# Divide and Conquer: Financial Ratios and Industry Returns Predictability*

Daniele Bianchi[†] Kenichiro McAlinn[‡]

**Abstract**

We study whether a large set of financial ratios provides valuable information about future excess stock returns. Confronted with a data-rich environment, we propose a novel "divide and conquer" methodology that allows to efficiently retain all of the information available to investors. In particular, our method does not assume, *a priori*, that some of the financial ratios may be irrelevant or easily reducible. We compare our methodology against standard, recursive sparse and dense predictive regression methodologies, as well as benchmark forecast combination strategies and non-linear machine learning methods. Forecasts based on our method, not only outperform in out-of-sample predictive comparisons, but translate into out-of-sample economic gains that are greater than the historical averages and all of the competing forecasting strategies. Our results lend strong support for using accounting-based information in forecasting stock returns, both at the industry and at the market level.

**Keywords:** Financial ratios, forecast combination, machine learning, returns predictability, data-rich models, industry portfolios.

**JEL codes:** C11, G11, G17, C32, C53, C55.

*This draft: November 16, 2020. We thank Roberto Casarin, Marco Del Negro, Dimitris Korobilis, Andrew Patton, Davide Pettenuzzo, Stijn Van Nieuwerburgh, Mike West (discussant), and seminar participants at the University Cá Foscari of Venice, as well as participants at the NBER-NSF Seminar on Bayesian Inference in Econometrics and Statistics at Stanford, the Barcelona GSE Summer Forum in Time Series Econometrics and Applications for Macroeconomics and Finance, the 10th ECB Workshop on Forecasting Techniques, the world meeting of the International Society of Bayesian Analysis in Edinburgh, and the Workshop on Forecasting at the National Bank of Poland in Warsaw for their helpful comments and suggestions. A previous version of this paper was circulating as "Large-scale dynamic predictive regressions".

†School of Economics and Finance, Queen Mary University of London, London, UK. Email: d.bianchi@qmul.ac.uk Web: whitesphd.com

‡Department of Statistical Science, Fox School of Business, Temple University, PA, USA. Email: kenichiro.mcalinn@temple.edu

# 1 Introduction

Understanding the dynamics of future excess returns, and its implication for portfolio allocation decisions, is one of the fundamental goals of empirical asset pricing. However, risk premia are notoriously difficult to measure, since their time series variation is often buried in the noise of realised returns. The issue of having a low signal-to-noise ratio is addressed by searching for additional sources of conditional information, such as financial ratios, firm characteristics, and/or aggregate macroeconomic variables.[1] As a result, investors are often left with tens – if not hundreds – of possible observables and predictors, each one of them, *a priori*, may provide useful information to capture future expected returns (see Harvey and Liu, 2019 and references therein).

Theoretical models, developed for these studies, offer guidance in identifying which predictors or risk factors may matter. However, it is also the case that these models are often too stylized to explicitly describe all sources of time series variation in financial returns. As a result, investors and researchers, alike, typically face a trade-off: they could pre-select a set of candidate predicting variables for returns by appealing to economic theories, existing empirical literature, and a variety of heuristic arguments, with the risk of omitting important predictors, or, alternatively, they could use the entire set of available predictors in a way that hopefully captures the effective signals.

Even when enlarging the investors' information set, different papers argue in favour of different groups of predictors that capture the time series variation of risk premia, and there is overall little agreement amongst studies (see, e.g., Green et al., 2017, Kelly et al., 2019, DeMiguel et al., 2020, and Freyberger et al., 2020). The difference stems from the fact that existing methods either assume, *a priori*, that only an unknown sub-set of variables carry most of the predictive power – such as shrinkage methods, e.g., lasso and its extensions – or they are based on the assumption that all predictors could bring useful information, although the impact of some of these might be small – such as data compression methods, e.g., PCA and its extensions. [2]

Methodologically, we contribute to the literature on returns predictability and decision making in a data-rich environment by proposing a novel class of Bayesian predictive techniques. We take a significantly different approach – a divide and conquer approach – towards the bias-variance trade-off by breaking a large dimensional regression into a set of smaller dimensional ones. More specifically, we retain all of the information available and *decouple* a large predictive model into a set of smaller regressions, constructed by clustering the set of financial ratios into $J$ different groups, each sub-regression containing significantly fewer regressors than the whole.[3] As a result, rather

[1]See, e.g., Daniel et al. (1997), Lewellen (2004), Green et al. (2017), Kelly et al. (2019), DeMiguel et al. (2020), and Freyberger et al. (2020).

[2]Both of these approaches entail either an implicit or explicit reduction of the model space with the intention to arbitrarily lower model complexity in order to potentially minimize predictive loss. For instance, in penalised regressions, increasing the tuning parameter (i.e., increasing shrinkage) leads to a higher bias, thus, by utilizing cross-validation, the researcher aims to balance the bias-variance trade-off by adjusting the tuning parameter. Similarly, in factor models, the optimal number of latent common components is chosen by using some information criterion to reduce the model variance at the cost of increasing the bias (see, e.g., Bai and Ng, 2002).

[3]Throughout the paper we use the terms "sub-model", "sub-regression" or "Group-specific predictive regressions"

than assuming, *a priori*, the existence of a sparse structure, or few latent common components, we make use of all of the information by estimating $J$ different predictive regressions – one for each group of predictors – for the same target variable and *recouple* them, dynamically, to generate aggregate coherent predictions for decision making. By decoupling a large predictive regression model into smaller, less complex regressions, we keep the aggregate model variance low, while sequentially learning and correcting for the mis-specification bias that characterize each group of predictors.

Empirically, we apply our proposed methodology, which we call decouple-recouple synthesis (DRS), to a typical investment decision problem, whereby a representative investor faces the choice of investing in a variety of industry-specific portfolios based on signals extracted from a large set of financial ratios grouped into seven different categories. Surprisingly enough, the use of financial ratios in time series returns predictability has been mostly relegated to few measures of stock prices relative to fundamentals, such as the dividend yield, the book-to-market ratio, and the earnings-to-price ratio. The main reason is that because each of these ratios has price in the denominator, the ratio should be related to expected returns, either through a mispricing channel, i.e., ratios are low when stocks are overpriced, or through a rational-pricing mechanism, whereby the ratios track the time variation in discount rates (see, e.g., Cochrane, 2011). However, existing research show that, for instance, profitability (see Fama and French, 2006 and Ball et al., 2016), liquidity (see Pástor and Stambaugh, 2003), leverage (see Gomes and Schmid, 2010), and R&D investments (see Li, 2011) – all indices that, to a large extent, can be classified as accounting ratios – turn out to be key determinants of expected future returns on stocks.

Our focus on industry-sorted portfolios is motivated by keen interests from researchers (see, e.g., Fama and French, 1997) and practitioners (for example, the increasing popularity of industry ETFs). While there is a vast literature examining the out-of-sample predictability of U.S. aggregate returns (see, e.g., Elliott and Timmermann, 2004, Timmermann, 2004, Goyal and Welch, 2008, Rapach et al., 2010, Dangl and Halling, 2012, Johannes et al., 2014, Pettenuzzo et al., 2014, and Rossi, 2018, among others), the question of whether industry-specific returns are predictable, out of sample, has received little attention so far. Yet, the implications of industry returns predictability are far from trivial. If all industries are unpredictable, then the market return should also be unpredictable; the evidence of aggregate market return predictability, thus, implies that at least some market component returns are predictable. Furthermore, industry return predictability could have important implications for dynamic asset pricing models, since they ultimately affect the efficient allocation of capital across sectors (e.g., Stambaugh, 1983, Campbell, 1987, Huang, 1987, Kirby, 1998, and, more recently, Lewellen et al., 2010).

Studies on industry return predictability, however, are few. Early exceptions are Ferson and Harvey (1991), Ferson and Korajczyk (1995) and Ferson and Harvey (1999), which use a set of industry portfolio as test assets to look at the in-sample explanatory power of macroeconomic risk

interchangeably.

factors. Using a Bayesian approach, Avramov (2004) explores the predictive content of standard Fama-French risk factors for a handful of industry portfolios and investigate the implications for asset allocation decisions. More recently, Cohen and Frazzini (2008), Menzly and Ozbas (2010) and Rapach et al. (2015) investigate the in-sample cross-industry return behavior and returns predictability. Relative to these studies, which largely focus on in-sample predictability using a small number of predictors, we analyze the out-of-sample predictability of all industry components based on a large set of financial ratios. In this respect, to the best of our knowledge, this paper represents the first comprehensive study on industry returns predictability within a large-scale regression setting.

In terms of alternative forecasting frameworks, we compare our model against a set of mainstream forecast combination techniques, such as standard Bayesian model averaging (BMA), in which the forecast densities are combined with respect to recursively updated posterior model probabilities (see, e.g., Harrison and Stevens, 1976, Sect 12.2 West and Harrison, 1997 and Pettenuzzo and Ravazzolo, 2016), as well as against simpler, equal-weighted averages of the model-specific forecast densities using linear pools, i.e., arithmetic means of forecast densities, with some theoretical and practical underpinnings (see, e.g., West 1984 and Diebold and Shin, 2017). In addition to the equal-weight forecast combination scheme, we follow Rapach et al. (2010) and also consider the median sub-regression forecast and the trimmed equal-weight mean forecast, which sets to zero the weight for the smallest and largest individual forecasts and assume equal weights for the remaining sub-regressions. While some of these strategies might seem overly simplistic, they have been shown to dominate more complex aggregation strategies in many contexts, and are considered benchmarks in the literature (Genre et al., 2013).

Together with standard forecast combination strategies, we compare the forecasts from our DRS strategy with a set of linear penalised regressions, such as lasso (see, e.g., Tibshirani, 1996), ridge (see, e.g., Hsiang, 1975), and elastic net (see, e.g., Zou and Hastie, 2005), as well as PCA based latent factor modeling (see, e.g., Stock and Watson, 2002 and McCracken and Ng, 2016). Finally, we also consider two conventional non-linear machine learning predictive strategies, such as random forest and a shallow long short-term memory neural network (see, e.g., Rossi and Timmermann, 2015; Rossi, 2018 and Gu et al., 2020).

As suggested by Campbell and Thompson (2007), we benchmark each of these strategies against the simple historical average (HA) of the returns. We employ forecast encompassing tests to elucidate the econometric sources of the benefits of our decouple-recouple predictive strategy. These tests are both based on the conditional mean forecasts, i.e., the expected value of future returns, as well as on conditional density forecasts, for given time indices $t$. Both tests produce evidence of statistically significant information in financial ratios in forecasting stock excess returns, both at the industry and at the aggregate market level. Consistent with existing evidence on the aggregate stock market (see, e.g., Rapach et al., 2010), we show that, also for industry portfolios, returns tend to be more predictable during recessions than expansionary periods.

With regard to the out-of-sample economic performance, we test our predictive system on two main portfolio optimization settings. First, we implement a mean-variance portfolio choice, where a representative investor seeks to maximise her risk-return trade-off. Second, we test each predictive strategy within the context of a power-utility representative investor. We follow DeMiguel et al. (2014) and consider 10 basis points of linear transaction costs when calculating the returns on each predictive strategy. The empirical results show that an investor, using our DRS predictions to construct trading signals, consistently outperforms the competition. In practice, this translates to an investor that is willing to pay a significant fee to have access to the predictions of the DRS, compared to both the competing predictive models and the simple historical mean. Such outperformance is confirmed by assuming both unconstrained and short-sales constrained investments.

The conceptual innovation of our decouple/recouple estimation strategies has been nucleated by Zhao et al. (2016), though in a very different context. Since then, the utility to allow for the sequential estimates to be decoupled – enabling fast, more efficient processing – and then ensembled for decision making has been outlined in the context of Cholesky-type volatility modeling (see, e.g., Lopes et al., 2016 and Shirota et al., 2017), as well as for dynamic latent factor models (see, e.g., Irie et al., 2019). Our methodology differs from existing model combination schemes by utilizing the theoretical foundations and recent developments in dynamic forecasts with multiple forecasts/models (see, e.g., McAlinn and West, 2019). Under this framework, the inter-dependencies between the group-specific predictive densities, as well as biases within each group, are effectively treated as separate latent states that can be sequentially learned, updated, and corrected; information that is critical for decision making, though lost in typical model combination techniques. Along this line, Clemen (1989), Makridakis (1989), Diebold and Lopez (1996), and Stock and Watson (2004) have pointed out that individual forecasting models are likely to be subject to mis-specification bias of unknown form. Even in a stationary world, the data generating process is likely to be far more complex than assumed by the best forecasting model and it is unlikely that the same set of regressors dominates all others at all points in time. As a result, our forecasting approach can be viewed as a way to robustify the aggregate prediction against model mis-specification and measurement errors affecting each individual forecasts.

