of models became popular. *Sparse* modelling 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. In the Bayesian literature, a prominent example is given by George and McCulloch (1993) (and more recently, Ročková and George 2016 and Ročková 2018), which introduced variable selection through a data-augmentation approach. Similarly, regularised models take a large number of predictors and introduces penalisation to discipline the model space. LASSO-type regularisation and ridge regressions are by far the most used in both research and practice. A second class of models fall under the heading of *dense* modelling; this is based on the assumption that, a priori, all variables could bring useful information for 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 modelling (see, e.g., Stock and Watson 2002 and De Mol et al. 2008 and the references therein), which is highly popular in applied macroeconomics.

Both these approaches entail either an implicit or explicit reduction of the model space that is in-

¹See, e.g., Timmermann (2004), De Mol, Giannone, and Reichlin (2008), Mönch (2008), Bai and Ng (2010), Belloni, Chen, Chernozhukov, and Hansen (2012), Billio, Casarin, Ravazzolo, and van Dijk (2013), Elliott, Gargano, and Timmermann (2013), Manzan (2015), Harvey, Liu, and Zhu (2016), Freyberger, Neuhierl, and Weber (2017), Giannone, Lenza, and Primiceri (2017), and McAlinn and West (2017), just to name a few.

tended to mitigate the curse of dimensionality. However, the question of which one of these techniques is best is still largely unresolved. 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. For instance, a portfolio manager interested in constructing a long-short investment strategy might not find useful to use latent factors that she cannot clearly identify as meaningful sources of risk, or similarly would not want critical, economically sound, predictors to be shrunk to zero. More importantly, Giannone et al. (2017) recently show, in a Bayesian setting, that the posterior distribution of the parameters of a large dimensional linear regression do not concentrate on a single sparse model, but instead spreads over different types of models depending on priors elicitation. These problems possibly undermine the usefulness of exploiting data-rich environments for economic and financial decision making.

In this paper, we propose a class of data-rich predictive synthesis techniques and contribute to the literature on predictive modelling and decision making with big data. Unlike sparse modelling, we do not assume a priori that there is sparsity in the set of predictors. For example, suppose we are interested in forecasting the one-step ahead excess returns on the stock market based on, say, a hundred viable predictors. Using standard LASSO-type shrinkage— a typical solution— will implicitly impose a dogmatic prior that only a small subset of those regressors is useful for predicting stock excess returns and the rest is noise, i.e., sparsity is pre-assumed. Yet, there is no guarantee that the small subset is consistent, or smooth, over time. Similarly, even with such a moderate size, the model space is about $1e+30$ possible combinations of the predictors, which makes difficult to claim any reasonable convergence within the class of standard stochastic search variable selection algorithms (see, e.g., Giannone et al. 2017).

We, in turn, retain all of the information available and *decouple* a large predictive model into a set of much smaller predictive regressions, which are constructed by similarity among the set of regressors. Precisely, suppose these predictors can be classified into J different subgroups, each one containing fewer regressors, according to their economic meaning. Rather than assuming a sparse structure, we retain all of the information by estimating J different predictive densities— separately and sequentially— one for each class of predictors, and *recouple* them dynamically using the predictive synthesis approach.

One comment is in order. The term and general concept of "decouple/recouple" stems from the

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.

The way the subgroups of regressors are classified in the first step is independent of the decoupling-recoupling strategy. In the empirical application we classified groups of variables according to their economic meaning. However, one can use correlation-based clustering algorithms such as, K-means, fuzzy C-means, hierarchical clustering, mixture of Gaussians, or other nearest neighbour classifications to construct the set of smaller dimensional regressions.

Our proposed approach significantly differs from model combination of multiple small models (e.g. multiple LASSO models with different tuning parameters), such as Stevanovic (2017), by utilising the theoretical foundations and recent developments of Bayesian predictive synthesis (BPS: West and Crosse, 1992; West, 1992; McAlinn and West, 2017). This makes our decouple-recouple strategy theoretically and conceptually coherent, as it regards the decoupled models as separate latent states that are learned and calibrated using the Bayes theorem in an otherwise typical dynamic linear modelling framework (see West and Harrison 1997). Under this framework, the dependencies between subgroups, as well as biases within each subgroup, can be sequentially learned; information that is critical, though lost in typical model combination techniques.

The intuition why our predictive strategy could improve the forecasting performance compared to shrinkage methods and factor models is fairly simple. To fix ideas, we reconsider the bias-variance tradeoff; a well known statistical property where an increase in model complexity increases variance and lowers bias and vice versa. The goal in both shrinkage methods and factor models is to arbitrarily lower model complexity to balance bias and variance, in order to potentially minimise predictive loss. In terms of LASSO-type shrinkage, increasing the tuning parameter (i.e. increasing shrinkage) leads to increased bias, so using cross-validation aims to balance the bias-variance tradeoff by balancing the tuning parameter. Similarly, in factor model the optimal number of latent factors is chosen to reduce the variance by reducing the model dimensionality at the cost of increasing the bias. Our proposed

method takes a significantly different approach towards the bias-variance tradeoff by breaking a large dimensional problem into a set of small dimensional ones, while at the same time exploiting the fact that our methodology can learn the biases and inter-dependencies via Bayesian learning. As this is the case, recoupling step benefits from biased models, as long as the bias has a signal that can be learned. More specifically, by decoupling the model into smaller, less complex models, we adjust for the bias— that characterise each group— that is sequentially learned and corrected, while maintaining the low variance from each model. This flips the bias-variance tradeoff around, exploiting the weakness of low complexity models to an advantage in the recoupling step, potentially improving predictive performance.

We calibrate and implement the proposed methodology, which we call decouple-recouple synthesis (DRS), 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 empirical exercise involves a balanced panel of 119 monthly macroeconomic and financial variables, which are classified into eight main groups: Output and Income, Labor Market, Consumption, Orders and Inventories, Money and Credit, Interest Rate and Exchange Rates, Prices, and Stock Market.

The second application relates to forecasting monthly year-on-year total excess returns across different industries from 1970/1 to 2015/12, based on a large set of predictors, which have been chosen 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, Strauss, and Zhou 2010, and Dangl and Halling 2012, among others). More specifically, we collect monthly data on more than 70 pre-calculated financial ratios for all U.S. companies across eight different categories. 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 aggregation of the single-firm predictors. Industry aggregation is based on the four-digit SIC codes of the existing firm at each time t . Those 70 ratios are classified into eight main categories: Valuation, Profitability, Capitalisation,

Financial Soundness, Solvency, Liquidity, Efficiency Ratios, and Other. Together with industry-specific predictors, we use additional 14 aggregate covariates obtained from existing research, which are divided in two categories; aggregate financials and macroeconomic variables (see, Goyal and Welch 2008 and Rapach et al. 2010).

To evaluate our approach empirically, we compare forecasts against standard Bayesian model averaging (BMA), in which the forecast densities are mixed with respect to sequentially updated model probabilities (e.g. Harrison and Stevens, 1976; West and Harrison, 1997, Sect 12.2), 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 (e.g. West 1984). In addition, we compare the forecasts from our setting with a state-of-the-art LASSO-type regularisation, which constraints the coefficients of least relevant variables to be null leading to sparse models *ex-post*, and PCA based latent factor modelling (Stock and Watson, 2002; McCracken and Ng, 2016). While some of these strategies might seem overly simplistic, they have been shown to dominate some more complex aggregation strategies in some contexts, at least in terms of direct point forecasts in empirical studies (Genre, Kenny, Meyler, and Timmermann, 2013). Finally, we also compare our decouple-recouple model synthesis scheme against the marginal predictive densities computed from the group-specific set of predictors taken separately. Forecasting accuracy is primarily assessed by evaluating the out-of-sample log predictive density ratios (LPDR); at horizon k and across time indices t . Although we mainly focus on density forecasts in this paper, we also report the root mean squared forecast error (RMSFE) over the forecast horizons of interest, which, combined with the LPDR results, paints a fuller picture of the results.

Irrespective of the performance evaluation metrics, our decouple-recouple model synthesis scheme emerges as the best for forecasting the annual inflation rate for the U.S. economy. This holds for both one-step ahead and three-step ahead forecasts. It significantly out-performs both sequential BMA and the equal-weighted linear pooling of predictive densities. Interestingly, the LASSO performs worst among the model combination/shrinkage schemes, in terms of density forecasts. The sequential estimates of the latent inter-dependencies across classes of macroeconomic predictors show that pressure on the labor market and price levels tend to dominate other groups of predictors, with labor market being a dominant component in early 2000s, while prices tend to increase their weight in the aggregate predictive density towards the end of the test period.

The results are possibly even more pronounced concerning the prediction of the yearly total excess returns across different industries. The differences in the LPDRs are rather stark and clearly shows a performance gap in favour of DRS. None of the alternative specifications come close to DRS when it comes to predicting one-step ahead. While the equally-weighted linear pooling turns out to be a challenging benchmark to beat, we show that LASSO-type shrinkage estimators and PCA perform poorly out-of-sample, especially when it comes to predicting the one-step ahead density of excess returns. This result is consistent with the recent evidence in Diebold and Shin (2017), which show the sub-optimality of LASSO estimators in out-of-sample real-time forecasting exercises. We also compare our model combination scheme against the competitors outlined above on the basis of the economic performance assuming a representative investor with power utility preferences.

The comparison is conducted for the unconstrained as well as short-sales constrained investor at the monthly horizons, for the entire sample. We find that the economic constraints lead to higher Certainty Equivalent (CER) values at all horizons and across practically all competing specifications. Specifically, the short-sale constraint results in a higher CER (relative to the unconstrained case) of more than 100 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 under portfolio constraints. In addition, we show that DRS allows to reach a higher CER both in the cross-section and in the time-series, which suggests that there are economically important gains by using our methodology.

The structure of this paper is as follows. Section 2 introduces our decouple-recouple methodology for the efficient synthesis of predictive densities. Section 3 presents the core of the paper and report the empirical results related to both the U.S. annual inflation forecasts and the total stock returns predictability across industries in the U.S. Section 4 concludes the paper with further discussion.

2 Decouple-Recouple 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 (such as interest rates). 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 time series linear predictive regression (see, e.g., Stambaugh 1999, Pesaran and Timmermann 2002, Avramov 2004, Lewellen 2004, Goyal and Welch 2008, and Rapach et al. 2010, among others);

$$y_t = \boldsymbol{\beta}' \mathbf{x}_t + \epsilon_t, \quad \epsilon_t \sim N(0, \nu), \quad (1)$$

where y_t is the quantity of interest, \mathbf{x}_t is a p -dimensional vector of predictors, which could have its own dynamics, $\boldsymbol{\beta}$ is the p -dimensional vector of betas, and ϵ_t is some observation noise (Gaussian and constant over time here to fix ideas).

In many practically important applications, the dimension of predictors relevant to \mathcal{D} 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 the decision making process of \mathcal{D} . Under this setting, regularisation 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 Elias (2005), while using all of the predictors available, reduces them to a small preset number of latent factors that are typically difficult to interpret or control, in the sense of optimal decision making.

Our decouple-recouple strategy 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

$$y_t = \boldsymbol{\beta}'_1 \mathbf{x}_{t,1} + \dots + \boldsymbol{\beta}'_j \mathbf{x}_{t,j} + \dots + \boldsymbol{\beta}'_J \mathbf{x}_{t,J} + \epsilon_t, \quad \epsilon_t \sim N(0, \nu). \quad (2)$$

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 sensible estimates. The first step of our model combination strategy is to

decouple Eq. (2) into J smaller models, such as,

$$y_t = \beta_j' \mathbf{x}_{t,j} + \epsilon_{t,j}, \quad \epsilon_{t,j} \sim N(0, \nu_j), \quad (3)$$

for all $j = 1:J$, producing forecast distributions $p(y_{t+k}|\mathcal{A}_j)$, where \mathcal{A}_j denotes each subgroup, and $k \geq 1$ is the forecast horizon. Since Eq. (3) is a linear projection of data from each subgroup, we can consider, without loss of generality, that $p(y_{t+k}|\mathcal{A}_j)$ is reflecting the information arising from that subgroup regarding 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 $p(y_{t+k})$ reflecting and incorporating all of the information that arises from each subgroup. 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 for a further discussion);

$$y_{t+k} = w_1 p(y_{t+k}|\mathcal{A}_1) + \dots + w_j p(y_{t+k}|\mathcal{A}_j) + \dots + w_J p(y_{t+k}|\mathcal{A}_J), \quad (4)$$

where weights $w_{1:J}$ are estimated by the decision maker based on past observations (e.g. using $w_{1:J}$ proportional to the marginal likelihood). The main difference between BMA and linear pooling is about the domain of $w_{1:J}$ and the estimation approach adopted.

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. Arguably each subgroup $p(y_{t+k}|\mathcal{A}_j)$ is always biased unless one of them is the data generating process, which is something we cannot expect in applications in economics or finance. 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. As a matter of fact, optimal combination of weights should be chosen to minimise the expected loss of the *combined* forecast, which, by definition, reflects both the forecasting accuracy of each sub-model and the correlation across single 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. The linkages between liquidity, solvency, and aggregate macroeconomic

variables changed before and after the great financial crisis of 2008/2009. Thus, an effective recoupling step must be able to sequentially learn and recover the latent biases and inter-dependencies between the subgroups/sub-models.

To address these issues, we build on the theoretical foundations and recent developments proposed in West and Crosse (1992); West (1992); McAlinn and West (2017). Each subgroup is considered to be a latent state, whereby $p(y_{t+k}|\mathcal{A}_j)$ represents a distinct prior on state $j = 1, \dots, J$. As BPS treats the latent states within the Bayesian paradigm, the biases and inter-dependencies between the latent states can be learned and recovered via standard Bayesian updating. The difference between BPS and more general latent factor models, such as PCA, is that BPS allows to pin down each latent state, using priors $p(y_{t+k}|\mathcal{A}_j)$ at each time t , to a group that \mathcal{D} specifies. In this respect, the underlying assumption of BPS is that each latent state reflects information from each subgroup/sub-model, and thus retains interpretability, which is the key component of \mathcal{D} 's decision making process.

2.1 Bayesian Predictive Synthesis

In the general framework of BPS, the decision maker \mathcal{D} is interested in predicting some quantity y and aims to incorporate information from J individual models labeled \mathcal{A}_j , ($j = 1:J$). \mathcal{D} has some prior information $p(y)$ about the quantity of interest, and each \mathcal{A}_j provides their own prior distribution about what they believe the outcome of the quantity is in the form of a predictive distribution $h_j(x_j) = p(y|\mathcal{A}_j)$; the collection of which defines the information set $\mathcal{H} = \{h_1(\cdot), \dots, h_J(\cdot)\}$. The question BPS tackles is this: how should a Bayesian decision maker consolidate these prior distributions (\mathcal{D} 's own and of $\mathcal{A}_{1:J}$) and learn, update, and calibrate in order to improve forecasts?

A formal prior-posterior updating scheme posits that, for a given prior $p(y)$, and (prior) information set \mathcal{H} provided by $\mathcal{A}_{1:J}$, we can update using the Bayes theorem to obtain a posterior $p(y|\mathcal{H})$. Due to the complexity of \mathcal{H} —a set of J density functions with cross-sectional time-varying dependencies as well as individual biases— $p(y, \mathcal{H}) = p(y)p(\mathcal{H}|y)$ is impractical since $p(\mathcal{H}|y)$ is difficult to define. The works of West and Crosse (1992) and West (1992) extend the basic theorem of Genest and Schervish (1985), in the context of incorporating multiple prior information provided by experts, to show that, under a specific consistency condition, \mathcal{D} 's posterior density takes the form

$$p(y|\mathcal{H}) = \int \alpha(y|\mathbf{x})h(\mathbf{x})d\mathbf{x} \quad \text{where} \quad h(\mathbf{x}) = \prod_{j=1}^J h_j(x_j). \quad (5)$$

Here, $\mathbf{x} = x_{1:J} = (x_1, \dots, x_J)'$ is a J -dimensional latent vector of states with priors provided by each \mathcal{A}_j , and $\alpha(y|\mathbf{x})$ is a conditional density function, which reflects how the decision maker believes these latent states to be synthesised. The only requirement of Eq. (5), so that it is a coherent Bayesian posterior, is that it must be consistent with \mathcal{D} 's prior, i.e.,

$$p(y) = \int \alpha(y|\mathbf{x})m(\mathbf{x})d\mathbf{x} \quad \text{where} \quad m(\mathbf{x}) = E[h(\mathbf{x})], \quad (6)$$

the expectation in the last formula being over \mathcal{D} 's belief of what $p(\mathcal{H})$ should be. Critically, the representation of Eq. (5) does not require a full specification of $p(y, \mathcal{H})$, but only the conditional density $\alpha(y|\mathbf{x})$ and the marginal expectation function $m(\mathbf{x})$. These two functions alone allows to incorporate any prior knowledge in the form of models' predictions in terms of biases, predictive accuracy, and more importantly, inter-dependencies among each other. It is important to note that the theory does not specify the form of $\alpha(y|\mathbf{x})$. In fact, McAlinn and West (2017) show that many forecast combination methods, from linear combinations (including Bayesian model averaging) 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 BPS.

Now, suppose \mathcal{D} is interested in the more critical and relevant task of one-step ahead forecasting. \mathcal{D} wants to predict y_t and receives current forecast densities $\mathcal{H}_t = \{h_{t1}(x_{t1}), \dots, h_{tJ}(x_{tJ})\}$ from the set of models. The full information set used by \mathcal{D} is thus $\{y_{1:t-1}, \mathcal{H}_{1:t}\}$, the past data of y and historical information of predictive distributions coming from $\mathcal{A}_{1:J}$. Extending Eq. (5) to a dynamic context (as done in McAlinn and West, 2017), \mathcal{D} has a dynamic posterior distribution of the forecast of y_t at time $t - 1$ of the form

$$p(y_t|\Phi_t, \mathbf{y}_{1:t-1}, \mathcal{H}_{1:t}) \equiv p(y_t|\Phi_t, \mathcal{H}_t) = \int \alpha_t(y_t|\mathbf{x}_t, \Phi_t) \prod_{j=1:J} h_{tj}(x_{tj})dx_{tj} \quad (7)$$

where $\mathbf{x}_t = x_{t,1:J}$ is a J -dimensional latent agent state vector at time t , $\alpha_t(y_t|\mathbf{x}_t, \Phi_t)$ is \mathcal{D} 's conditional synthesis function for y_t given the latent states \mathbf{x}_t , and Φ_t represents some time-varying parameters learned and calibrated over $1:t$.

This general framework implies that \mathbf{x}_t is the realisation of the inherent dynamic latent factors at time t and a synthesis is achieved by recoupling these separate latent predictive densities to the time

series y_t through the time-varying conditional distribution $\alpha_t(y_t|\mathbf{x}_t, \Phi_t)$. Though the theory does not specify $\alpha_t(y_t|\mathbf{x}_t, \Phi_t)$, a natural choice— as with McAlinn and West (2017)— is to impose linear dynamics, such that,

$$\alpha_t(y_t|\mathbf{x}_t, \Phi_t) = N(y_t|\mathbf{F}'_t\boldsymbol{\theta}_t, v_t) \quad \text{with} \quad \mathbf{F}_t = (1, \mathbf{x}'_t)' \quad \text{and} \quad \boldsymbol{\theta}_t = (\theta_{t0}, \theta_{t1}, \dots, \theta_{tJ})', \quad (8)$$

where $\boldsymbol{\theta}_t$ represents a $(J + 1)$ –vector of time-varying synthesis coefficients. Observation noise is reflected in the innovation variance term v_t , and the general time-varying parameters Φ_t is defined as $\Phi_t = (\boldsymbol{\theta}_t, v_t)$. The evolution of these parameters is needed 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 traditional in the Bayesian time series literature (see West and Harrison 1997; Prado and West 2010). Thus, we consider

$$y_t = \mathbf{F}'_t\boldsymbol{\theta}_t + \nu_t, \quad \nu_t \sim N(0, v_t), \quad (9a)$$

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(0, v_t\mathbf{W}_t), \quad (9b)$$

where $v_t\mathbf{W}_t$ 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 residuals ν_t and the evolution *innovations* $\boldsymbol{\omega}_s$ are independent over time and mutually independent for all t, s . The dynamics of \mathbf{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 $\gamma_t \sim \text{Beta}(\delta n_{t-1}/2, (1 - \delta)n_{t-1}/2)$ 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 parameter.

With the \mathbf{x}_t vectors in each \mathbf{F}_t treated as latent variables, a dynamic latent factor model is defined through Eqs. (9). When forecasting each t , the latent states are conceived as arising as single draws from the set of models' predictive densities $h_{tj}(\cdot)$, the latter becoming available at time $t - 1$ for forecasting y_t . Note that x_{tj} are drawn *independently* (for t) from

$$p(\mathbf{x}_t|\Phi_t, \mathbf{y}_{1:t-1}, \mathcal{H}_{1:t}) \equiv p(\mathbf{x}_t|\mathcal{H}_t) = \prod_{j=1:J} h_{tj}(x_{tj}) \quad (10)$$

with $\mathbf{x}_t, \mathbf{x}_s$ conditionally independent for all $t \neq s$. Importantly, the independence of the x_{tj} , condi-

tional on h_{tj} , must not be confused with the question of modelling and estimation of the dependencies among predictive densities. \mathcal{D} 's modelling and estimation of the biases and inter-dependencies among these models are effectively mapped on and reflected through the parameters $\Phi_t = (\theta_t, v_t)$.