Delving further into the economic mechanism underlying our predictive model, we provide evidence of three main findings of interest. First, we show that there is substantial time series variation in the dynamics of the correlation amongst forecasts from different groups of financial ratios, with visual differences pre- and post-financial crisis. This suggest that the "robustifying" benefits of aggregating individual forecasts correlates with the aggregate macroeconomic conditions. Second, we show that the marginal effects of different groups of financial ratios in forecasting industry stock returns change both in the time series and across different industries. When averaged across industries, such marginal effects align with the value-weighted market portfolio. From an asset pricing perspective, this suggests that different factors are priced differently across industries, while, almost by construction, the aggregate market portfolio represents a value-weighted average of the industry-specific pricing kernels (see, e.g., Fama and French, 1997 and Huang et al., 2015). Third, we show that the aggregate forecasting bias and the uncertainty around the expected future stock returns tend to move in opposite directions. That is, while the predictive content of financial ratios tend to increase during recessions, the associated forecast uncertainty also increases. These findings lend empirical support for our time-varying forecast combination strategy and its implications for industry returns predictability.

# 2 Modeling Expected Returns

A canonical and relevant approach to investigate if and which characteristic and/or risk factor help to predict risk premia is to consider a basic linear regression;

$$y_t = \alpha + \sum_{i=1}^{n} \beta_i C_{i,t-1} + \epsilon_t, \tag{1}$$

where $y_t$ is the return in excess of the risk-free rate at time $t$, $C_{i,t-1}$, for $i = 1, \ldots, n$ is a given predictor in the previous period, $\beta_i$ is the corresponding slope coefficient, $\alpha$ is the intercept (or bias term), and $\epsilon_t$ is some observation noise. A linear regression of the form in Eq. (1) still represents a benchmark approach to capture the time series variation of stock excess returns (see, e.g., Campbell and Thompson, 2007, Goyal and Welch, 2008, and Rapach et al., 2010).

Although simple to implement, a standard multiple linear predictive regression comes with many pitfalls. In many practically important investment decisions, the information set relevant to make an informed choice is large, possibly too large, to directly fit something as simple as an ordinary linear regression. In almost all contexts, at least *a priori*, all of the available predictors could provide relevant information. However, the out-of-sample performance of standard estimation techniques for linear regressions, such as ordinary least squares, maximum likelihood, or Bayesian inference with uninformative priors, tend to deteriorate as the dimension of the data increases, which is the well known curse of dimensionality. Cochrane (2011) describes and elaborates many of the challenges that multiple linear regressions face in the context of data-rich environments, and suggests "we will have to use different methods".

## 2.1 Current conventional approaches

To address the curse of dimensionality in standard predictive models of the form in Eq. (1), a variety of alternative methods have been proposed in the literature. Broadly speaking, four classes of methods emerged. The first, labelled *penalised regressions*, focuses on the selection/shrinkage 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. The second class of models falls under the heading of *model combination* techniques. The central idea in the context of stock returns predictability has been outlined in Timmermann (2004) and Rapach et al. (2010). That is, even if the best set of predictors can be identified at each point in time, combination of different predictions (from different subset of variables) for the same variable of interest may still be attractive. The third class of models, dubbed *data compression* techniques, 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 predictors are captured by a smaller set of common latent components, which could be either static or dynamic. Departing from linearity, a fourth class of models relates to the idea that there may be an unknown, possibly sparse, non-linear mapping between returns and predictors. Examples used in empirical finance are random forest regressions and neural networks (see, e.g., Rossi, 2018, Gu et al., 2020, and Bianchi et al., 2020). In the following, we briefly review the four classes of methods. We leave to Appendix C a more detailed explanation of each method and estimation strategy.

**2.1.1 Penalised regressions.** The central idea of penalised regressions is to add a penalty term to the minimisation of an otherwise standard loss function, with the explicit goal of shrinking small coefficients towards zero, while leaving large coefficients mostly intact, i.e.,

$$\mathcal{L}\left(\boldsymbol{\beta};\cdot\right) = \underbrace{\mathcal{L}\left(\boldsymbol{\beta}\right)}_{\text{Loss Function}} + \underbrace{\phi\left(\boldsymbol{\beta};\cdot\right)}_{\text{Penalty Term}}, \tag{2}$$

where the loss function, $\mathcal{L}\left(\boldsymbol{\beta}\right)$, is commonly represented by the sum of squared residuals. Depending on the functional form of the penalty term $\phi\left(\boldsymbol{\beta};\cdot\right)$, the regression coefficients can be regularised and shrunk towards zero, completely set to be zero, or a combination of the two. The penalised minimisation problem in Eq. (2) can be rewritten as a Bayesian hierarchical model by using a specific shrinkage prior. The main advantage of adopting a Bayesian approach is that the penalisation parameters are treated as random variables with their own prior and posterior distributions. This allows to make inference on the key shrinkage parameters in a transparent fashion, without relying on often hard-to-implement cross-validation techniques. For instance, a ridge regression can be simply cast as a Normal prior on the regression coefficients of the form, $\beta_i|\tau^2, \sigma^2 \sim N\left(0, \tau^2\sigma^2\right)$, where $\tau^2$ is the global shrinkage parameter (see, Hsiang, 1975). In addition to the ridge, we implement both a sparse regression method in the form of a Bayesian lasso as well as a hybrid shrinkage prior in the form of an elastic net (see, Li et al., 2010). The latter mitigates the lasso concerns related to $n > T$ and the presence of group-wise correlated regressors (see, Zou and Hastie, 2005).

**2.1.2 Model averaging.** From a purely theoretical perspective, unless a particular forecasting model can be identified *ex-ante*, forecast combination offers diversification gains that make it attractive (see, e.g., Bates and Granger, 1969 and Clemen, 1989). In its simplest form, the individual forecast from the $jth$ group of financial ratios, i.e., $p(y_t|\boldsymbol{C}^j_{t-1})$, can be aggregated via linear pooling (see, e.g., Geweke and Amisano, 2011):

$$p\left(y_t|\boldsymbol{C}_{t-1}\right) = \sum_{j=1}^{J} w_j p\left(y_t|\boldsymbol{C}_{t-1}^{j}\right), \tag{3}$$

with $\boldsymbol{C}_{t-1}^{j} = \left(C_{1t-1}^{j}, \dots, C_{n_j t-1}^{j}\right)$ represents a set of $n_j$ predictors in the group $j = 1, \dots, J$ of financial ratios and $w_{1:J}$ is the set of model weights.

Among others, two main reasons make forecast combination particularly appealing in the context of financial forecasting. First, individual predictive regressions may be subject to misspecification bias of unknown form, since the data generating process of the returns is unobservable, and any individual forecasting model merely represents a local approximation, which is unlikely to dominate consistently over time (see, e.g., Diebold and Lopez, 1996). Second, individual models can be vastly, differently affected by non-stationarities, such as structural breaks and parameter changes. Both aspects are particularly relevant within the context of financial returns, which are plagued by both low signal-to-noise ratios, as well as time-varying exposure to proxies of systematic risk (see, e.g., Dangl and Halling, 2012).

We follow existing research and use established benchmarks for the model weights. The first approach involves a simple equal-weighting scheme, whereby we simply assume $w_j = 1/J$, for $J$ number of clusters of predictors. Despite being theoretically suboptimal, the equal weighting scheme has been shown to generate substantial outperformance with respect to optimal weights based on log-score and/or in-sample calibration (see, e.g., Smith and Wallis, 2009, and Diebold and Shin, 2017). The second forecast combination scheme is based on a large and increasing literature in time series forecasting, within the context of Bayesian methods: Bayesian Model Averaging (BMA; Harrison and Stevens, 1976; West and Harrison, 1997; Avramov, 2002; Johannes et al., 2014), where the weight, $w_j$, is chosen based on the marginal predictive density score.[4] The third and fourth forecast combination methods are borrowed from Rapach et al. (2010). In particular, we give weight one to the median forecast at each time $t$ or implement a trimmed equal-weight average scheme, whereby the largest and smallest forecasts are discarded and the remaining forecasts are given a weight equal to $w_j = 1/(J-2)$, for $j = 1, \dots, J$.

#### 2.1.3 Data compression methods.

A third class of models used to address the proliferation of predictors is based on the assumption that few latent components summarise most of the time series variation in the data. The approach of treating conditioning information to characterise expected

[4] In our context, we estimate a separate predictive regression for each group of financial ratios and estimate its weight in the aggregate forecast as

$$w_j \propto \frac{p(y_t|\boldsymbol{C}_{t-1}^{j})}{\sum_{j=1}^{J} p(y_t|\boldsymbol{C}_{t-1}^{j})}, \qquad j = 1, \dots, J. \tag{4}$$

with the weights restricted to be inside the unit circle and their sum restricted to one. The restrictions on the weights, $w_j$, are necessary and sufficient to assure that $p\left(y_t|\boldsymbol{C}_{t-1}^{j}\right)$ is a density function for all values of the weights and all arguments of the group-specific predictive regressions (see, e.g., Geweke and Amisano, 2011).

returns as latent and use factor analytic techniques, such as principal components analysis (PCA), has been pioneered by Chamberlain and Rothschild (1983) and Connor and Korajczyk (1986), and became highly popular, especially in data-rich macroeconomic contexts (see, e.g., Stock and Watson, 2002, 2004; De Mol et al., 2008; Manzan, 2015, and Stevanovic, 2017, and the references therein). Information from a large panel of predictors is condensed into few statistical factors, which are extracted from no previous knowledge of the dynamics of stock returns. Often, the optimal number of factors is selected at each time $t$ using some information criterion (see, e.g., Bai and Ng, 2002).[5] Given all other methods are static in nature, that is each predictive regression does not have time-varying betas, we consider a recursively estimated static factor model.[6]

**2.1.4 Non-linear machine learning methods.** In the main empirical analysis, we restrict, on purpose, our focus to standard linear predictive frameworks. The reason is three-fold: first, we want to make our results comparable, as much as possible, with linear predictive systems in small and large data settings (see, e.g., Barberis, 2000; Lewellen, 2004; Pastor and Stambaugh, 2009; Johannes et al., 2014; Huang et al., 2015). Second, from a methodological perspective, our main aim is to show the advantages of our approach towards the bias-variance trade-off relative to classical approaches, such as principal component and shrinkage regressions, that have been commonly used to deal with data-rich environments (see, e.g., Rapach et al., 2015; Feng et al., 2019). Third, we believe that focusing on linear models allow to isolate, in a transparent way, the underlying mechanism of the predictive strategy compared to non-linear machine learning methods.

Nevertheless, we consider it to be instructive to compare our DRS predictive strategy to several widely used and popular non-linear machine learning methods, namely regression trees and neural networks (see, e.g., Rossi and Timmermann, 2015; Rossi, 2018; Gu et al., 2020; Bianchi et al., 2020, among others). In particular, we consider two non-linear forecasting strategies; random forests and recurrent neural networks. Appendix C gives a detailed description of both methods and the algorithmic procedures implemented.