Further discussion on the dynamic synthesis function is in order. While we choose a simple and flexible dynamic form for the synthesis function, $\alpha_t(y_t|\mathbf{x}_t, \Phi_t)$, in theory we do not need to imply any certain structure for the synthesis of model-specific predictive densities. For instance, one can set cross-sectional correlations to be high if different models are known to give identical predictions; similarly, if we believe there are clear regime changes that favour certain models at given periods of time, a regime switching approach or an indicator in the state equation might be suitable. We also note that most methods in the forecast combination literature focus on weights that are restricted to the unit simplex, as well as the weights summing to one. For weights summing to one, we can apply the technique used in Irie and West (2016), where the sum of weights are always restricted to the same value. For weights restricted to the unit simplex but not summing to one, it is significantly more complicated, as we now have a non-linear state space model. Although the benefit of having weights restricted to the unit simplex is interpretability, there is no real gain in terms of forecasting accuracy in such restriction (Diebold, 1991), just as portfolios allowed to hold short positions can improve on long only portfolios. In the dynamic setting in Eqs. (9), restricting the weights possibly leads to an under-performance compared to the unrestricted case. For example, consider the case where all models overestimate the quantity of interest by some positive value. Under the restrictive case, there is no combination of weights that can achieve that quantity, while the unrestrictive case can by imposing some negative coefficients. For these reasons, we utilise the unrestricted dynamic weighting scheme implied by Eqs. (9) instead of the conventional restricted variations.

2.2 Estimation Strategy and Forecasting

Estimation for the decouple step is straightforward, depending on the model assumptions for each subgroup. For (dynamic) linear regressions, we can sample each $h_j(x_j) = p(y|\mathcal{A}_j)$ using conjugate updates. As for the recouple step using BPS, some discussion is needed. In particular, the joint posterior distribution of the latent states and the structural parameters is not available in closed form. In our framework, the latent states are represented by the predictive densities of the models, $\mathcal{A}_j, j = 1, \dots, J$, and the synthesis parameters, Φ_t . We implement a Markov Chain Monte Carlo (MCMC) approach

using an efficient Gibbs sampling scheme, which is detailed in Appendix A. Marginal posterior distributions of quantities of interest are computed as mixtures of the model-dependent marginal predictive densities weighted by the synthesis implied by $\alpha_t(y_t|\mathbf{x}_t, \Phi_t)$. Integration over the models space is performed using our MCMC scheme, which provides consistent estimates of the latent states and parameters.

Posterior estimates of the latent states \mathbf{x}_t provide insights into the nature of the conditional dependencies across the subgroups, as well as subgroup characteristics. The MCMC algorithm involves a sequence of standard steps in a customised two-component block Gibbs sampler: the first component simulates from the conditional posterior distribution of the latent states given the data, past forecasts from the subgroups, and the synthesis parameters. This is the “learning” step, whereby we learn the biases and inter-dependencies of the latent states. The second step samples the predictive synthesis parameters, that is, we “synthesise” the models’ predictions by effectively mapping the biases and inter-dependencies learned in the first step onto parameters in a dynamic manner. The second step 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). At each iteration of the sampler we sequentially cycle through the above steps. 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 $\boldsymbol{\theta}_0 \sim N(\mathbf{m}_0, \mathbf{C}_0 v_0/s_0)$ and $1/v_0 \sim G(n_0/2, n_0 s_0/2)$, and discount factors (β, δ) are specified.

In terms of forecasting, at time t , we generate predictive distributions of the object of interest as follows: (i) For each sampled Φ_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. (9b)– this gives a draw $\Phi_{t+1} = \{\boldsymbol{\theta}_{t+1}, v_{t+1}\}$ from $p(\Phi_{t+1}|y_{1:t}, \mathcal{H}_{1:t})$; (ii) draw \mathbf{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. (9a). Repeating, this generates a random sample from the 1-step ahead 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. The BPS modelling framework provides a natural and flexible procedure to recouple subgroups over multiple horizons.

In the BPS framework, there are two ways to forecast over multiple horizons, through traditional DLM updating or through customised 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, customised 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 customised 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.

3 Empirical Study

3.1 Research Design

To shed light on the predictive ability of our decouple-recouple model synthesis strategy, we calibrate and test the models in two different scenarios: (1) a macroeconomic application, which relates to the monthly forecast on the U.S. annual inflation using a large set of macroeconomic and financial variables, and (2) a finance application concerning the forecasting of the monthly year-on-year stock returns in excess of the risk-free rate across different industries. For both applications, for the decouple step we use a dynamic linear model (DLM: West and Harrison, 1997; Prado and West, 2010), for each subgroup, $j = 1:J$,

$$y_t = \boldsymbol{\beta}'_{tj} \mathbf{x}_{tj} + \epsilon_{tj}, \quad \epsilon_{tj} \sim N(0, \nu_{tj}), \quad (11)$$

$$\boldsymbol{\beta}_{tj} = \boldsymbol{\beta}_{t-1,j} + \mathbf{u}_{tj}, \quad \mathbf{u}_{tj} \sim N(0, \nu_{tj} \mathbf{U}_{tj}), \quad (12)$$

where the coefficients follow a random walk and the observation variance evolves with discount stochastic volatility, see, e.g., Dagl and Halling (2012), Koop and Korobilis (2013), Gruber and West (2016), Gruber and West (2017) and Zhao, Xie, and West (2016). Priors for each decoupled predictive regression are assumed rather uninformative, such as $\boldsymbol{\beta}_{0j}|v_{0j} \sim N(\mathbf{m}_{0j}, (v_{0j}/s_{0j})\mathbf{I})$ with $\mathbf{m}_{0j} = \mathbf{0}'$ and

$1/v_{0j} \sim G(n_{0j}/2, n_{0j}s_{0j}/2)$ with $n_{0j} = 10, s_0 = 0.01$. The discount factors for the conditional volatilities in Eq. (11) are set to $(\beta, \delta) = (0.95, 0.99)$. For the recouple step, we follow the synthesis function in Eq. (8), with the following marginal priors: $\boldsymbol{\theta}_0|v_0 \sim N(\mathbf{m}_0, (v_0/s_0)\mathbf{I})$ with $\mathbf{m}_0 = (0, \mathbf{1}'/J)'$ and $1/v_0 \sim G(n_0/2, n_0s_0/2)$ with $n_0 = 10, s_0 = 0.01$. The discount factors are the same as in the decouple step.

For both studies, we compare our framework against a variety of competing predictive strategies. First, we compare the aggregate predictive density from the BPS (see Eq.(9a)-(9b)) against the predictive densities from each subgroup regressions calculated from Eq.(11)-(12). That is we test the benefits of the recoupling step and the calibration of the aggregate model prediction upon latent bias and interdependencies. 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 penalised least-squares regression, i.e.,

$$\hat{\boldsymbol{\beta}}_{LASSO} = \arg \min_{\boldsymbol{\beta}} \| y_{1:t} - \boldsymbol{\beta}' \mathbf{x}_{1:t} \|_2^2 + \lambda \sum_{i=1}^n |\beta_i|$$

where the shrinkage parameter λ is calibrated by a leave-one-out cross-validation, that is the model is trained by using the whole sample up to $t - 1$ (cross-validation training set) and the shrinkage parameter is selected based on the prediction accuracy at time t (cross-validation test set). 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 modelling where factors are latent and extracted from the set of predictors. That is we compare our DRS against a dense modelling benchmarking framework (see, e.g., Stock and Watson 2002). More precisely, the factor model relates each y_t to an underlying vector of $q < n$ of random variables \mathbf{f}_t , the latent common factors, via

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

where (i) the factors \mathbf{f}_t are independent with $\mathbf{f}_t \sim N(0, I_q)$, (ii) the ϵ_t are independent and normally distributed with a discount-factor volatility dynamics, (iii) $u_t \perp \mathbf{f}_s \forall s, t$, 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 by a BIC information criterion. Also, we assume that the factor regression betas on the latent factors are time-varying and follow a dynamic linear model consistent with the DRS specification above. Precisely, at each time t we replace \mathbf{x}_{tj} with \mathbf{f}_t in Eq. (11) and the slope parameters have a random walk dynamics as in Eq. (12).

The fourth competing strategy is a relatively standard Bayesian Model Averaging in which the forecast densities are mixed with respect to sequentially updated model probabilities (e.g. Harrison and Stevens, 1976; West and Harrison, 1997, Sect 12.2). In particular, subgroup-specific predictive density can be interpreted as a model combination scheme, whereby the weights are restricted to be inside the unit circle and the j th sub-model is restricted to have weight equal to one. This allows to compare the benefit of the predictive density calibration that is featured the recoupling step of the Bayesian predictive synthesis framework underlying our DRS.

A fifth competing predictive strategy is linear pooling of predictive densities such that,

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

where the restrictions on the weights w_i are necessary and sufficient to assure that $p(y_{t+k})$ is a density function for all values of the weights and all arguments of the sub-model density functions (see, e.g., Geweke and Amisano 2011). Choice of weights in any forecast combination is widely regarded as a difficult and important question. Existing literature showed 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., Clemen 1989, Timmermann 2004, Smith and Wallis 2009, and Diebold and Shin 2017). For this reason, we opt for a specification in which pooling of predictive densities is such that each sub-model has the same weight in the aggregate forecast, i.e., $w_j = 1/J$. Finally, in the finance application related to predicting future stock returns we also compare DRS against the prediction from the historical average, as suggested by Campbell and Thompson (2007) and Goyal and Welch (2008).

3.2 Statistical Performance and Economic Significance

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 a Root Mean Squared Error (RMSE). Ideally, one also wants to compare the predictive densities across strategies. As a matter of fact, performance measures based on the obtained 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. We compare predictive strategies based on the log predictive density ratios (LPDR); at horizon k and across time indices t , that is,

$$\text{LPDR}_t(k) = \sum_{i=1}^t \log\{p(y_{i+k}|y_{1:i}, \mathcal{M}_s)/p(y_{i+k}|y_{1:i}, \mathcal{M}_0)\}, \quad (13)$$

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.

While it is not obvious how to measure economic gains when it comes to assess inflation predictions, it is fairly natural to isolate the economic benefits of our DRS strategy against competing benchmark within the context of predicting future stock returns. As often in the empirical finance literature we evaluate the economic significance of return forecasts 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. 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

$$\omega_\tau^* = \arg \max_{\omega_\tau} E[U(\omega_\tau, y_{\tau+1}) | \mathcal{H}_\tau], \quad (14)$$

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

to have power utility

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

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 his available information set at that time. Taking expectations with respect to the predictive density in Eq. (7), we can rewrite the optimal portfolio allocation as

$$\omega_\tau^* = \arg \max_{\omega_\tau} \int U(\omega_\tau, y_{\tau+1}) p(y_{\tau+1} | \mathcal{H}_\tau) dy_{\tau+1}, \quad (16)$$

As far as DRS is concerned, the integral in Eq. (16) can be approximated using the draws from the predictive density in Eq. (7). The sequence of portfolio weights ω_τ^* , $\tau = 1, \dots, t$ is used to compute the investor's realised utility for each model-combination scheme. Let $\hat{W}_{\tau+1}$ represent the realised wealth at time $\tau + 1$ as a function of the investment decision,

$$\hat{W}_{\tau+1} = \left[(1 - \omega_\tau^*) \exp(r_\tau^f) + \omega_\tau^* \exp(r_\tau^f + y_{\tau+1}) \right], \quad (17)$$

The certainty equivalent return (CER) for a given model is defined as the annualised value that equates the average realised utility. We follow Pettenuzzo, Timmermann, and Valkanov (2014) and compare the the average realised utility of DRS \hat{U}_τ to the average realised utility of the model based on the alternative predicting scheme i , over the forecast evaluation sample:

$$CER_i = \left[\frac{\sum_{\tau=1}^t \hat{U}_{\tau,i}}{\sum_{\tau=1}^t \hat{U}_\tau} \right]^{\frac{1}{1-\gamma}} - 1, \quad (18)$$

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. (17). We interpret a negative CER_i as evidence that model i generates a lower (certainty equivalent) return than our DRS predictive modelling.

3.3 Macroeconomic application: Forecasting Inflation

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 1986/1 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 follows; first, the decoupled models are analysed in parallel over 1986/1-1993/6 as a training period, simply estimating the DLM in Eq. (11) to the end of that period to estimate the forecasts from each subgroup. This continues over 1993/7-2015/12, 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 1993/7 up to time t . We discard the forecast results from 1993/7-2000/12 as training data and compare predictive performance from 2001/1-2015/12. 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. These periods exhibit sharp shocks to the U.S. economy in general, and possibly provide shifts in relevant predictors and their inter-dependencies. We consider both 1- and 3-step ahead forecasts, in order to reflect interests and demand in practice.

Panel A of Table 1 shows that our decouple-recouple strategy using BPS 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.

[Insert Table 1 about here]

Similarly, Panel B of Table 1 shows that DRS for the 3-step ahead forecasts reflect a critical benefit of using BPS for the recoupling step 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 changing.

Delving further into the dynamics of the LDPR, Figure 1 shows the one-step ahead out-of-sample performance of DRS in terms of predictive density. The figure makes clear that the out-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 predict 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 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.

[Insert Figure 1 about here]

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 regularisation 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. However, 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.

Top panel of Figure 2 highlights a first critical component of using BPS in the recouple step,

namely learning the latent inter-dependencies among and between the subgroups in order to maintain economic interpretability and reduce the overall model variance. Precisely, the figure reports the latent BPS coefficients rescaled such that they are bounded between zero and one and sum to one. This allows to give a clearer interpretability of the relative importance of these latent interdependencies through time. 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.

[Insert Figure 2 about here]

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 1-step ahead forecasts, bottom panel of Figure 2 shows that there are specific differences in the dynamics of the latent interdependencies 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 which is 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 BPS for multi-step ahead modelling.

Looking at the overall bias, i.e., the conditional intercept, Figure 3 clearly show that switch sign in the aftermath of the short recession in the early 2000s and the financial crisis of 2008/2009.

[Insert Figure 3 about here]

Since the parameters of the recoupling step are considered to be latent states, the conditional intercept

can be interpreted as a free-roaming component, which is not directly pinned down by any group of predictors. In this respect, and for this application, the time variation in the conditional intercept can be thought of as a reflection of unanticipated economic shocks, which then affect inflation forecasts with some lag. 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 1-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, both forecasts bias 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.

3.4 Finance application: Forecasting Industry Stock Returns

We consider a large set of predictors to forecast monthly total excess returns across different industries from 1970/1 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 70 pre-calculated financial ratios for all U.S. companies across eight different categories. 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 aggregation 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. Those 70 ratios are classified in eight main categories: Valuation, Profitability, Capitalisation, Financial Soundness, Solvency, Liquidity, Efficiency Ratios, and Other.

Together with industry-specific predictors, we use additional 14 aggregate explanatory variables which are divided in two additional categories; aggregate financials and macroeconomic variables. In particular, following Goyal and Welch (2008) and Rapach et al. (2010), the market-level, aggregate, financial predictors consist of the monthly realised 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 capitalisation (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.

As far as the aggregate macroeconomic predictors are concerned, we utilise 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 authorised 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 DLM specification in Eq.(11) is attractive due to its parsimony, ease to compute, and the smoothness it induces to the parameters (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). For the recouple step, we follow the synthesis function in Eq. (8), with the following priors: $\theta_{0n}|v_{0n} \sim N(\mathbf{m}_{0n}, (v_{0n}/s_{0n})\mathbf{I})$ with $\mathbf{m}_{0n} = \mathbf{0}'$ and $1/v_{0n} \sim G(n_{0n}/2, n_{0n}s_{0n}/2)$ with $n_{0n} = 12, s_{0n} = 0.01$. The discount factors are $(\beta, \delta) = (0.99, 0.95)$.

The empirical application is designed similarly to the macroeconomic study. We used, as training period for the decoupled models, the sample 1970/1-1992/9, fitting the liner regression in a expanding window manner for each industry. Over the period 1992/10-2015/12 we continue the calibration of the recouple strategies. We discard the forecast results from 1993/7-2000/12 as training data and compare predictive performance from 2001/1-2015/12. 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 in both biases and the dynamics of 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 HA. These results hold for all the ten industries under investigation.

[Insert Table 2 about here]

Similar to the macroeconomic study, the performance gap in favour 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 the LPDR, Figure 4 shows the whole out-of-sample path of density forecasting accuracy across modelling 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 favourably against alternative predictive strategies. Similar results appear in other sectors.

As a whole, Figure 4 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.

[Insert Figure 4 about here]

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, as highlighted by the DLM-type of dynamics in Eqs. (9). 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.

Figure 5 shows that there is a substantial flexibility in the DRS coefficients and some interesting aspects related to returns predictability emerge.² For instance, 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.

[Insert Figure 5 about here]

²As above, the figure reports the latent interdependencies rescaled such that they are bounded between zero and one and sum to one.

Although the interpretation of the dynamics of the latent interdependencies is not always clean, some interesting picture emerge. Take the Other sector as an example; in the 10-industry classification we used, the Other sector is composed by business services, constructions, building materials, financial services, and banking. The financial capacity of all these industries, especially the banking and finance sector, has been significantly affected after the collapse of Lehman in the fall of 2008. As a matter of fact, on the one hand, anecdotal evidence and policy making commentaries highlighted how the increasing burden, due to a huge amount of non-performing loans in the banking sectors, ultimately affected those sectors more dependent on bank financing, such as construction and building materials. On the other hand, while the regime of low policy rates might have, in the short term, helped to prevent a disorderly adjustment of balance sheets in distressed banks and provided relief in terms of lower interest payments in those more exposed to mortgages, they also weakened the incentive to repair balance sheets of banks and building societies in the first place. As a result, the joint effect of moral-hazard issues and the massive amount of non-performing loans and the subsequent risk capacity of financial intermediaries represented significant sources of financial risk.

Although there are some similarities in the recoupling dynamics across industries, some cross-sectional heterogeneity emerge as well. For instance, for few industries, e.g., other, manufacturing, and consumer non-durable, profitability tend to play a significant role in the aggregate predictive density until the great financial crisis of 2008/2009.

As a whole, Figure 5 provide substantial evidence on the out-of-sample instability in the latent interdependencies across group of predictive densities over time. However, one comment is in order. It should be clear that our goal here is not to over-throw other results from the empirical finance literature with respect to the correlation among predictors, but to deal with the crucial aspect of modelling the dynamic interplay between different, economically motivated, predictive densities in forecasting excess stock returns.

The time variation in the latent interdependencies is reflected in the aggregate dynamic bias which is sequentially corrected within the BPS framework. Figure 6 shows the dynamics of the calibrated bias across different industries.

[Insert Figure 6 about here]

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. That is, the aggregate predictive density

that is synthesised from each class of predictors is significantly recalibrated around periods of market turmoil.

3.4.1 Economic Significance. We now investigate the economic significance of our DRS compared to the competing predictive strategies. Throughout the empirical analysis we take the perspective of a representative investor with power utility and moderate relative risk aversion, $\gamma = 5$. Panel A of Table 3 shows the results for portfolios with unconstrained weights, i.e. short sales are allowed to maximise the portfolio returns.

[Insert Table 3 about here]

The economic performance of our decouple-recouple strategy is rather stark in contrast to both group-specific forecasts and the competing forecasts combination schemes. The realised CER from DRS is much larger than virtually any of the other model specifications across different industries. Not surprisingly, given the statistical accuracy of a simple recursive historical mean model is not remarkable, the HA model leads to a very low CER. 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 favour of DRS is again confirmed under the restriction that the portfolio weights have to be positive, i.e., long-only strategy. Our decouple-recouple model synthesis scheme allows a representative investor to obtain 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

$$CER_{i\tau} = \left[\frac{\hat{U}_{\tau,i}}{\hat{U}_{\tau}} \right]^{\frac{1}{1-\gamma}} - 1, \quad (19)$$

Similarly to Eq (18), we interpret a negative $CER_{i\tau}$ 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 annualised, single-period CER for the forecasting sample for an unconstrained investor. The results

show that the out-of-sample performance is robustly in favour 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.

[Insert Table 4 about here]

Panel B shows the results for a short-sales constrained investor. Although the gap between DRS and the competing forecast combination schemes is 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 learn latent interdependencies and biases improve the out-of-sample economic performance within the context of typical portfolio allocation example. To parallel the LPDR in Eq. (13), we also inspect the economic performance of the individual model combination schemes by reporting the cumulative sum of the CERs over time:

$$CCER_{it} = \sum_{\tau=1}^t \log(1 + CER_{i\tau}), \quad (20)$$

where CER_{it} is calculated as in Eq. (19). Figure 7 shows the out-of-sample cumulative CER across the forecasting sample and for the Consumer durable, Consumer non-durable, Telecomm, Health, Shops and Other industrial sectors. Except few nuances, e.g., the pre-crisis period for Telecomm and Other, the DRS combination scheme constantly outperforms the other predictive strategies.