## 2.2 Our "divide and conquer" approach

Our methodology exploits an existing intuition in multivariate time series analysis, whereby a large set of multivariate outcomes is broken up into smaller-dimensional univariate and multivariate models, and the full multivariate structure (i.e., the covariance matrix) is recovered via a post-process (see, e.g., Gruber and West, 2016 and the recent developments in Gruber and West, 2017;

[5]More precisely, a factor model is built on the idea that $y_t$ relates to an underlying vector of $q < n$ latent variables $\boldsymbol{f}_t = (f_{1t}, \ldots, f_{qt})$ extracted from the set of $n$ characteristics in $\boldsymbol{C}_t$, via

$$y_t = \alpha + \sum_{j=1}^{q} \beta_j f_{jt-1} + \epsilon_t.$$

[6]To explicit a given dynamics, one could model factor loadings as a linear function of exogenous variables (see the IPCA method proposed by Kelly et al., 2019), or assume a stochastic dynamic for the factor loadings (see, e.g., Bernanke et al., 2005; Stock and Watson, 2005)

Chen et al., 2017). Our idea applies and extends this intuition to problems with large predictors: a potentially large $n-$dimensional vector of predictors can be partitioned into smaller – possibly inter-dependent – clusters, $j = 1{:}J$, modifying Eq. (1) to

$$y_t = \alpha + \sum_{i=1}^{n_1} \beta_i^1 C_{it-1}^1 + \ldots + \sum_{i=1}^{n_j} \beta_i^j C_{it-1}^j + \ldots + \sum_{i=1}^{n_J} \beta_i^J C_{it-1}^J + \epsilon_t, \tag{5}$$

where $n = \sum_{j=1}^{J} n_j$ is the total number of predictors, $C_{it-1}^j$ is the $ith$ predictor within the $jth$ group, and $\beta_i^j$ is the corresponding slope coefficient.[7]

Our predictive regression approach works in two intuitive steps. The *first step* is to *decouple* Eq. (5) into $J$ smaller predictive models with the same target variable $y_t$, such as,

$$y_t = \sum_{i=1}^{n_j} \beta_i^j C_{it-1}^j + \epsilon_{tj}, \tag{6}$$

for all $j = 1{:}J$, producing forecast distributions $p(y_t|\boldsymbol{C}_{t-1}^j)$, where $\boldsymbol{C}_{t-1}^j$ denotes each group of characteristics. Since Eq. (6) is a linear projection of data from each group of financial ratios, we can consider, without loss of generality, that $p(y_t|\boldsymbol{C}_{t-1}^j)$ is reflecting the group-specific information regarding the future behavior of excess returns on a given industry portfolio.

In the *second step*, we *recouple* the densities $p(y_t|\boldsymbol{C}_{t-1}^j)$ for $j = 1{:}J$ in order to obtain a unified/synthesized forecast distribution $p(y_t|\boldsymbol{C}_{t-1})$, reflecting and incorporating all of the information that arises from each group of predictors. We describe the features of each of these two steps. A detailed description of the estimation procedure is provided in Appendix B.

**2.2.1 Group-specific predictive regressions.** In order to be able to disentangle the pure effect of our decouple-recouple modeling framework, we consider a simple setting for the individual, group-specific, predictive regressions, $p(y_t|\boldsymbol{C}_{t-1}^j)$, for $j = 1, \ldots, J$;

$$y_t = \boldsymbol{\beta}_j' \boldsymbol{C}_{t-1}^j + \epsilon_{tj}, \quad \epsilon_{tj} \sim N(0, \nu_j), \tag{7}$$

where the coefficients are assumed to have a conjugate Normal-inverse Gamma prior distribution, such that $\boldsymbol{\beta}_j|v_j \sim N(\boldsymbol{m}_j, (v_j/s_j)\boldsymbol{I})$ and $1/v_j \sim G(n_j/2, n_j s_j/2)$. Note that our model would work exactly the same if the individual predictive regressions are assumed to have time-varying betas with, for instance, a random-walk type of dynamics. While such time-varying specification is appealing, since it reflects the dynamic predictive content of financial ratios,[8] this would make the

[7]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.

[8]See, e.g., Jostova and Philipov (2005), Nardari and Scruggs (2007), Adrian and Franzoni (2009), Pastor and Stambaugh (2009), Binsbergen et al. (2010), Dangl and Halling (2012), Pastor and Stambaugh (2012), and Bianchi et al. (2017), among others.

comparison with penalised regressions and machine learning methods unfair, as they assume betas are constant in the data generating process. As a result, we consider a static predictive regression for each cluster of financial ratios, which is more consistent with the underlying assumption of other competing methods. In addition, this gives a more direct comparison with some of the existing research on the predicting content of financial ratios and the predictability of industry stock returns, more generally (see, e.g., Lewellen, 2004, Rapach et al., 2010, Rapach et al., 2015 among others).

**2.2.2 Predictive synthesis.** Arguments against using forecast combination strategies are as numerous as the arguments in favour. For instance, whereas the instability in the underlying data generating process can be an argument in favour of forecast combination, it can also lead to instabilities in the weights and cause great difficulty in deriving a set of combination weights that perform well (see, e.g., Clemen and Winkler, 1999, 2007), which, again, is why equal-weighting is considered a hard benchmark to beat. In addition, it may be reasonable to assume that information sets used in the primitive forecasting regressions may overlap, leading to a dependence between $p(y_t|\boldsymbol{C}_{t-1}^j)$ and $p(y_t|\boldsymbol{C}_{t-1}^q)$, for $j \neq q$. This, in turn, could affect the estimates of the combination weights that are performance based and often do not explicitly take into account such dependence (see, e.g., Eq. 4). Finally, the individual forecasts, $p\left(y_t|\boldsymbol{C}_{t-1}^j\right)$, are possibly affected by omitted variable biases, being built on a smaller set of predictors.

To address these issues, efficient forecast combination weights should be chosen to minimize the expected loss of the *combined* forecast, which, by definition, reflects both the forecasting accuracy of each sub-model and the correlation across forecasts.[9] This is a key feature of our econometric strategy. In particular, the prediction from each group of predictors in the first step is considered to be a latent state, such that $x_{tj} \sim p(y_t|\boldsymbol{C}_{t-1}^j)$ represents a distinct prior on the individual predictive models. That is, each $x_{tj}$ represents a random sample from the predictive content of a group of financial ratios regarding the expected excess returns of the given industry portfolio; the collection of which defines the information set $\mathcal{X}_t = \{p(y_t|\boldsymbol{C}_{t-1}^1), \ldots, p(y_t|\boldsymbol{C}_{t-1}^J)\}$. These latent states are then calibrated and learned using standard prior-posterior updating rules. More precisely, for a given prior, $p(y_t)$, and (prior) information set, $\mathcal{X}_t$, we can update using the Bayes' theorem to obtain a posterior $p(y_t|\mathcal{X}_t)$.

Due to the complexity of $\mathcal{X}_t$ – a set of $J$ predictions with cross-sectional time-varying dependencies, as well as individual biases – the synthesised prediction is often difficult to define. We build on West and Crosse (1992) and West (1992) (which extend the basic theorem of Genest and Schervish, 1985), and the recent developments by McAlinn and West (2019) that show that, under

[9]For instance, it is evident that the marginal predictive power of, for example, valuation ratios may be correlated with variables that are directly related to solvency. In addition, correlations across predictive densities are arguably latent and dynamic. For instance, the spillover effects between solvency, liquidity, and aggregate financial variables possibly changed before and after the great financial crisis of 2008/2009.

a specific consistency condition, the time-varying posterior density takes the form,

$$p(y_t|\mathbf{\Phi}_t, \mathcal{X}_t) = \int \pi(y_t|\boldsymbol{x}_t, \mathbf{\Phi}_t) \prod_{j=1:J} p(y_t|\boldsymbol{C}_{t-1}^j) dx_{tj}, \tag{8}$$

where $\boldsymbol{x}_t = (x_{t1}, \ldots, x_{tJ})$ can be thought of as a $J-$dimensional latent state vector at time $t$, $\pi(y_t|\boldsymbol{x}_t, \mathbf{\Phi}_t)$ is a conditional *synthesis* function, which reflects how the latent states, $\boldsymbol{x}_t$, are to be synthesised, and $\mathbf{\Phi}_t$ represents some time-varying parameters.

Although the theory does not specify $\pi(y_t|\boldsymbol{x}_t, \mathbf{\Phi}_t)$,[10] a natural choice is to impose linear dynamics, such that $\pi(y_t|\boldsymbol{x}_t, \mathbf{\Phi}_t) = N(y_t|\boldsymbol{F}_t'\boldsymbol{\theta}_t, \sigma_t^2)$, where $\boldsymbol{F}_t = (1, \boldsymbol{x}_t')'$ and $\boldsymbol{\theta}_t = (\theta_{0t}, \theta_{t1}, ..., \theta_{tJ})'$ represents a $(J+1)-$vector of time-varying synthesis coefficients (see, e.g., McAlinn and West, 2019; McAlinn et al., 2020). More precisely, we assume that both $\boldsymbol{\theta}_t$ and $\sigma_t^2$ evolve as a random walk to allow for stochastic changes over time and recast Eq. (5) as a Dynamic Linear Model (DLM) of the form (see West and Harrison 1997; Dangl and Halling 2012),

$$y_t = \boldsymbol{F}_t'\boldsymbol{\theta}_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_t^2), \tag{9a}$$

$$\boldsymbol{\theta}_t = \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(0, \sigma_t^2 \boldsymbol{W}_t), \tag{9b}$$

such that the time-varying latent parameters are defined as $\mathbf{\Phi}_t = \left(\boldsymbol{\theta}_t, \sigma_t^2\right)$. Few comments are in order. First, by representing forecast combination as a state-space model, the "observation noise", $\epsilon_t$, can be directly interpreted as the measurement error. As a result, the latent parameter $\sigma_t^2$ increases (decreases) when the synthesised forecast is less (more) uncertain; thus providing a direct check on the quality of the forecast combination scheme, something that existing methods do not provide. Second, the time-varying correlations of individual forecasts are explicitly controlled for, through the dynamics of the state equation Eq. (9b); take, for instance, $Cov_t\left(\theta_{it}, \theta_{jt}\right) = \sigma_t^2 Cov_t\left(\omega_{it}, \omega_{jt}\right) = \sigma_t^2 W_{ij,t}$, with $W_{ij,t}$ the $ij$-th element of the state covariance matrix $\boldsymbol{W}_t$. Third, McAlinn and West (2019); McAlinn et al. (2020) show that the aggregate intercept, $\theta_0$, can be interpreted as the "bias" in the predictive system, Eqs. (9a)-(9b). In this respect, and given the nature of the forecasting problem, $\theta_0$ can be interpreted as the fraction of expected excess returns on industry portfolios that is not explained by financial ratios and additional predictors, i.e., the omitted variable bias of the entire set. Fourth, despite the fact that the individual predictive regressions are constant within windows of observations, the time-varying synthesis in Eq. (9a)-(9b) implies that our method, similar to model averaging techniques, can respond to non-stationarities, such as structural breaks caused, for example, by changes in the fundamental relationship between financial ratios and stock returns (see, e.g., Timmermann, 2004 for a related discussion on forecast combination).

[10] Note that McAlinn and West (2019) show that many forecast combination methods, from linear combinations (including equal-weighting and BMA) to more advanced density pooling methods (e.g. Aastveit et al., 2014; Kapetanios et al., 2015; Pettenuzzo and Ravazzolo, 2016), are special cases of Eq. (8).

**2.2.3 Estimation strategy.** Estimation for the decouple step is straightforward and depends on the model assumptions for each group-specific model. For instance, for a typical linear regression, we can compute draws from each of the forecasting densities, $x_{tj} \sim p(y_t|\boldsymbol{C}_{t-1}^j)$, using standard conjugate Bayesian updating.[11]

As for the recoupling step, some discussion is needed. In particular, the joint posterior distribution of the latent states and the structural parameters is not available in closed form. We implement a Markov Chain Monte Carlo (MCMC) approach using an efficient Gibbs sampling scheme. In our framework, the latent states are represented by the predictions from the models, $x_{tj}$, for $j = 1, ..., J$, and the synthesis parameters, $\boldsymbol{\Phi}_t = \left(\boldsymbol{\theta}', \sigma_t^2\right)$. In a nutshell, our MCMC algorithm involves a sequence of standard steps in a customized two-component block Gibbs sampler; the first component simulates from the conditional posterior distribution of the latent states given the synthesis parameters and the data, and the second component simulates the synthesis parameters given the latent states and the data.

The residual $\epsilon_t$ and the evolution *innovation* $\boldsymbol{\omega}_s$ are independent over time and mutually independent for all $t, s$. The dynamics of $\boldsymbol{W}_t$ is imposed by a standard, single discount factor specification, as in West and Harrison (1997) (Ch.6.3) and Prado and West (2010) (Ch.4.3). The residual variance, $\sigma_t^2$, follows a beta-gamma random-walk volatility model, such that $\sigma_t^2 = \sigma_{t-1}^2 \delta/\gamma_t$, where $\delta \in (0, 1]$ is a discount parameter, and $\gamma_t \sim \text{Beta}\left(\delta n_t/2, (1-\delta)\, n_t/2\right)$ are innovations independent over time and independent of $\sigma_s, \boldsymbol{\omega}_r$ for all $t, s, r$, with $n_t = \delta n_{t-1} + 1$, the degrees of freedom.