[Insert Figure 7 about here]

Interestingly, although initially generate a good certainty equivalent return, the LASSO failed to adjust to the abrupt underlying changes in the predictability of industry returns around the crisis. As a matter of fact, despite the initial cumulative CER is slightly in favour 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, the 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 8 shows the out-of-sample cumulative CER across the forecasting sample and 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, i.e. no-short sale constraints.

[Insert Figure 8 about here]

The picture that emerges is the same. Except 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.

4 Conclusion

In this paper, we propose a framework for predictive modelling 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 summarise the whole information in an unsupervised manner, e.g., PCA and factor models.

We calibrate and implement the proposed methodology on both a macroeconomic and a finance application. We compare forecasts from our framework against sequentially updated Bayesian model averaging (BMA), equal-weighted linear pooling, LASSO-type regularization, as well as a set of simple predictive regressions, one for each class of predictors. 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 total excess returns across different industries in the U.S market.

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. 2010. Instrumental variable estimation in a data rich environment. *Econometric Theory* 26:1577–1606.
- Belloni, A., D. Chen, V. Chernozhukov, and C. Hansen. 2012. Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica* 80:2369–2429.
- Bernanke, B. S., J. Boivin, and P. Elias. 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* published online July 10. ArXiv:1607.02655.
- 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.
- DeMiguel, V., L. Garlappi, and R. Uppal. 2007. Optimal versus naive diversification: How inefficient is the $1/N$ portfolio strategy? *The review of Financial studies* 22:1915–1953.

- Diebold, F. X. 1991. A note on Bayesian forecast combination procedures. In *Economic Structural Change*, pp. 225–232. Springer.
- Diebold, F. X., and M. Shin. 2017. Beating the Simple Average: Egalitarian LASSO for Combining Economic Forecasts .
- Elliott, G., A. Gargano, and A. Timmermann. 2013. Complete subset regressions. *Journal of Econometrics* 177:357–373.
- Freyberger, J., A. Neuhierl, and M. Weber. 2017. Dissecting characteristics nonparametrically. Tech. rep., National Bureau of Economic Research.
- 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.
- Irie, K., and M. West. 2016. Bayesian emulation for optimization in multi-step portfolio decisions. *ArXiv* 1607.01631.
- Jagannathan, R., and T. Ma. 2003. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance* 58:1651–1683.
- 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., and D. Korobilis. 2013. Large time-varying parameter VARs. *Journal of Econometrics* 177:185–198.
- 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.
- Manzan, S. 2015. Forecasting the distribution of economic variables in a data-rich environment. *Journal of Business & Economic Statistics* 33:144–164.
- McAlinn, K., K. A. Aastveit, J. Nakajima, and M. West. 2017. Multivariate Bayesian Predictive Synthesis in Macroeconomic Forecasting. *arXiv preprint arXiv:1711.01667* .
- 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.
- Mönch, E. 2008. Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented VAR approach. *Journal of Econometrics* 146:26–43.
- 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.
- Pesaran, M. H., and A. Timmermann. 2002. Market timing and return prediction under model instability. *Journal of Empirical Finance* 9:495–510.
- 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.
- Ročková, V. 2018. Bayesian estimation of sparse signals with a continuous spike-and-slab prior. *The Annals of Statistics* 46:401–437.
- Ročková, V., and E. I. George. 2016. The spike-and-slab lasso. *Journal of the American Statistical Association* .
- 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.
- Stambaugh, R. F. 1999. Predictive regressions. *Journal of Financial Economics* 54:375–421.
- Stevanovic, D. 2017. Macroeconomic forecast accuracy in a data-rich environment. *Working Paper* .

- 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. 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.
- Zhao, Z. Y., M. Xie, and M. West. 2016. Dynamic dependence networks: Financial time series forecasting & portfolio decisions (with discussion). *Applied Stochastic Models in Business and Industry* 32:311–339. ArXiv:1606.08339.

Appendix

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 accumulates. 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 $\mathbf{F}_t = (1, x_{t1}, \dots, x_{tJ})'$ 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 $p(\mathbf{x}_{1:T}, \mathbf{\Phi}_{1:T} | y_{1:T}, \mathcal{H}_{1:T})$. These two conditional posteriors and algorithmic details of their simulation are as follows.

A.2 Sampling the synthesis parameters $\Phi_{1:T}$

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

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

has known elements $\mathbf{F}_t, \mathbf{W}_t$ and specified initial prior at $t = 0$. The implied conditional posterior for $\Phi_{1:T}$ then does not depend on $\mathcal{H}_{1:T}$, reducing to $p(\Phi_{1:T} | \mathbf{x}_{1:T}, y_{1:T})$. 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:*

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

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

2. *Update to time t prior:*

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

with (unchanged) point estimates \mathbf{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:* $y_t | \mathbf{x}_{1:t}, y_{1:t-1} \sim T_{\beta n_{t-1}}(f_t, q_t)$ where

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

4. *Filtering update to time t posterior:*

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

with defining parameters as follows:

- i. For $\boldsymbol{\theta}_t|v_t$: $\mathbf{m}_t = \mathbf{m}_{t-1} + \mathbf{A}_t e_t$ and $\mathbf{C}_t = r_t(\mathbf{R}_t - q_t \mathbf{A}_t \mathbf{A}_t')$,
- 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’’) $\mathbf{A}_t = \mathbf{R}_t \mathbf{F}_t / q_t$, 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 $p(\boldsymbol{\Phi}_T|\mathbf{x}_{1:T}, y_{1:T})$ as follows. First, draw v_T^{-1} from $G(n_T/2, n_T s_T/2)$, and then draw $\boldsymbol{\theta}_T$ from $N(\mathbf{m}_T, \mathbf{C}_T v_T/s_T)$.
- b. *Recurse back over times $t = T - 1, T - 2, \dots, 0$:* At time t , sample $\boldsymbol{\Phi}_t = (\boldsymbol{\theta}_t, v_t)$ 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 $p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, v_t, \mathbf{x}_{1:T}, y_{1:T})$ with mean vector $\mathbf{m}_t + \delta(\boldsymbol{\theta}_{t+1} - \mathbf{m}_t)$ and variance matrix $\mathbf{C}_t(1 - \delta)(v_t/s_t)$.

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

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

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

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 $\mathbf{H}_t = \text{diag}(H_{t1}/\phi_{t1}, H_{t2}/\phi_{t2}, \dots, H_{tJ}/\phi_{tJ})$, 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 \mathbf{x}_t is then sampled, given ϕ_{tj} , from

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

where $c_t = y_t - \theta_{t0} - \mathbf{h}_t' \boldsymbol{\theta}_{t,1:J}$, $g_t = v_t + \boldsymbol{\theta}_{t,1:J}' \mathbf{q}_t \boldsymbol{\theta}_{t,1:J}$, and $\mathbf{b}_t = \mathbf{q}_t \boldsymbol{\theta}_{t,1:J} / g_t$. Here, given the previous values of ϕ_{tj} , we have $\mathbf{H}_t = \text{diag}(H_{t1}/\phi_{t1}, H_{t2}/\phi_{t2}, \dots, H_{tJ}/\phi_{tJ})$. Then, conditional on these new samples of \mathbf{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 $1:T$.

B Further Results on Latent Interdependencies

Finally, we explore the retrospective dependencies of the latent states for the one-step ahead inflation forecasting exercise. For this, we measure the MC-empirical R^2 , which is the variation of one of the retrospective posterior latent states explained by the other latent states. Retrospective, here, means that these measures are computed using all of the data in the testing period, rather than the one-step ahead coefficients of Figure 2. Figure B.1 shows the MC-empirical R^2 for one of the latent states, given all of the other latent states; e.g., variation of Output and Income given Labor Market, Consumption and Orders, etc. There are some clear patterns that emerge. Most latent states are highly dependent with each other, with Output and Income, Labor Market, Orders and Inventories, Money and Credit, and Prices grouping up over the whole period, with increased dependencies measure after the crisis of 2008/2009.

[Insert Figure B.1 about here]

We also note that there are clear trends in terms of decrease in dependencies before the crisis and sharp increase after. This is indicative of the closeness of these groups, as well as how they shift through different economic paradigms. Most interesting is how Interest Rate and Exchange Rates increase during the dot.com bubble, almost to the level of the other highly dependent states, and drops down, and then syncs almost perfectly with Stock Market after 2008. We can infer from this that the dependency characteristics of Interest Rate and Exchange Rates and Stock Market have changed dramatically over the testing period, with the Stock Market being significantly less dependent to the

broader macroeconomy, including Interest Rate and Exchange Rates, the crisis of 2008/2009 shifting the two characteristics to be similar, and finally tapering off at the end again to be less dependent to the other latent states (though we note this is a general trend in all of the latent states).

Figure B.2 further explores the retrospective dependencies showing the pairwise MC-empirical R^2 , which measures the variation explained of one state given another, but now focusing solely on the pair of states. Based on the results in Table 1 we focus on two of the most prominent states: Labor Market (top panel) and Prices (bottom panel). Notice that, due to the symmetry in the dependence structure of the latent predictive densities, the relationship between Labor Market vs Prices and Prices vs Labor Market are the same. The rest have relatively low dependence, with some notable exceptions.

[Insert Figure B.2 about here]

For one, we find that Labor Market and Output and Income to be highly dependent around the build up of the sub-prime mortgage bubble and the consequent great financial crisis of 2008/2009. Money and Credit almost has an inverse relationship, with it decreasing during that period and increasing otherwise. On the other hand, we find that, in terms of Prices, there is a gradual increase of Money and Credit and Orders and Inventories. These changes in coefficients, as well as the retrospective dependencies, are indicative of the structural changes in the economy brought on by crises and shocks, showing that recoupling using BPS successfully learns these trends and is able to provide economic interpretability to the analysis, compared to, for example, BMA, which degenerated to one of the groups, or LASSO, which dogmatically shrinks certain factors to zero.