The discount factors for the conditional volatilities in Eqs. (9a)-(9b) are set to $\delta = (0.95, 0.99)$. Results are qualitatively the same using different specifications. Priors for each decoupled predictive regression are assumed to be fairly uninformative, such as $\boldsymbol{\beta}_j|v_j \sim N(\boldsymbol{m}_j, (v_j/s_j)\boldsymbol{I})$ with $\boldsymbol{m}_j = \boldsymbol{0}'$ and $1/v_j \sim G(n_j/2, n_j s_j/2)$ with $n_j = 10, s_0 = 0.01$. Similarly, for the recouple step, we use the following uninformative marginal priors: $\boldsymbol{\theta}_0|v_0 \sim N(\boldsymbol{m}_0, (v_0/s_0)\boldsymbol{I})$ with $\boldsymbol{m}_0 = (0, \boldsymbol{1}'/J)'$ and $1/v_0 \sim G(n_0/2, n_0 s_0/2)$ with $n_0 = 10, s_0 = 0.01$. A more detailed description of the algorithm and how forecasts are generated can be found in Appendix B.

# 3 Empirical Findings

The empirical analysis for DRS is conducted as follows. First, the individual predictions (cf., Section 2.2.1) are analysed in parallel over 1970:01-2000:01 as a training period, simply estimating the regression in Eq. (7) to the end of that period to estimate the forecasts from each subgroup. This continues over 2000:01-2018:12, but with the calibration of the recouple strategy, which, at each quarter $t$ during this period, is run with the MCMC-based DRS analysis using data from 2000:02 up to time $t$. We discard the forecast results from 2000:02-2004:01 as training data and compare predictive performance from 2004:02-2018:12. Similarly, for the other methods, the forecasts are

[11] Notice that $x_{tj} \sim p(y_t|\boldsymbol{C}_{t-1}^j)$ are forecasting *densities* and not merely *conditional means*, as we work under a standard prior-posterior Bayesian updating scheme.

recursively calculated starting from 2004:01 using information from 1970:01. That means that each predictive system uses the same in-sample information to generate the one-step ahead forecast at any given time $t$, and is tested across the same out-of-sample period.[12]

Importantly, we assume the $J$ groups of predictors are determined *ex-ante* based on their economic meaning. In this respect, we let the economics guide the construction of the modeling framework, instead of using a more data-driven approach based on the classification explained below (see, e.g., Bianchi et al., 2020 for an application of a similar concept on neural networks). Alternative algorithmic classification procedures, such as the $k$-nearest neighbors algorithm, can be used to group variables based on their correlation structure. Yet, exactly the same methodology would apply, since neither the model specification, nor the estimation, is dependent on the method used to cluster the financial ratios.

### 3.1 Data

To approximate the largest possible set of financial ratios, we use the WRDS Financial Ratio (WFR, henceforth) database. Such data collects, in total, over 70 financial ratios grouped into seven categories: Valuation, Profitability, Capitalization, Financial soundness, Solvency, Liquidity, Efficiency, and Other ratios.

Table A.1 in Appendix A provides a full description of the variables, how they are constructed, as well as their group classification. Capitalization measures the debt component of a firm's capital structure; Efficiency measures capture the effectiveness of a firm's usage of assets and liabilities; Financial soundness and Solvency measures capture a firm's ability to meet long-term obligations, whereas Liquidity measures focus more on short-term obligations; Profitability measures capture the ability of a firm to generate profit, whereas Valuation measures estimate the attractiveness of a firm's stock. Finally, Other measures contain miscellaneous variables, such as R&D-related ratios.

In order to mitigate the effect of small- and micro-cap stocks, the aggregation of financial ratios at the industry level is constructed by taking the value-weighted mean of firm-specific values.[13] Industry aggregation for both portfolio returns and financial ratios 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. All original accounting variables are obtained from Compustat Quarterly and Annual file, whereas pricing related data are obtained from CRSP and Compustat. Earnings-related variables are obtained from IBES. We also consider a market-wide aggregation, where we consider a value-weighted aggregation of the financial ratios and returns. The sample size is from 1970:01 to 2018:12, monthly. The length of the time series is limited by the data provider.

In addition to financial ratios, we also consider a ninth group of variables that we labeled as

[12] The time frame of the testing period includes key events, 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 was led by the burst of the sub-prime mortgage crisis.

[13] We winsorize values at the 1% and 99% percentiles to mitigate the effect of outliers.

Macro, which contains five additional aggregate predictors that are taken from existing studies, such as Goyal and Welch (2008), Rapach et al. (2010), and Dangl and Halling (2012). In particular, we consider 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 capitalization (ntis), the default yield spread (dfy) calculated as the difference between BAA and AAA-rated corporate bond yields, the term spread (tms) calculated as the difference between the long term yield on government bonds and the T-bill, and the growth rate of inflation (infl).

## 3.2 Forecasting performance

We compare the forecasts obtained from each methodology to a naïve prediction based on the historical mean of excess stock returns. In particular, we calculate the out-of-sample predictive $R^2$, as suggested by Campbell and Thompson (2007). The $R^2_{oos}$ is akin to the in-sample $R^2$ and is calculated as

$$R^2_{oos} = 1 - \frac{\sum_{t_0=2}^{T} \left(y_t - \widehat{y}_t\left(\mathcal{M}_s\right)\right)^2}{\sum_{t_0=2}^{T} \left(y_t - \overline{y}_t\right)^2}, \tag{10}$$

where $\overline{y}_t$ is the historical mean and $\widehat{y}_t\left(\mathcal{M}_s\right)$ is the expected forecast from model $\mathcal{M}_s$, for each industry, and $t_0$ is the date of the first prediction. Forecast errors are obtained by comparing the excess industry portfolio returns during the 2004:02 - 2018:12 period.

We also build a portfolio-level return forecast from the individual industry forecasts produced by our models. We construct forecasts of an equally-weighted portfolio, $\widehat{y}_t^{(EW)} = \frac{1}{N}\sum_{n=1}^{N} \widehat{y}_t^{(n)}\left(\mathcal{M}_s\right)$, where $\widehat{y}_t^{(n)}\left(\mathcal{M}_s\right)$ is the forecast for a given industry, at a given time $t$, from the model $\mathcal{M}_s$. We compute $R^2_{oos,EW}$ by constructing forecast errors using the realised return $y_t^{(EW)} = \frac{1}{N}\sum_{n=1}^{N} y_t^{(n)}$ and comparing it to the historical cross-sectional mean of sample average returns for each industry, i.e., $\overline{y}_t^{(EW)} = \frac{1}{N}\sum_{n=1}^{N} \overline{y}_t^{(n)}$.

Testing the null hypothesis, $R^2_{oos} \leq 0$, against the alternative hypothesis, $R^2_{oos} > 0$, is tantamount to testing whether the predictive model has a significantly lower mean squared prediction error (MSPE) than the historical average benchmark forecast. Thus, to test whether $R^2_{oos}$ is significant, we implement the MSPE-adjusted Clark and West (2007) statistic.

**3.2.1 Out-of-sample $R^2_{oos}$.** Table 1 reports both the $R^2_{oos}$ and the corresponding p-value for the null hypothesis, $R^2_{oos} \leq 0$, against the alternative hypothesis, $R^2_{oos} > 0$. We report the p-value only if $R^2_{oos} > 0$, which is equivalent to testing whether the predictive model has a significantly lower mean squared prediction error (MSPE) than the historical average benchmark forecast.

With regard to the individual industries, three results emerge. First, both non-linear methods and forecast combination strategies tend to outperform recursive sparse predictive regressions, with lasso and elastic-net type penalties, and the full OLS model, i.e., no penalty, performing the worst. This result echoes the findings in Rossi (2018); Gu et al. (2020) for equity and Bianchi et al.

(2020) for bond returns. Second, forecast combination methods, which consider only the median prediction or a trimmed mean of all models, considerably outperform standard linear pooling, in turn. The performance of more "regularised" model averaging techniques is consistent with Rapach et al. (2010), especially for the aggregate stock market. Third, and more importantly, although with differences across industries, our decouple-recouple strategy is the only one that consistently delivers positive and significant out-of-sample $R^2_{oos}$.[14] When looking at the $R^2_{oos}$ and their statistical significance, the empirical evidence suggests that predicting industry stock returns is somewhat harder than forecasting the aggregate stock market, even using conventional predictive strategies. The results at the industry level are confirmed by looking both at the equal-weight portfolio of all industries. That is, although the median and the trimmed combination slightly outperform the other competing methods, our DRS model delivers the highest performance with regard to mean squared predictive error.

**3.2.2 Cumulative sum of square residuals.** Upon reporting the complete out-of-sample forecasting results, we follow Goyal and Welch (2008) and present time series plots of the differences between the cumulative square prediction error for the historical average benchmark forecast and the cumulative square prediction error for the forecasts based on four of the most representative competing predictive strategies in Figure 1.[15] As pointed out in Rapach et al. (2010), this is an informative graphical device that provides a visual impression of the consistency of an individual predictive regression model's out-of-sample forecasting performance over time.

In particular, we compare our DRS method against some of the best performing competing forecasting schemes: median, equal-weight pooling, BMA, and trimmed average forecast. When the curve Figure 1 is positive, the corresponding predictive regression model outperforms the historical average benchmark, while the opposite holds when the curve is negative.[16]

The left panel of Figure 1 reports the results averaged across industries, whereas the right panel reports the results for the aggregate market.[17] For ease of exposition, the results for all competing strategies and our DRS method are reported on two different scales, on the left and on the right of the graph, respectively.

[14] The only exception are Durables and Hi-tech industries, whereby the positive $R^2_{oos}$ is only marginally statistically significant, i.e., significant below a 10% threshold.

[15] More precisely, the cumulative sum of squared errors (SSE) for the historical average versus the *ith* model is computed as:

$$\Delta\text{CumSSE}_{t,i} = \sum_{\tau=\underline{t}}^{t} (e_{\tau,HA})^2 - \sum_{\tau=\underline{t}}^{t} (e_{\tau,i})^2 ,$$

where $e_{\tau,HA}$ is the prediction error from the historical average forecast and $e_{\tau,i}$ is the prediction error of the competing strategy at time $\tau$.

[16] Essentially, we compare the height of the curve at the two points corresponding to the beginning and end of a given out-of-sample period: if the curve is higher (lower) at the end of the out-of-sample period than at the beginning, the predictive regression model (historical average) has a lower predictive error over the out-of-sample period.

[17] Although there is some heterogeneity across industries, the qualitative conclusion of the results do not change by looking at the cross-sectional average.

Three interesting facts emerge. First, the historical mean forecast turns out to be a challenging benchmark for typical model averaging techniques, such as BMA and equal-weight pooling. Noticeably, although BMA improves with respect to the historical mean at the beginning of the sample, the latter performs worse throughout the great financial crisis of 2008/2009. This result is in line with previous research, such as Campbell and Thompson (2007) and Goyal and Welch (2008). Second, both selecting the median forecast and trimming the average prediction seems to pay off in terms of forecasting accuracy, especially in the aftermath of the great financial crisis of 2008/2009. Third, our dynamic predictive synthesis strategy turns out to consistently outperform the HA benchmark throughout the sample, with a significantly better forecasting performance starting from the great financial crisis to the end of the sample. When looking at the aggregate stock market (right panel of Figure 1) the evidence suggests that forecasting industry portfolios is much harder than predicting the market. As a matter of fact, while standard forecast combination strategies still underperform our DRS approach, their performance compares much more favourably against the recursive sample mean.

**3.2.3 Industry returns predictability over the business cycle.** Recent studies, such as Rapach et al. (2010), Henkel et al. (2011), and Dangl and Halling (2012), report that the predictability of aggregate stock market returns is primarily concentrated in economic recessions, while it is largely absent during economic expansions. Similarly, Sarno et al. (2016), Gargano et al. (2019), and Bianchi et al. (2020) show that, for bond returns as well, there are typically larger gains during times of high macroeconomic uncertainty and/or drop in aggregate economic activity. These findings are important, as they suggest that return predictability may be linked to the cyclical variation in economic fundamentals and/or firm characteristics.

To test if industry returns predictability varies over the business cycle, we split the data into recession and expansionary periods using the NBER dates of peaks and troughs. This information is used *ex-post* and is not used at any time in the estimation of the predictive models.[18] Table 2 shows the $R^2_{oos}$ separately for expansions (Panel A) vs recessions (Panel B). Consistent with the existing evidence on the aggregate stock market returns, the results suggest that the out-of-sample returns predictability is higher in recessions compared to expansions. This is true, not only for our DRS predictive strategy, but also for the forecast combination methods and the random forest. Interestingly, the increasing forecasting performance is stronger at the industry level than at the aggregate market level for the equal-weight portfolio of industries. In fact, while the $R^2_{oos}$ turns from negative to positive for a relatively substantial set of forecast combination techniques at the industry level, for both the market portfolio and the equal-weight aggregation, the same models still underperform the prediction from the recursive sample mean.

In order to test if the differences in the forecasting performance over the business cycle are statistically significant, we implement the bootstrap approach proposed by Gargano et al. (2019).

[18]Note that it is not our goal to predict business cycles. The rationale of this analysis is to investigate if and how each of the predictive models perform differently over the business cycle.

The idea is simply to exploit the monotonic relationship between the out-of-sample mean squared prediction error and the $R^2_{oos}$. For the interested reader, a detailed description of the bootstrap procedure is outlined in Appendix E. The null hypothesis is that the accuracy of a given forecasting model relative to the benchmark is the same across the business cycle, whereas the alternative hypothesis that the relative predictive accuracy is higher in expansions.[19]

The outcomes of the tests are indicated by highlighting in bold the results in the recessions panel of Table 2. In particular, we highlight those $R^2_{oos}$ which are higher, positive, and significant at the conventional 5% significance threshold. With the notable exception of sparse and dense linear predictive models, we find that not only is the fit of the industry returns prediction better in recessions than in expansions, but this difference is highly significant for the vast majority of forecast combination methods. Nevertheless, our DRS consistently outperforms, in terms of $R^2_{oos}$, each of the competing strategies for each of the industry portfolios, as well as for the aggregate stock market and the equal-weight industry aggregation.[20]

**3.2.4 Returns predictions and macroeconomic variables.** To shed more light on the relationship between industry returns predictability and the business cycle, we compare the forecasts from the DRS and the time series of few macroeconomic variables, which are available at the same frequency and have not been used as predictors. For ease of exposition, we compare the forecasts for both the equal-weight portfolio of industries and the value-weighted stock market against the monthly changes in industrial production and employment, as well as the index of economic policy uncertainty proposed by Baker et al. (2016) and the VIX index.[21] We observe some interesting patterns in Figure 2. In the case of both industrial production (top-left panel) and employment (top-right panel) we see that the large drop in economic activity occurred during the great financial crisis coincide with a substantial decrease in forecasted returns. This holds for both the equal-weight industry portfolio and the aggregate stock market. However, the strong positive correlation between forecasts and economic activity tends to deteriorate from early 2010 on, with industrial production diverging from the dynamics of expected industry returns and employment changes largely fluctuating around 0.1% monthly change. In fact, if we calculate the time series correlation between these two series and the model forecasts, we find a positive, but relatively small,

[19] More specifically, the null and alternative hypothesis can be defined as

$$H_0 : E\left[e^2_{HA,0} - e^2_{i,0}\right] = E\left[e^2_{HA,1} - e^2_{i,1}\right] \quad \text{vs} \quad H_1 : E\left[e^2_{HA,0} - e^2_{i,0}\right] < E\left[e^2_{HA,1} - e^2_{i,1}\right],$$

with $e^2_{HA,j}$ and $e_{i,j}$ the squared out-of-sample prediction errors for the simple recursive mean and a given model $i$, respectively. The subscript $j = 0, 1$ refers to the expansion (zero) vs recession (one) period. As highlighted by Gargano et al. (2019), by computing the relative mean squared prediction error with respect to the benchmark historical average forecast we control for the differences in returns variances in recessions vs expansions.

[20] In an unreported set of results, we also consider alternative indicators of expansions vs recessions, such as the unemployment gap recession indicator of Stock and Watson (2010). Except for few nuances, we continue to find that (1) returns predictability tends to be stronger in recessions than in expansions across industries, and (2) our DRS strategy consistently deliver positive $R^2_{oos}$ while a large fraction of competing models do not outperform the forecast from the recursive sample mean.

[21] We focus on the equal-weight aggregation of the industries and the aggregate value-weighted market portfolio for the ease of exposition. The results for the single industries are available upon request.

correlation of 0.3 and 0.5, respectively.

The bottom panels of Figure 2 shows the time series of both the Baker et al. (2016)'s economic policy uncertainty index (bottom-left panel) and the VIX (bottom-right panel). As far as economic policy uncertainty is concerned, there is no obvious correlation with the model-implied forecasts. Apart from a strong negative correlation during the crisis of 2008/2009, there is no signal of comovement to the end of the sample. On the other hand, the VIX strongly negatively correlates with industry and market returns. Not only is such negative correlation evident during the financial crisis, but also during the spikes in early 2010, late 2011, and 2016. This is particularly true regarding the DRS forecasts for the aggregate stock market, whereby the negative correlation with the VIX index takes a remarkable value of -0.75.

The time series comparison of the returns predictions obtained from our DRS strategy to a selection of aggregate macroeconomic variables show reasonable patterns for predictions. This is true both for the equal-weight aggregation of individual industry portfolios, as well as for the value-weighted market portfolio.

### 3.3 Economic value of forecasts

So far, our analysis concentrated on statistical measures of predictive accuracy. However, it is of paramount importance to evaluate the extent to which apparent gains in predictive accuracy translates into better investment performances. Existing evidence of out-of-sample economic gains of model averaging techniques are somewhat mixed (see, e.g., Rapach et al., 2010).

We contribute to this debate by computing the average utility for an investor with relative risk aversion parameter $\gamma$, who allocates his or her portfolio, monthly, between each industry portfolio and risk-free bills using forecasts from each model. Following existing literature (see, e.g., Campbell and Thompson, 2007; Goyal and Welch, 2008; Rapach et al., 2010; Dangl and Halling, 2012 and Pettenuzzo et al., 2014), we calculate realised utility gains for a mean-variance utility investor and a power utility investor, both in a single- and multiple-asset setting.

For the mean-variance utility, at each time $t$, the decision-maker selects the weights on each risky asset to maximize the quadratic utility $\mu_{t,p}\left(\mathcal{M}_s\right) - \frac{\gamma}{2}\sigma^2_{t,p}\left(\mathcal{M}_s\right)$, where $\mu_{p,t}\left(\mathcal{M}_s\right)$ and $\sigma^2_{t,p}\left(\mathcal{M}_s\right)$ are the sample mean and variance, respectively, on the portfolio formed using forecasts of the equity premium based on model $s$. More precisely, in the single-asset case, the investor selects between an industry portfolio and the risk-free return based on the expected return implied by a given model, $w_{t,s} = \frac{\widehat{y}_t(\mathcal{M}_s)}{\gamma\sigma_{t|t-1}}$, where $\widehat{y}_t\left(\mathcal{M}_s\right)$ is the returns forecast for a given industry, given model $\mathcal{M}_s$, and $\sigma_{t|t-1}$, an estimate of time-varying volatility of the returns based on the forecasting residuals. For the case with multiple assets, the vector of optimal portfolio weight, $\boldsymbol{w}_{t,s}$, for the mean-variance investor is expressed as $\boldsymbol{w}_{t,s} = \frac{1}{\gamma}\Sigma^{-1}_{t|t-1}\widehat{\boldsymbol{y}}_t\left(\mathcal{M}_s\right)$, where $\widehat{\boldsymbol{y}}_t\left(\mathcal{M}_s\right)$ is the vector of industry returns forecasts obtained using model $\mathcal{M}_s$, and $\Sigma_{t|t-1} = Var_t\left(\boldsymbol{y}_t - \widehat{\boldsymbol{y}}_t\left(\mathcal{M}_s\right)\right)$.

In addition to a mean-variance utility, we also consider an extended framework, whereby a representative investor has a power utility (Constant Relative Risk Aversion (CRRA) preferences) of the form, $\hat{U}_{t,s} = \hat{W}_t^{1-\gamma}(\mathcal{M}_s)/(1-\gamma)$, and $\hat{W}_t(\mathcal{M}_s)$, the wealth generated by the competing model, $s$, at time, $t$. Campbell and Viceira (2004) show that, under conditionally normal returns, the optimal portfolio allocation for a power utility investor can be expressed as $\boldsymbol{w}_t = \frac{1}{\gamma}\Sigma_{t|t-1}^{-1}\left[\widehat{\boldsymbol{y}}_t + \boldsymbol{\sigma}_{t|t-1}^2/2\right]$, with $\boldsymbol{\sigma}_{t|t-1}^2$ a vector containing the diagonal elements of $\Sigma_{t|t-1}$ (see, Gargano et al., 2019). Given the optimal weights, we compute the realised returns. We follow DeMiguel et al. (2014) and assume a flat 10 basis point transaction cost for each trade.[22] Then, following Fleming et al. (2001), we obtain the certainty equivalent gains (annualized and in percentages) from the realised returns by equating the average utility of the historical average forecast, $u_{t,HA}$, with the average utility of any of the alternative models, $u_{t,s}$. To test whether the certainty equivalent return (CER) values are statistically significant, we use the Diebold and Mariano (1995) test.[23]

This optimal portfolio allocation exercise requires the investor to use some sort of time-varying volatility model to estimate, for example, $\Sigma_{t|t-1}$. However, different predictive strategies have different estimates of the conditional variance of the returns. In fact, in the penalised regression models, the returns variance is assumed constant in the data generating process. To address this issue, and in order to mitigate the effect of alternative volatility estimates on the resulting optimal weights, we consider, for each predictive strategy, the same rolling sample variance estimator as in Thornton and Valente (2012): $\widehat{\Sigma}_{t+1|t} = \sum_{l=0}^{\infty} \Omega_{t-l} \odot \epsilon_{t-l}\epsilon_{t-l}'$, where $\epsilon_t'$ the vector of forecast errors, $\Omega_{t-l} = \delta \exp(-\delta)\,\mathbf{1}\mathbf{1}'$ is a symmetric matrix of weights, $\odot$ denotes element-by-element multiplication, and we set the decay rate $\delta$ to 0.05.[24] We also winsorize the weights for each of the industry portfolio to $-1 \leq w_t \leq 1.5$ to prevent extreme investments; however, we evaluate the robustness of our results to alternative assumptions about the portfolio weights. Finally, to make our results directly comparable to other studies (e.g., Pettenuzzo et al., 2014; Pettenuzzo and Ravazzolo, 2016; Johannes et al., 2014), we assume a risk aversion $\gamma = 5$.

**3.3.1 Certainty equivalent returns.** Table 3 reports the results for the mean-variance case by looking both at a single industry portfolio, a more general multivariate cross-industry investment, as well as at a value-weighted market portfolio. Similar to the predictive performance, we report both the certainty equivalent gain and the p-value for the null hypothesis that the gain is null, with the t-statistics calculated from Eq. (11), which follows the logic of Diebold and Mariano (1995).

Few comments are in order. First, from a pure economic standpoint, the forecast from a

[22]Considering we are not dealing with single stocks but with industry portfolios, a careful investigation of transaction costs would be rather prohibitive. In this respect, assuming industry portfolios can be thought of as passive investment, a 10 basis point flat fee is rather in line with conventional passive investment strategies.

[23]Specifically, to evaluate the allocation implied by a given predictive strategy, we estimate the following regression:

$$u_{t,s} - u_{t,HA} = \mu + \varepsilon_t, \tag{11}$$

where the test is based on the sample mean of the difference in utilities $\mu$. Standard error are robust to heteroskedasticity and autocorrelation by using a Newey and West (1987) estimator with lags as suggested in Lazarus et al. (2018).