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. (13). 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	-

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. (13). 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		PCA		DRS	
	HA	Value	Profit	Capital	Soundness	Solvency	Liquidity	Efficiency	Other	Aggregate Fin	Macro	Aggregate Fin	Macro	Aggregate Fin	Macro	Aggregate Fin	Macro	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	0.205	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	0.103	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	0.150	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	0.170	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	0.220	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	0.098	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	0.148	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	0.098	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	0.144	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	0.123	0.146	0.091	0.097	0.143	0.083			

Panel B: Log-Predictive Density Ratio

Industry	Group-Specific Models										LASSO				EW		PCA		DRS	
	HA	Value	Profit	Capital	Soundness	Solvency	Liquidity	Efficiency	Other	Aggregate Fin	Macro	Aggregate Fin	Macro	Aggregate Fin	Macro	Aggregate Fin	Macro	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	-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	-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	-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	-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	-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	-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	-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	-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	-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	-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
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
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

Figure 1. US inflation rate forecasting: Out-of-sample log predictive density ratio

This figure shows the dynamics of the out-of-sample Log Predictive Density Ratio (LPDR) as in Eq. (13) 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.

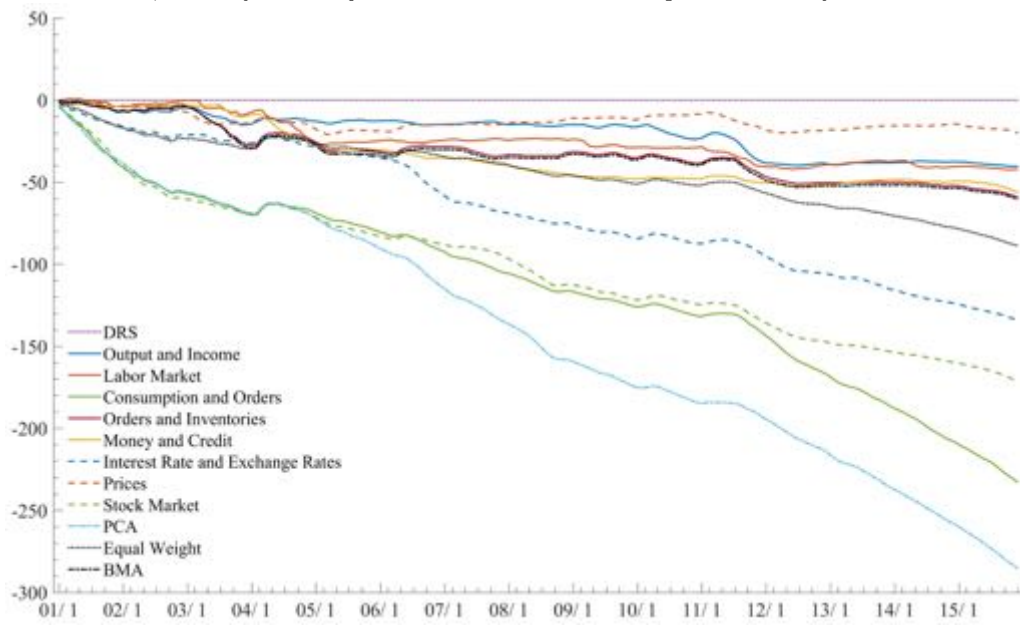
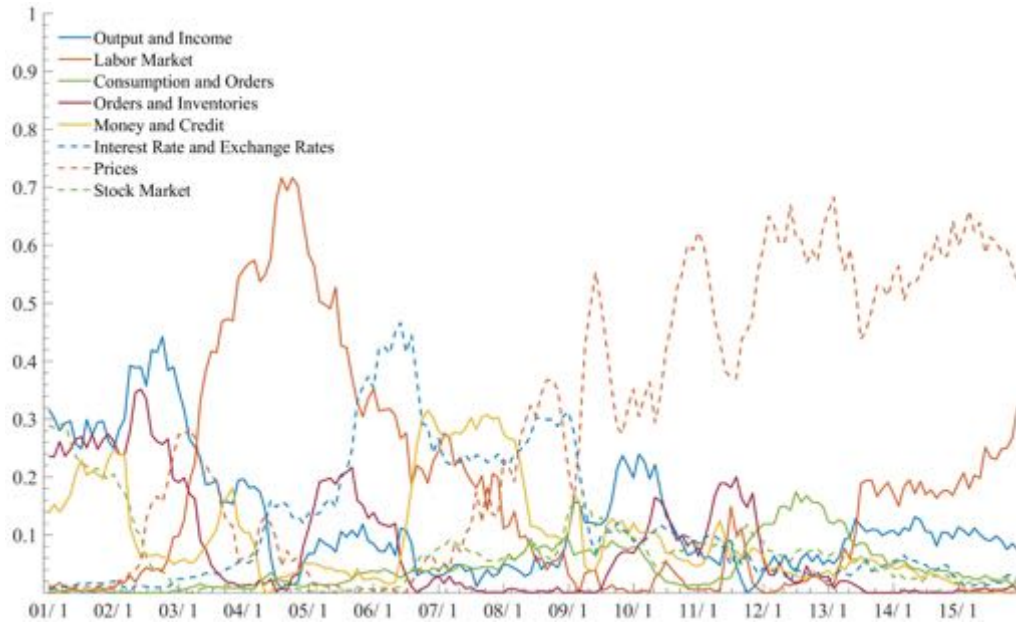
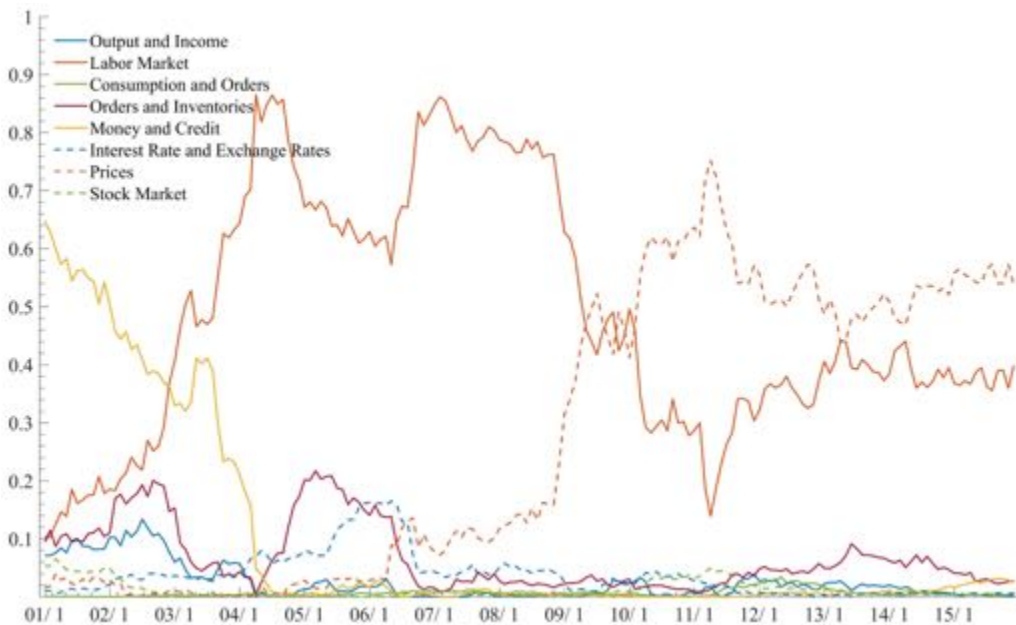


Figure 2. US inflation forecasting: Posterior means of rescaled latent interdependencies.

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. These latent components are sequentially computed at each of the $t = 1:180$ months then rescaled such that they are bounded between zero and one, and sum to one. Top panel shows the results for the one-step ahead forecasting exercise, while bottom panel shows the same results but now for a three-period ahead forecast objective function.



(a) 1-step ahead



(b) 3-step ahead

Figure 3. US inflation rate forecasting: Out-of-Sample Dynamic Predictive Bias

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 period is 01:2001-12:2015, monthly. The objective function is the one-step ahead density forecast of annual inflation.

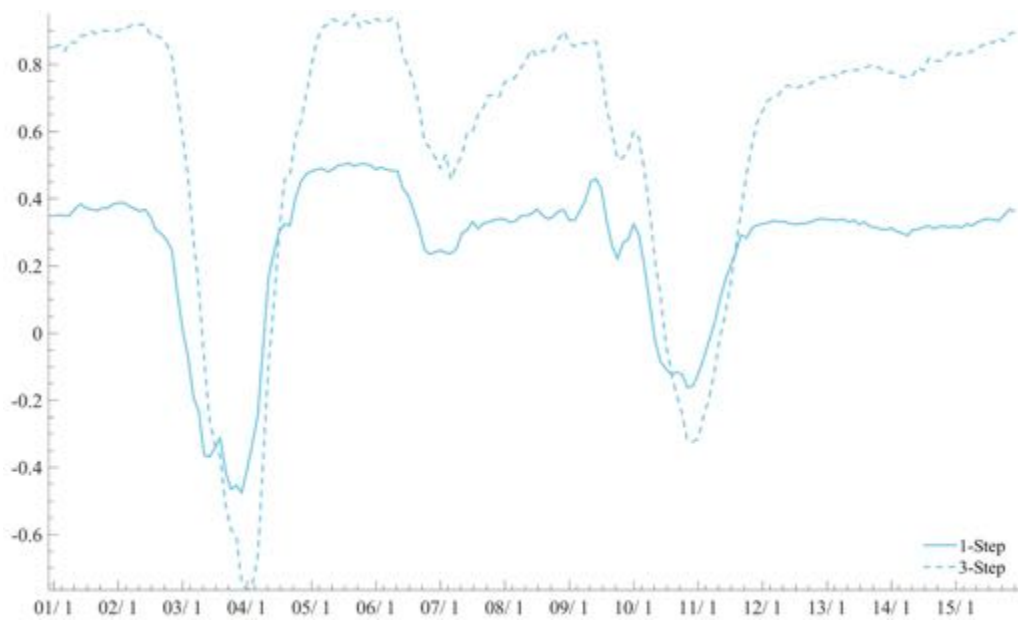


Figure 4. US equity return forecasting: Out-of-sample log predictive density ratio

This figure shows the dynamics of the out-of-sample Log Predictive Density Ratio (LPDR) as in Eq. (13) 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. The sample period is 01:1970-12:2015, monthly.