[24]This is the same value as in Thornton and Valente (2012) and within the range of those reported in studies like Fleming et al. (2001).

recursive mean are quite challenging to beat. Although a large fraction of methods deliver a positive certainty equivalent across industries, only for a handful of these the spread with respect to the certainty equivalent from the sample mean is statistically significant. This confirms previous evidence in, for example, Goyal and Welch (2008), Rapach et al. (2010), and Dangl and Halling (2012). Second, our DRS model outperforms both the model averaging and the recursive penalised regression estimates, even though some industry portfolios are borderline significant, with the health industry portfolio being the exception in terms of significance. For instance, when looking at the aggregate market portfolio, an investor is willing to pay 38 basis points to access the strategy based on a dynamic factor model, while the same investor would be willing to pay more than double, i.e., a 83 basis point fee, to have access to a strategy based on our DRS method. Consistent with the existing evidence (see, e.g., Avramov, 2002 and Rapach et al., 2010), forecast combination methods, such as equal-weight linear pooling, BMA, and both the median and the trimmed mean forecast represent two challenging benchmarks to beat in an economic sense. On the other hand, all penalised predictive regressions turn out to underperform the historical mean by a sizable margin, whereas dense factor models and non-linear machine learning methods are placed somewhat in between shrinkage methods and forecast combination strategies. More generally, the economic performance across industry portfolios confirm and extends some of the existing evidence at the market level (see, e.g., Pettenuzzo et al., 2014).

Similar to the statistical performance, we also report a more dynamic picture that shows the cumulative CERs gain (or loss) from the same competing strategies reported in Figure 1 against the historical mean. The top panels of Figure 3 reports the results for the mean-variance investor. The evidence that emerges from the average value across industries (left panel) and the value for the aggregate stock market (right panel), are similar. The trading signal coming from our DRS strategy tend to systematically outperform the signal obtained from the conditional mean.

As a robustness check, Table D.1 investigate the consistency of the results for a mean-variance investor when short sales are restricted, i.e., $0 \leq w_t \leq 1$. Although to a lesser extent, the results continue to be largely in favour of our DRS predictive strategy, vis-a-vis both forecast combination methods, non-linear machine learning methods, and sparse and dense penalised predictive regressions. More precisely, although the economic magnitude of the performance is reduced throughout, the pecking order and the rationale of the results remain intact. For instance, the implied fees an investor would be willing to pay to access the DRS forecasts for the aggregate market portfolio are around 40% higher than the one the very same investor would be willing to pay for either the median or the trimmed mean forecasts.

Turning to a more general power utility investor (see, e.g., Barberis, 2000, Johannes et al., 2014, Pettenuzzo et al., 2014, and Pettenuzzo and Ravazzolo, 2016), the general message of Table 3 is consistent. In fact, Table 4 confirms that our large-scale predictive regression framework tends to (1) significantly outperform the forecasts from a sample mean, and (2) outperform the competing model combination strategies, the recursive sparse and dense regression methods, and non-linear machine learning methods. These results hold both at the univariate level, i.e., investing in a single industry at a time, and at the multivariate level, i.e., investing in all industries at once, as well as at the market level, i.e., investing in a value-weighted portfolio. Interestingly, the performance of all competing methods tend to be slightly better within a power utility context than for mean-variance preferences. For instance, for the equal-weight linear pooling, a mean-variance investor would be willing to pay 9 basis points, net of fees, to access such strategy, whereas, for the same strategy, a CRRA investor would pay a doubled 17 basis points.

Table D.2 shows the economic performance of each model when restricting the portfolio weights to be non-negative. Similar to the mean-variance case, the main results of the paper are robust to the inclusion of short-sales constraints in the investment process. In fact, the economic performance slightly improves throughout, with only a few cases in which the forecast from the sample mean delivers better performances. That is, there is no substantial variation in the pecking order across models; sparse and dense regressions are outperformed by forecast combination techniques, which in turn are outperformed by our DRS predictive strategy. This can be seen by looking at the p-values in Panel B; although a large fraction of models deliver positive certainty equivalent returns with respect to the sample mean, such positive spread is statistical significant in only a handful of cases. The bottom panels of Figure 3 report the dynamics of the difference in CERs for the investor with CRRA preferences. The main results confirm the economic gains obtained by a mean-variance investor. Our DRS strategy tends to systematically outperform the signal obtained from the conditional mean.

**3.3.2 A discussion on the alternative predictive strategies.** The main empirical results suggest that our DRS strategy outperforms competing methods, both statistical and economically. More importantly, our strategy outperforms sparse regressions and standard forecast combination strategies, which have been shown to perform well, at least for the aggregate stock market (see, e.g., Rapach et al., 2010, Rapach et al., 2015, and Bianchi and Tamoni, 2019 among others).

One of the key aspects of financial ratios is that they may be highly correlated to each other. This affects the performance of sparse regression methods that make use of the whole set of regressors in a single fashion (see, e.g., Giannone et al., 2017 for a discussion on the "illusion of sparsity" in dense datasets). The left panel of Figure 4 shows the correlation structure of the predictors. For ease of exposition, the industry specific financial ratios are averaged across industries. Two interesting facts emerge. First, there is little correlation, on average, between industry financial ratios and aggregate macroeconomic indicators. In principle, this evidence lends support for the use of both industry-specific and aggregate macroeconomic predictors, which appear to not have too much overlapping information. Second, there are no signs of sparse correlation within industry-specific predictors. This lends support for explicitly taking into account dense modeling strategies, and in particular cross-correlations among different individual predictive regressions. Popular model selection and averaging priors do not explicitly model the correlation structure in the data when determining which variables are restricted to enter the regression (see, e.g., Giannone et al., 2017).

This could explain the underperformance of sparse penalised regressions.

A different, although related, issue is that even by clustering groups of financial ratios by their economic meaning, there is still substantial correlation among sub-model forecasts, which is not explicitly considered by standard forecast combination strategies. The right panel in Figure 4 shows this case in point. The figure shows the unconditional correlation among forecasts generated by using each of the groups of financial ratios, taken singularly. Notice this is not the DRS-implied sub-models correlation, but is simply the sample co-movement of each sub-model prediction when estimated separately. There is rather strong evidence that forecasts tend to co-move significantly over the sample. However, such correlation is completely discarded by non-parametric forecast combination strategies, such as the equal-weight linear pooling, the trimmed mean, and the median forecast (see, e.g., Timmermann, 2004 for some detailed discussion).

As a whole, Figure 4 suggests that one possible reason why our DRS strategy outperforms penalised regressions and forecast combination methods, for a large set of industry specific financial ratios, is that the dynamic synthesis, and the fact that smaller sets of ratios are efficiently synthesised, allows to mitigate identification concerns due to correlation (see also the discussion in Section 2.3). The next Section will delve further into the dynamics of the predictive synthesis implied by the DRS approach.

# 4 Dissecting Predictive Synthesis

In the previous sections, we have seen that the performance generated by our DRS strategy outperforms competing methods, both statistically and economically. This implies that, when synthesised carefully, financial ratios provide valuable information in forecasting stock returns, both at the industry and at the aggregate market level. We now delve further into the model dynamics, in an attempt to provide further insights on the key ingredients that possibly make our model successful in comparison to other alternative forecasting procedures.

To investigate the properties of our decouple-recouple predictive strategy, we show two different sets of results. First, we provide evidence of time variation in the covariance structure across forecasts. A key feature of our model is that it allows to capture, explicitly, the time-varying correlations of individual forecasts through the dynamics of the state equation, Eq. (9b). Second, we show that the dynamics of the aggregate bias, $\theta_{0t}$; the marginal effects of individual, group-specific, forecasts, $\theta_{tj}$, for $j = 1, \ldots, J$; and the uncertainty around the measurement error, $\sigma_t^2$, respond to aggregate macroeconomic conditions.

## 4.1 Model-implied covariance between sub-model forecasts

Figures 5-6 show the covariance among the individual forecasts for the aggregate stock market and the average industry at four different time periods; before, during, and after the great financial crisis of 2008/2009.[25]

Figure 5 shows three interesting results. First, there is substantial time series variation in the dynamics of the correlations amongst forecasts for the aggregate stock market. For instance, the difference in pre- and post-crisis periods and 2008 are rather clear from a visual perspective, with an increasing correlation between capitalization and financial soundness in January 2011, vis-a-vis a positive correlation between macroeconomic variables and profitability in August 2006. Second, the correlation across forecasts is primarily negative. This has important implications for the "diversification benefits" of forecast combination. Bunn (1985) shows that the aggregate loss, obtained by combining two models, is inversely related to the covariance between forecasts, that is, the more negative the correlation, the lower the aggregate forecasting loss in a mean squared sense (see, also Timmermann, 2004, for a discussion). Related to that, the third interesting result concerns the flip in the sign – from negative to positive – of some forecast covariance during the great financial crisis of 2008/2009. For instance, the covariance between the sub-models related to capitalization and value goes from negative to positive. From November 2008, there is evidence of increasing positive spillovers across the forecasts from different groups of financial ratios. This is consistent with the existing evidence on systemic risk and macroeconomic conditions, whereby increased correlation among financial and macroeconomic variables is observed during market crashes (see, e.g., Billio et al., 2012 and Bianchi et al., 2019).

A slightly different dynamic is shown in Figure 6, regarding the average industry. The cross-forecast covariances are primarily negative before the great financial crisis, suggesting significant diversification benefits in combining forecasts. However, a greater set of covariance switched from negative to positive, during the great financial crisis, the period until early 2011, and towards the end of the sample. This suggests that industry portfolios, on average, may be less affected by business cycles fluctuations.

## 4.2 Time-varying synthesis parameters

The dynamic properties of our decouple-recouple predictive model are summarised by the latent time-varying parameters $\mathbf{\Phi}_t = \left(\boldsymbol{\theta}', \sigma_t^2\right)$. In this section, we discuss the dynamics of $\mathbf{\Phi}_t$ against aggregate macroeconomic conditions.

**4.2.1 Aggregate bias and observation uncertainty.** Since the parameters of the recoupling step are considered to be latent states, the conditional intercept, $\theta_{0t}$, in Eq. (9a) can be interpreted as the aggregate bias, namely a free-roaming component, which is not directly pinned down by any group of predictors. Specifically, the time variation in the conditional intercept can be thought of as a reflection of unanticipated (by the group-specific models, and as an extension, the group

[25] We report the results for the average industry for ease of exposition. The industry-specific results are available upon request.

indicators) economic shocks, which then affect equity premium forecasts with some lag.[26]

For ease of exposition, the left panel of Figure 7 reports the posterior mean estimates of $\theta_{0t}$, averaged across industries (blue line with markers) and for the aggregate stock market (orange solid line). Two interesting facts emerge. First, there is discrepancy between the dynamics of the bias for the average industry, vis-a-vis the aggregate stock market, with a gap that is increasing in the aftermath of the financial crisis of 2008/2009. More specifically, the bias for the average industry is consistently lower than the aggregate market after 2008. This suggests that the variation underpinned by financial ratios at the industry level seems higher, on average, than the variation captured by the same financial ratios at the market level. The second finding is that there is a large drop in the bias during the market turmoil of 2008/2009. As the bias captures the variation in the data that is not underpinned by financial ratios, such drop in the bias suggests that the explanatory power of financial ratios is higher in recessions vs expansions, conditional on the intercept. This result confirms the unconditional evidence reported in Table 2 and expands some of the existing results related to the counter-cyclical returns predictability of the aggregate stock market (see, e.g., Rapach et al., 2010) by adding a different perspective to the dynamics of the structural parameters of the model, rather than at the $R^2_{oos}$.[27]

The right panel of Figure 7 reports another key parameter of our DRS predictive model, that is the standard deviation, $\sigma_t$, in the observation equation, Eq. (9a). By representing forecast combination as a state-space model, the "observation noise", $\epsilon_t$, can be directly interpreted as the measurement error. As a result, the latent parameter $\sigma_t$ should increase (decrease) when the aggregated forecast is more (less) uncertain. Again, one interesting aspect emerges. The uncertainty around the one-step ahead forecasts increases during the great financial crisis. This is not surprising, given such period corresponded to a profound market turmoil. However, consistent with the left panel in Figure 7, it turns out that the possibly lower forecasting power of financial ratios proxied by a higher $\theta_{0t}$ (orange solid line) also corresponds to a higher forecast uncertainty (see blue line with markers on the right panel). As a result, the dynamics of the $\theta_{0t}$, $\sigma_t$ parameters of our DRS model suggests substantial time variation in the pure forecasting ability of financial ratios over the business cycle. To the best of our knowledge, this result has not been previously shown in the literature.

**4.2.2 Marginal effects of sub-model forecasts.** The time variation in $\theta_{0t}$ is reflected in the dynamics of the latent inter-dependencies amongst individual forecasts through $\theta_{jt}$. More specifically, each $\theta_{jt}$ represents, conditional at time $t$, the marginal effect of an individual prediction for the target variable $y_t$, i.e., $\partial y_t / \partial x_{jt} = \theta_{jt}$ with $x_{tj} \sim p(y_t | \boldsymbol{C}^j_{t-1})$ (see Eqs. 9a-9b).

Figure 8 shows some of the posterior mean estimates of $\theta_{jt}$, for $j = 1, \ldots, J$, averaged across

[26] Note that the aggregate bias does not reflect, per se, the amount of predictability, which is a function of both the fitted value of the regression and the bias. In this respect, a reduced intercept does not necessarily translates into lower predictive performance, which is indeed the function, $F_t \boldsymbol{\theta}_t$.

[27] In a time-varying setting, an increasing $R^2_{oos}$ could be a function of a more significant intercept – that is capturing more precisely a time trend – rather than a function of a higher explanatory power of the predictors.

industries. The results show some interesting insights. First, there is a substantial time variation in the marginal effects of individual forecasts. In particular, abrupt changes in the relative effects of financial ratios can be identified around the great financial crisis, especially for the Manufacturing and Other industries. This is likely not due to idiosyncratic volatility effects, as we explicitly take into account time varying volatility for the unexpected returns for each of the group-specific regressions (see Eq. 9a). Second, the marginal effect of valuation and capital ratios tend to move in opposite directions. Although the interpretation of the dynamics of the latent inter-dependencies is not always clean, this possibly suggests there is some offsetting effects of both groups of financial ratios, especially in the aftermath of the great financial crisis. Similarly, the effects of financial ratios that are classified as liquidity and efficiency tend to diverge in the few years from 2008 to 2014, then converge towards the end of the sample. Third, except from an abrupt change in 2008, the marginal effect of profitability, financial soundness, and solvency remain rather stable. This is not a trivial implication and instead shows that the fact that we impose a random-walk dynamics to $\theta_{jt}$ does not prevent the model to be "spillovers" in individual forecasts to be stable over time.

Finally, the marginal effect of the forecast from macroeconomic variables tend to weigh negatively on the stock returns forecasts at the industry level. Note that, this does not imply that increasing macroeconomic conditions predict negative returns, but rather suggest that, conditional on financial ratios, forecasts from aggregate macroeconomic and financial variables tend to have diversification benefits on the aggregate forecasting loss in a mean squared sense (see, Timmermann, 2004).

Figure (9) shows the same marginal effects, but for the aggregate stock market. Except for few nuances, the picture that emerges from the model dynamics is similar to the average industry. For instance, financial ratios that are classified as liquidity, efficiency, financial soundness, and other, all have a marginal positive relationship with the aggregate forecast. Also, aggregate macroeconomic variables keep their negative marginal effect. However, differently from the average industry results, valuation, capitalization, and profitability variables all have a slightly negative marginal effect on the aggregate forecast.

As a whole, Figures (8)-(9) show that (1) there is a substantial time variation in the way individual predictions from groups of financial ratios interact to produce stock returns forecasts from the model, and (2) such time variation reflects aggregate macroeconomic conditions. This lends support for models with time-varying parameters in a data-rich environment, such as using a large set of financial ratios.

However, Figures (7)-(8) show the average effect across industries and do not suggest any insight into why different industries show different levels of predictability. To gain further insight into the cross-sectional heterogeneity of returns predictability, we now focus on a more granular representation of the time-varying synthesis parameters for each industry. For ease of exposition, Figure (10) reports two key parameters of interest, that is the aggregate bias and the observation uncertainty (top panels), and two representative examples of the marginal effects of sub-model forecasts (bottom panels).

Two interesting aspects emerge from the aggregate bias and the measurement error. First, a substantial time series variation is coupled with a fairly relevant cross-sectional dispersion in the parameters across industries. For instance, while the bias for sectors, such as Durables and Others, turn from positive to negative during the great financial crisis, other industries, such as Energy and Non-durables, consistently show a positive bias, albeit moving closer to zero during the 2008/2009 financial crisis. Second, there is a clear inverse relationship between the aggregate bias and the observation uncertainty. For instance, Durables, which shows the most volatile bias, also shows the highest observation uncertainty during the great financial crisis, while the opposite holds for the Non-durables industry.

Turning to the marginal effects of the forecasts, we report, as an example, the effect of the forecasts from sub-models, which include variables labelled as valuation ratios (left panel) and profitability (right panel). Again, one clear pattern emerge. There is a substantial heterogeneity in the marginal effects of valuation ratios across industries. For instance, while the vast majority of industries show a positive marginal effect, $\theta_{jt}$ for the Durables sector turned persistently negative after 2008/2009. This explains the effect on the aggregate stock market – see Figure (9) –, given the weight of the Durables sector on the market portfolio. A similar argument holds for the average industry – see Figure (8) –, although the effect is diluted, given all industries carry the same weight in the cross-sectional average. Similarly, the marginal effect of profitability largely varies across industries. For instance, while $\theta_{jt}$ becomes increasingly positive after the 2008/2009 for the Others and Manufacturing industries, the same does not hold for the vast majority of the other industries, whereby there is a rather downward trending effect. Again, this reflects both in the marginal effect of profitability on the aggregate stock market portfolio – see top-right panel of Figure (9) –, and the average industry – see top-right panel of Figure (9).

As a whole, Figure (10) suggests that the difference in the statistical and economic significance of our DRS predictive strategy in forecasting industry portfolio returns – see Tables (1)-(4) – can be due to an substantial heterogeneity in the way groups of financial ratios interact when forecasting returns. From an asset pricing standpoint, such heterogeneity suggests that different factors are priced differently across industries. In addition, the fact that the average marginal effect and the marginal effects on the market portfolio somewhat align, suggest that, almost by construction, the aggregate market portfolio represents a value-weighted average of the industry-specific pricing kernels (see, e.g., Fama and French, 1997 and Huang et al., 2015).

## 5 Conclusion

We are interested in the predictability of the equity premium across different industries in the US, based on a large set of financial ratios. Methodologically, we contribute to the literature by proposing a novel model combination scheme that retains all of the information available to investors.

The empirical results suggest that financial ratios, above and beyond typical measures of stock prices relative to fundamentals, provide valuable information for forecasting stock returns, both at the industry and at the aggregate market level. More precisely, our forecasting approach significantly outperforms, both statistically and economically, forecasts from a variety of competing strategies, such as sparse and dense regressions, forecast combination methods, and non-linear machine learning methods.

Delving further into the key features of the model, we show that the dynamic properties of our model correlate with aggregate macroeconomic conditions and differ across industries, lending support to industry-specific asset pricing models.

## References

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

Adrian, T. and Franzoni, F. (2009). Learning about beta: Time-varying factor loadings, expected returns, and the conditional capm. *Journal of Empirical Finance*, (16):537–556.

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

Avramov, D. (2004). Stock return predictability and asset pricing models. *Review of Financial Studies*, 17(3):699–738.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring economic policy uncertainty. *The quarterly journal of economics*, 131(4):1593–1636.

Ball, R., Gerakos, J., Linnainmaa, J. T., and Nikolaev, V. (2016). Accruals, cash flows, and operating profitability in the cross section of stock returns. *Journal of Financial Economics*, 121(1):28–45.

Barberis, N. (2000). Investing for the long-run when returns are predictable. *The Journal of Finance*, 55(1):225–264.

Bates, J. M. and Granger, C. W. J. (1969). The combination of forecasts. *Operational Research Quarterly*, 20:451–468.

Bernanke, B. S., Boivin, J., and Eliasz, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (favar) approach. *The Quarterly journal of economics*, 120(1):387–422.

Bianchi, D., Billio, M., Casarin, R., and Guidolin, M. (2019). Modeling systemic risk with markov switching graphical sur models. *Journal of econometrics*, 210(1):58–74.

Bianchi, D., Büchner, M., and Tamoni, A. (2020). Bond risk premia with machine learning. *The Review of Financial Studies*, (forthcoming).

Bianchi, D., Guidolin, M., and Ravazzolo, F. (2017). 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(1):110–129.

Bianchi, D. and Tamoni, A. (2019). Sparse predictive regressions: Statistical performance and economic significance. *Working Paper*.

Billio, M., Getmansky, M., Lo, A. W., and Pelizzon, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of financial economics*, 104(3):535–559.

Binsbergen, V., Jules, H., and Koijen, R. S. (2010). Predictive regressions: A present-value approach. *The Journal of Finance*, 65(4):1439–1471.

Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2):123–140.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Bunn, D. W. (1985). Statistical efficiency in the linear combination of forecasts. *International Journal of Forecasting*, 1(2):151–163.

Campbell, J. Y. (1987). Stock returns and the term structure. *Journal of financial economics*, 18(2):373–399.

Campbell, J. Y. and Thompson, S. B. (2007). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, 21(4):1509–1531.

Campbell, J. Y. and Viceira, L. M. (2004). Long-horizon mean-variance analysis: A user guide. *Manuscript, Harvard University, Cambridge, MA*.

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 51:1281–1304.

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

Clark, T. E. and West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of econometrics*, 138(1):291–311.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5(4):559–583.

Clemen, R. T. and Winkler, R. L. (1999). Combining probability distributions from experts in risk analysis. *Risk Analysis*, 19:187–203.

Clemen, R. T. and Winkler, R. L. (2007). Aggregating probability distributions. In W. Edwards, R. M. and von Winterfeldt, D., editors, *Advances in Decision Analysis: From Foundations to Applications*, chapter 9, pages 154–176. Cambridge University Press.

Cochrane, J. H. (2011). Presidential address: Discount rates. *The Journal of finance*, 66(4):1047–1108.

Cohen, L. and Frazzini, A. (2008). Economic links and predictable returns. *The Journal of Finance*, 63(4):1977–2011.

Connor, G. and Korajczyk, R. A. (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics (JFE)*, 15(3).

Dangl, T. and Halling, M. (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics*, 106(1):157–181.

Daniel, K., Grinblatt, M., Titman, S., and Wermers, R. (1997). Measuring mutual fund performance with characteristic-based benchmarks. *The Journal of finance*, 52(3):1035–1058.

De Mol, C., Giannone, D., and Reichlin, L. (2008). Forecasting using a large number of predictors: Is bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics*, 146(2):318–328.

DeMiguel, V., Martin-Utrera, A., Nogales, F. J., and Uppal, R. (2020). A transaction-cost perspective on the multitude of firm characteristics. *The Review of Financial Studies*, 33(5):2180–2222.

DeMiguel, V., Nogales, F. J., and Uppal, R. (2014). Stock return serial dependence and out-of-sample portfolio performance. *The Review of Financial Studies*, 27(4):1031–1073.

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

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & economic statistics*, 20:134–144.

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

Elliott, G., Gargano, A., and Timmermann, A. (2013). Complete subset regressions. *Journal of Econometrics*, 177(2):357–373.

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

Fama, E. F. and French, K. R. (1997). Industry costs of equity. *Journal of financial economics*, 43(2):153–193.

Fama, E. F. and French, K. R. (2006). Profitability, investment and average returns. *Journal of financial economics*, 82(3):491–518.

Feng, G., Giglio, S., and Xiu, D. (2019). Taming the factor zoo: A test of new factors. Technical report, National Bureau of Economic Research.

Ferson, W. E. and Harvey, C. R. (1991). The variation of economic risk premiums. *Journal of political economy*, 99(2):385–415.

Ferson, W. E. and Harvey, C. R. (1999). Conditioning variables and the cross section of stock returns. *The Journal of Finance*, 54(4):1325–1360.

Ferson, W. E. and Korajczyk, R. A. (1995). Do arbitrage pricing models explain the predictability of stock returns? *Journal of Business*, pages 309–349.

Fleming, J., Kirby, C., and Ostdiek, B. (2001). The Economic Value of Volatility Timing. *Journal of Finance*, 56(1):329–352.

Freyberger, J., Neuhierl, A., and Weber, M. (2020). Dissecting characteristics nonparametrically. *The Review of Financial Studies*, 33(5):2326–2377.

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

Gargano, A., Pettenuzzo, D., and Timmermann, A. (2019). Bond return predictability: Economic value and links to the macroeconomy. *Management Science*, 65(2):508–540.

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

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

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

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

Gomes, J. F. and Schmid, L. (2010). Levered returns. *The Journal of Finance*, 65(2):467–494.

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

Green, J., Hand, J. R. M., and Zhang, X. F. (2017). The Characteristics that Provide Independent Information about Average U.S. Monthly Stock Returns. *The Review of Financial Studies*, 30(12):4389–4436.