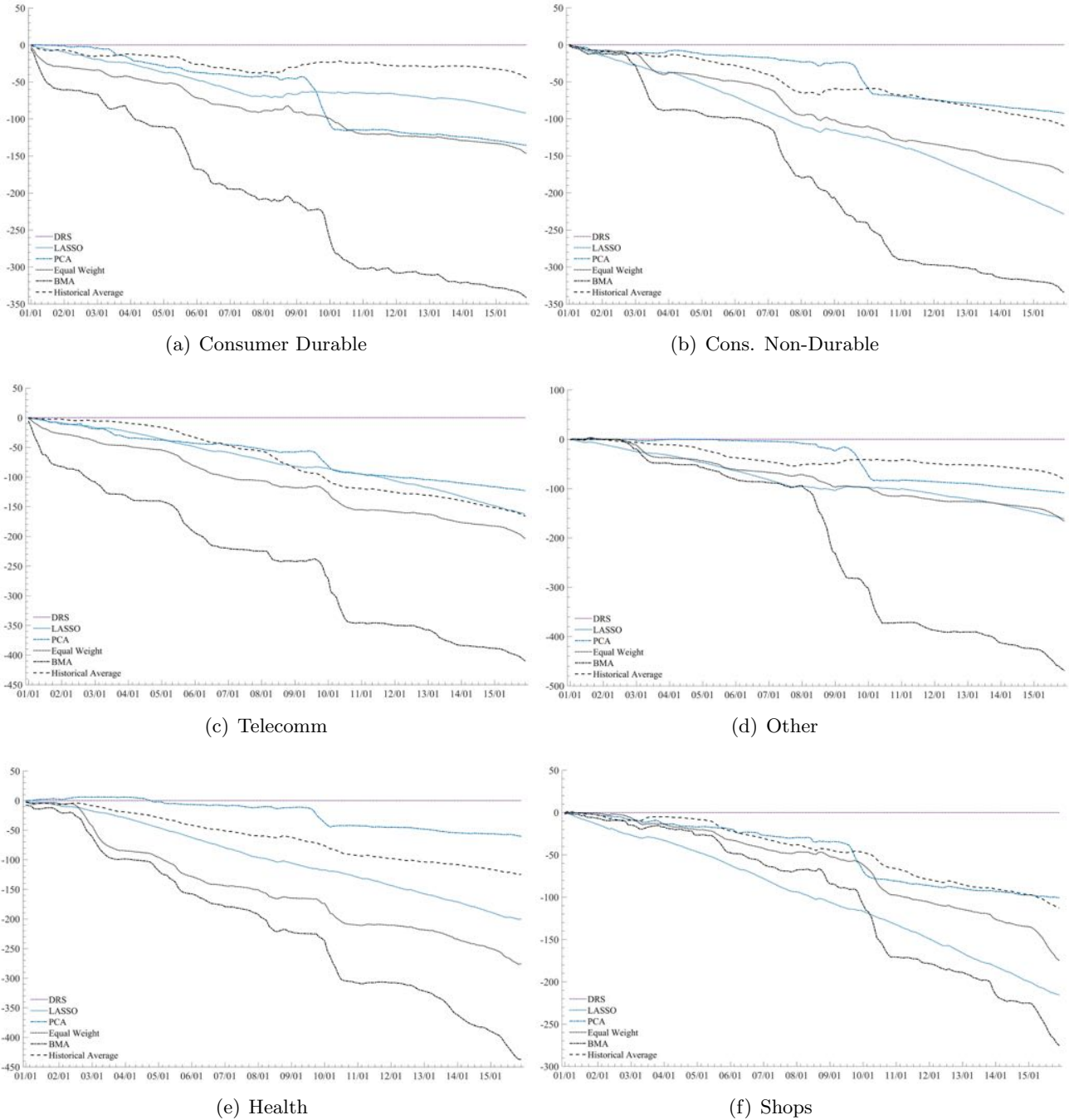
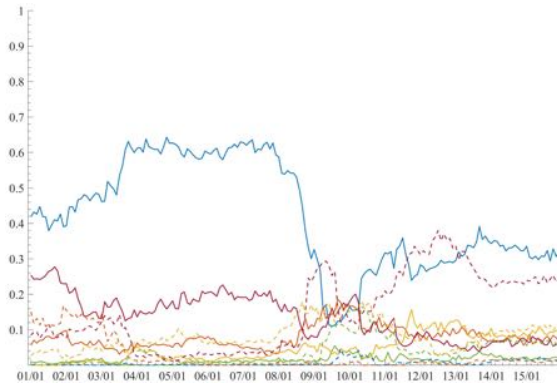
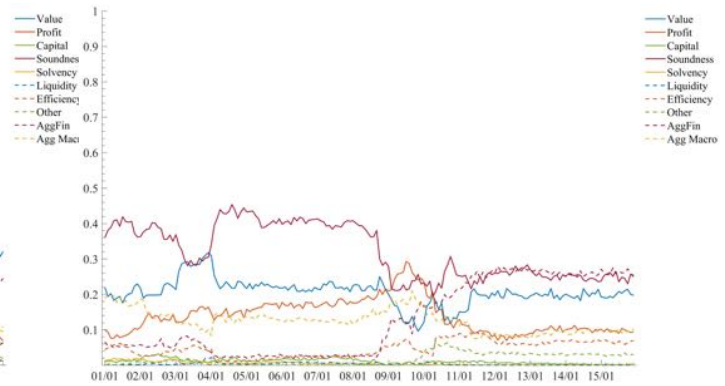


Figure 5. US equity return forecasting: Posterior means of rescaled latent interdependencies.

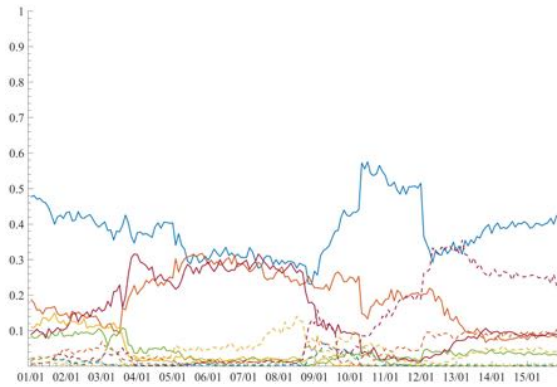
This figure shows the one-step ahead latent interdependencies across groups of predictive densities—measured through the predictive coefficients—used in the recoupling step. For the ease of exposition we report the results for four representative industries, namely, Consumer Durables, Consumer non-Durables, 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.



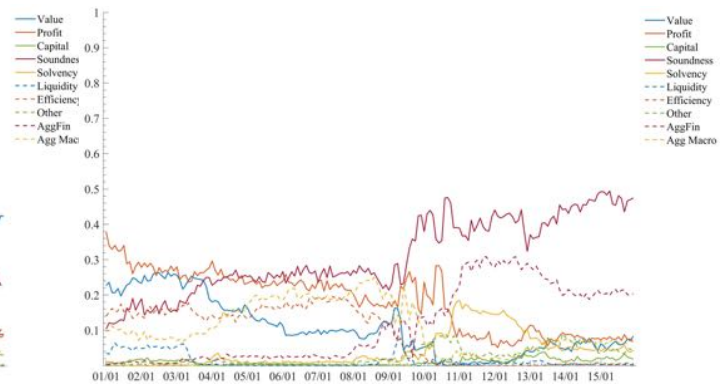
(a) Consumer Durable



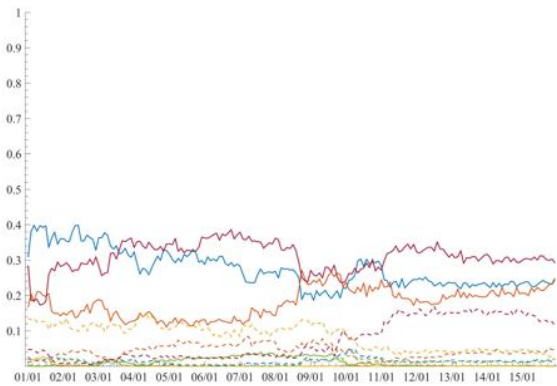
(b) Cons. Non-Durable



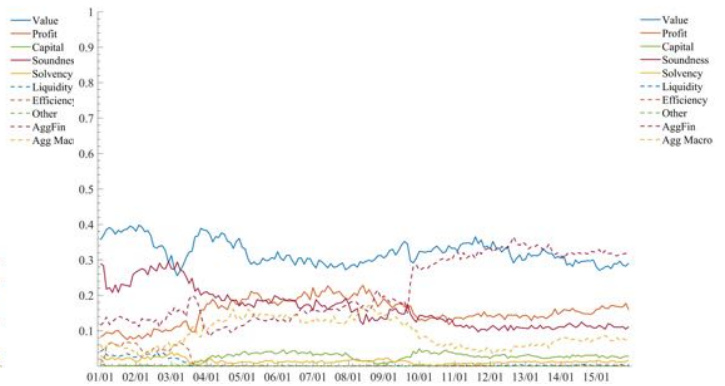
(c) Manufacturing



(d) Other



(e) Utils



(f) Shops

Figure 6. US equity return forecasting: Out-of-Sample Dynamic Predictive Bias

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.

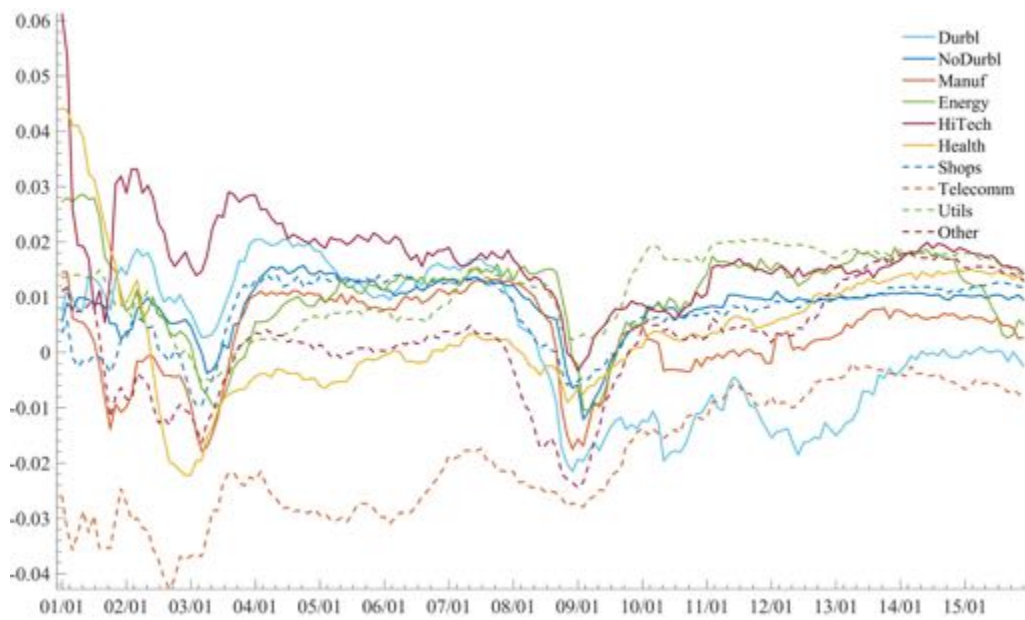


Figure 7. US equity return forecasting: Out-of-sample cumulative CER without Constraints

This figure shows the dynamics of the out-of-sample Cumulative Certainty Equivalent Return (CER) for an unconstrained as in Eq. (20) 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. The sample period is 01:1970-12:2015, monthly.

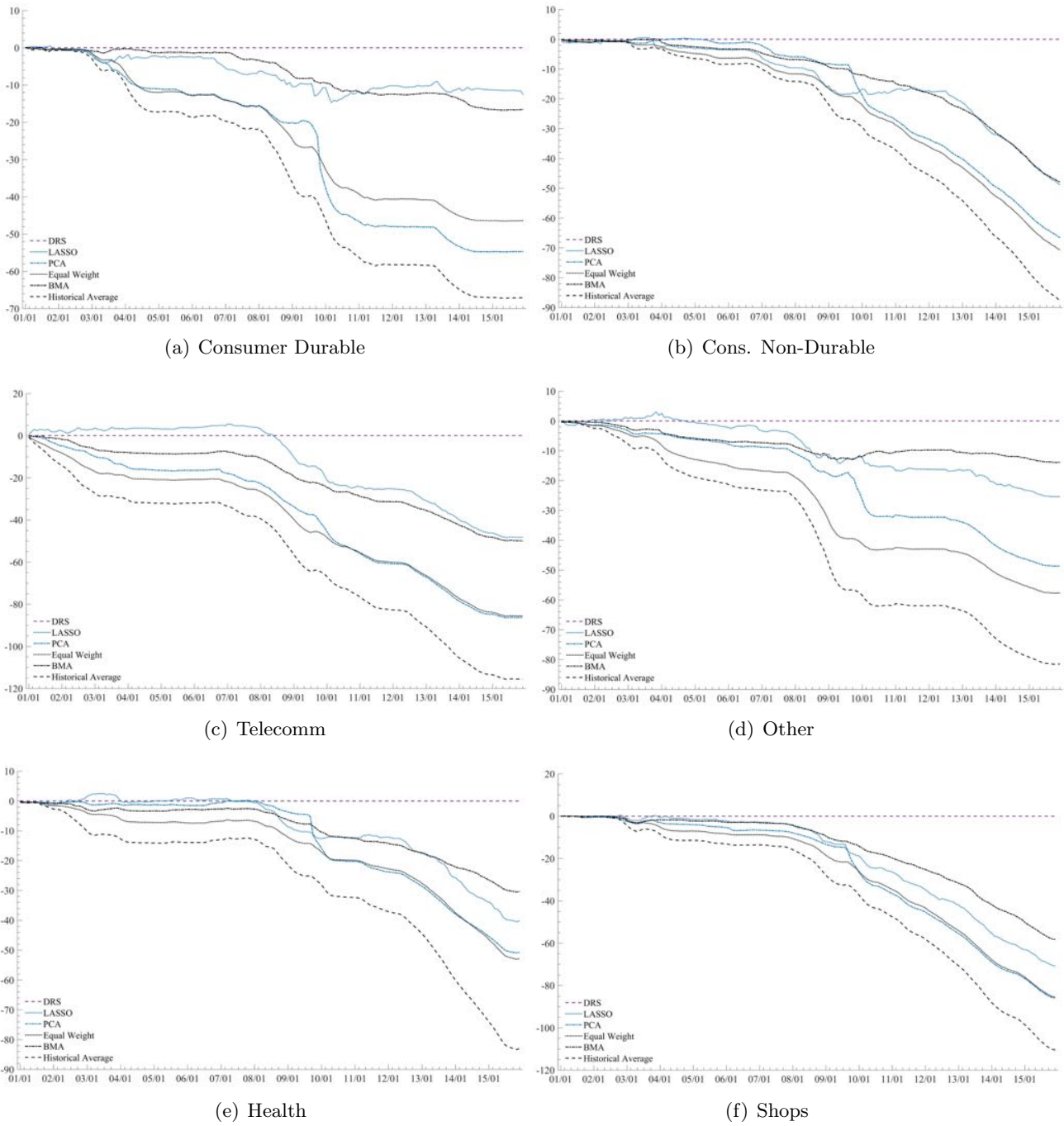


Figure 8. US equity return forecasting: Out-of-sample cumulative CER with short-sale constraints

This figure shows the dynamics of the out-of-sample Cumulative Certainty Equivalent Return (CER) for a short-sale constrained investor as in Eq. (20) 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. The sample period is 01:1970-12:2015, monthly.

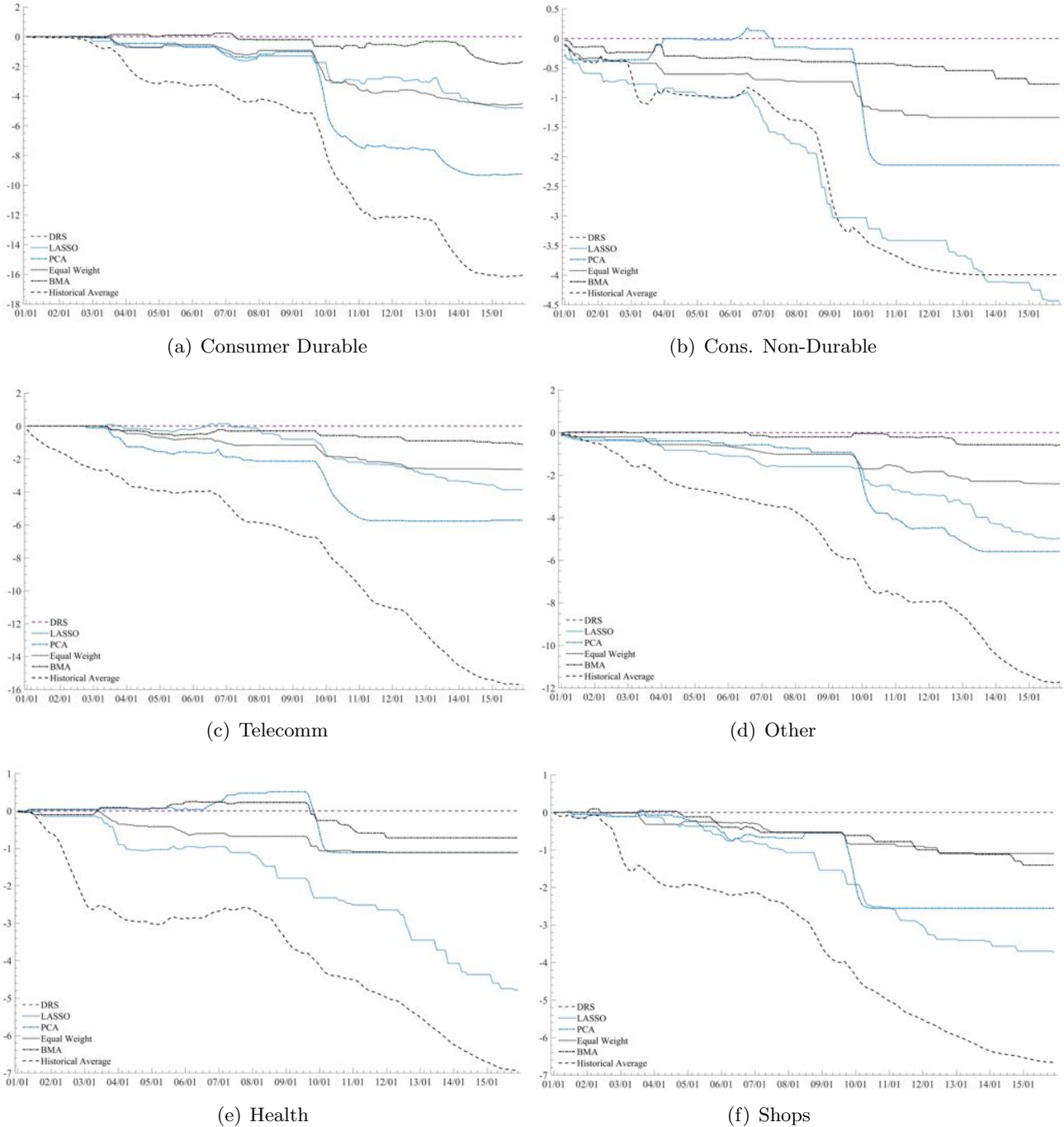


Figure B.1. US inflation rate forecasting: Retrospective latent dependencies

This figure shows the retrospective latent interdependencies 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.

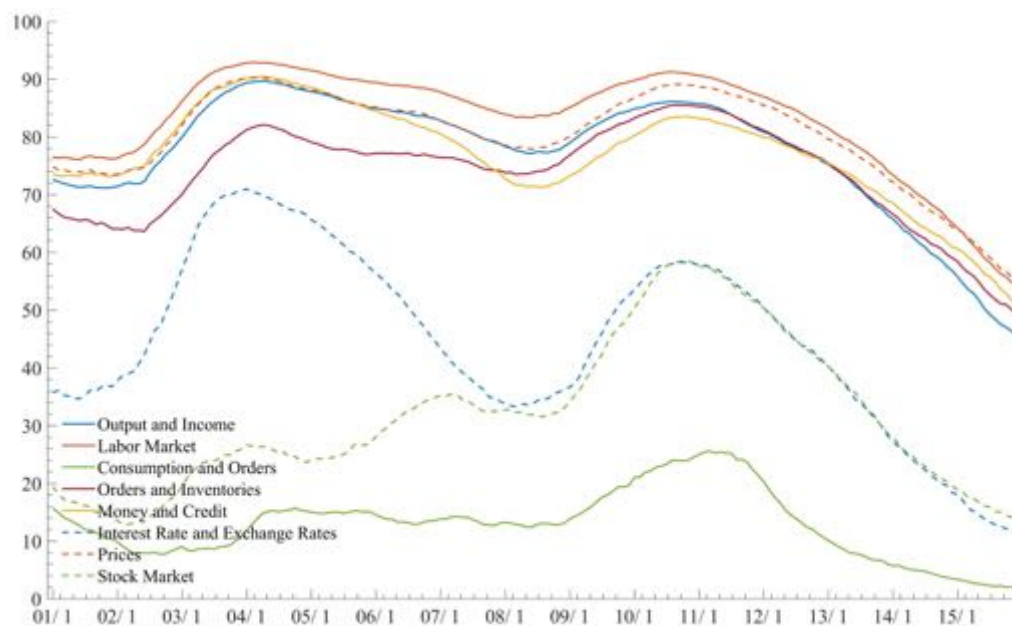


Figure B.2. US inflation rate forecasting: Retrospective latent dependencies (paired)

This figure shows the retrospective paired latent interdependencies 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.