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

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

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5):2223–2273.

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

Harvey, C. R. and Liu, Y. (2019). A census of the factor zoo. *Available at SSRN 3341728*.

Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of forecasting*, 13(2):281–291.

Henkel, S. J., Martin, J. S., and Nardari, F. (2011). Time-varying short-horizon predictability. *Journal of financial economics*, 99(3):560–580.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.

Hsiang, T. (1975). A bayesian view on ridge regression. *Journal of the Royal Statistical Society: Series D*, 24(4):267–268.

Huang, C.-f. (1987). An intertemporal general equilibrium asset pricing model: The case of diffusion information. *Econometrica: Journal of the Econometric Society*, pages 117–142.

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

Irie, K., West, M., et al. (2019). Bayesian emulation for multi-step optimization in decision problems. *Bayesian Analysis*, 14(1):137–160.

Johannes, M., Korteweg, A., and Polson, N. (2014). Sequential Learning, Predictability, and Optimal Portfolio Returns. *Journal of Finance*, 69(2):611–644.

Jostova, G. and Philipov, A. (2005). Bayesian analysis of stochastic betas. *Journal of Financial and Quantitative Analysis*, 40:747–778.

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

Kelly, B. T., Pruitt, S., and Su, Y. (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, 134(3):501–524.

Kirby, C. (1998). The restrictions on predictability implied by rational asset pricing models. *The Review of Financial Studies*, 11(2):343–382.

Lazarus, E., Lewis, D. J., Stock, J. H., and Watson, M. W. (2018). Har inference: Recommendations for practice. *Journal of Business & Economic Statistics*, 36(4):541–559.

Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics*, 74(2):209–235.

Lewellen, J., Nagel, S., and Shanken, J. (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial economics*, 96(2):175–194.

Li, D. (2011). Financial constraints, r&d investment, and stock returns. *The Review of Financial Studies*, 24(9):2974–3007.

Li, Q., Lin, N., et al. (2010). The bayesian elastic net. *Bayesian analysis*, 5(1):151–170.

Lopes, H. F., McCulloch, R. E., and Tsay, R. S. (2016). Parsimony inducing priors for large scale state-space models. *Bayesian Anal.*

Makridakis, S. (1989). Why combining works? *International Journal of Forecasting*, 5(4):601–603.

Manzan, S. (2015). Forecasting the distribution of economic variables in a data-rich environment. *Journal of Business & Economic Statistics*, 33(1):144–164.

McAlinn, K., Aastveit, K. A., Nakajima, J., and West, M. (2020). Multivariate bayesian predictive synthesis in macroeconomic forecasting. *Journal of the American Statistical Association*, 115(531):1092–1110.

McAlinn, K. and West, M. (2019). Dynamic bayesian predictive synthesis in time series forecasting. *Journal of econometrics*, 210(1):155–169.

McCracken, M. W. and Ng, S. (2016). Fred-md: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, 34(4):574–589.

Menzly, L. and Ozbas, O. (2010). Market segmentation and cross-predictability of returns. *The Journal of Finance*, 65(4):1555–1580.

Nardari, F. and Scruggs, J. (2007). Bayesian analysis of linear factor models with latent factors, multivariate stochastic volatility, and apt pricing restrictions. *Journal of Financial and Quantitative Analysis*, 42(4):857–891.

Newey, W. K. and West, K. D. (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, pages 777–787.

Park, T. and Casella, G. (2008). The bayesian lasso. *Journal of the American Statistical Association*, 103(482):681–686.

Pastor, L. and Stambaugh, R., F. (2009). Predictive systems: Living with imperfect predictors. *The Journal of Finance*, (4):1583–1628.

Pastor, L. and Stambaugh, R., F. (2012). Are stocks really less volatile in the long-run? *The Journal of Finance*, (2):431–477.

Pástor, L. and Stambaugh, R. F. (2003). Liquidity risk and expected stock returns. *Journal of Political Economy*, 111(3):642–685.

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

Pettenuzzo, D., Timmermann, A., and Valkanov, R. (2014). Forecasting stock returns under economic constraints. *Journal of Financial Economics*, 114(3):517–553.

Polson, N. G. and Scott, J. G. (2010). Shrink globally, act locally: Sparse bayesian regularization and prediction. *Bayesian statistics*, 9:501–538.

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

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

Rapach, D. E., Strauss, J., Tu, J., and Zhou, G. (2015). Industry interdependencies and cross-industry return predictability.

Rossi, A. G. (2018). Predicting stock market returns with machine learning. Technical report, Working paper.

Rossi, A. G. and Timmermann, A. (2015). Modeling covariance risk in merton's icapm. *The Review of Financial Studies*, 28(5):1428–1461.

Sarno, L., Schneider, P., and Wagner, C. (2016). The economic value of predicting bond risk premia. *Journal of Empirical Finance*, 37:247–267.

Shirota, S., Omori, Y., Lopes, H. F., and Piao, H. (2017). Cholesky realized stochastic volatility model. *Econometrics and Statistics*, 3:34–59.

Smith, J. and Wallis, K. F. (2009). A simple explanation of the forecast combination puzzle. *Oxford Bulletin of Economics and Statistics*, 71(3):331–355.

Stambaugh, R. F. (1983). Arbitrage pricing with information. *Journal of Financial Economics*, 12(3):357–369.

Stevanovic, D. (2017). Macroeconomic forecast accuracy in a data-rich environment. *Working Paper*.

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167–1179.

Stock, J. H. and Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23(6):405–430.

Stock, J. H. and Watson, M. W. (2005). Implications of dynamic factor models for var analysis. Technical report, National Bureau of Economic Research.

Stock, J. H. and Watson, M. W. (2010). Modeling inflation after the crisis. Technical report, National Bureau of Economic Research.

Thornton, D. L. and Valente, G. (2012). Out-of-sample predictions of bond excess returns and forward rates: An asset allocation perspective. *The Review of Financial Studies*, 25(10):3141–3168.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288.

Timmermann, A. (2004). Forecast combinations. In Elliott, G., Granger, C. W. J., and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 1, chapter 4, pages 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 Crosse, J. (1992). Modelling of probabilistic agent opinion. *Journal of the Royal Statistical Society (Series B: Methodological)*, 54:285–299.

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

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

Zhao, Z. Y., Xie, M., and West, M. (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.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320.

# Appendix

This Appendix provides additional details regarding the data, the estimation strategy for our methodology as well as for some of the competing methods for which estimation is non-standard, some additional out-of-sample empirical results as well as the results from a simple simulation study. Unless otherwise specified, all notations and model definitions are similar to those in the main text.

## A Description of the Financial Ratios

WRDS Financial Ratio (WFR hereafter) is a collection of most commonly used financial ratios by academic researchers. There are in total over 70 financial ratios grouped into the following seven categories: Capitalization, Efficiency, Financial Soundness/Solvency, Liquidity, Profitability, Valuation and Others. Table A.1 provides a detailed description of all financial ratios used in the empirical analysis.

Industry aggregation is based on the four-digit SIC codes of the existing firm at each timet. We use the ten industry classification codes obtained from Kenneth French's website. Aggregation of firm-specific characteristics is constructed by taking the value-weighted mean of the firm-specific values within a given industry.

The final outputs for both individual firm and industry-level aggregated value are at monthly frequency. In order to populate the data to monthly frequency, we carry forward the most recent quarterly or annual data item, whichever is most recently available at a given time stamp, to the subsequent months before the next filing data becomes available.

In addition, in order to make sure that all data is publicly available at the monthly time stamp, we lag all observations by two months to avoid any look ahead bias.

## B MCMC Algorithm

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

In our sequential learning and forecasting context, the full MCMC analysis is performed in an extending window manner, re-analyzing the data set as time and data accumilates. We detail

MCMC steps for a specific time $t$ here, based on all data up until that time point.

### B.1 Initialization:

First, initialize by setting $\boldsymbol{F}_t = (1, x_{t1}, ..., x_{tJ})'$ for each $t = 1{:}T$ at some chosen initial values of the latent states. Initial values can be chosen arbitrarily, though following McAlinn and West (2019) we recommend sampling from the priors, i.e., from the forecast distributions, $x_{tj} \sim p(y_t|\boldsymbol{C}^j_{t-1})$ independently for all $t = 1{:}T$ and $j = 1{:}J$.

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

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

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

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

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

#### B.2.1 Forward filtering:.

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

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

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

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

2. *Update to time $t$ prior:*

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

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

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

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

4. *Filtering update to time t posterior:*

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

with defining parameters as follows:
   i. For $\boldsymbol{\theta}_t|v_t : \boldsymbol{m}_t = \boldsymbol{m}_{t-1} + \boldsymbol{A}_t e_t$ and $\boldsymbol{C}_t = r_t(\boldsymbol{R}_t - q_t\boldsymbol{A}_t\boldsymbol{A}_t')$,
   ii. For $v_t : n_t = \beta n_{t-1} + 1$ and $s_t = r_t s_{t-1}$,

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

**B.2.2 Backward sampling:.** Having run the forward filtering analysis up to time $T$, the backward sampling proceeds as follows.

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

b. *Recurse back over times* $t = T-1, T-2, \ldots, 0$ : At time $t$, sample $\boldsymbol{\Phi}_t = (\boldsymbol{\theta}_t, v_t)$ as follows:
   i. Simulate the volatility $v_t$ via $v_t^{-1} = \beta v_{t+1}^{-1} + \gamma_t$ where $\gamma_t$ is an independent draw from $\gamma_t \sim G((1-\beta)n_t/2, n_t s_t/2)$,
   ii. Simulate the state $\boldsymbol{\theta}_t$ from the conditional normal posterior $p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, v_t, \boldsymbol{x}_{1:T}, y_{1:T})$ with mean vector $\boldsymbol{m}_t + \delta(\boldsymbol{\theta}_{t+1} - \boldsymbol{m}_t)$ and variance matrix $\boldsymbol{C}_t(1-\delta)(v_t/s_t)$.

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

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

$$p(\boldsymbol{x}_t|\boldsymbol{\Phi}_t, y_t, \mathcal{H}_t) \propto N(y_t|\boldsymbol{F}_t'\boldsymbol{\theta}_t, v_t)\prod_{j=1:J} p(y_t|\boldsymbol{C}_{t-1}^j) \quad \text{where} \quad \boldsymbol{F}_t = (1, x_{t1}, x_{t2}, ..., x_{tJ})'. \tag{B.1}$$

Since $x_{tj} \sim p(y_t|\boldsymbol{C}_{t-1}^j)$ has density $T_{n_{tj}}(h_{tj}, H_{tj})$, we can express this as a scale mixture of Normal, $N(h_{tj}, H_{tj})$, with $\boldsymbol{H}_t = \text{diag}(H_{t1}/\phi_{t1}, H_{t2}/\phi_{t2}, ..., H_{tJ}/\phi_{tJ})$, where $\phi_{tj}$ are independent over $t, j$ with gamma distributions, $\phi_{tj} \sim G(n_{tj}/2, n_{tj}/2)$.

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

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

where $c_t = y_t - \theta_{0t} - \boldsymbol{h}_t'\boldsymbol{\theta}_{t,1:J}$, $g_t = v_t + \boldsymbol{\theta}_{t,1:J}'\boldsymbol{q}_t\boldsymbol{\theta}_{t,1:J}$, and $\boldsymbol{b}_t = \boldsymbol{q}_t\boldsymbol{\theta}_{t,1:J}/g_t$. Here, given the previous values of $\phi_{tj}$, we have $\boldsymbol{H}_t = \text{diag}(H_{t1}/\phi_{t1}, H_{t2}/\phi_{t2}, ..., H_{tJ}/\phi_{tJ})$ Then, conditional on these new samples of $\boldsymbol{x}_t$, updated samples of the latent scales are drawn from the implied set of conditional gamma posteriors $\phi_{tj}|x_{tj} \sim G((n_{tj}+1)/2, (n_{tj}+d_{tj})/2)$ where $d_{tj} = (x_{tj} - h_{tj})^2/H_{tj}$, independently for each $t, j$. This is easily computed and then sampled independently for each 1:$T$ to provide resimulated agent states over 1:$T$.

### B.4 Forecasting

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

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

In general, there are two ways to forecast over multiple horizons, through traditional DLM updating or through customized synthesis. The former, direct approach follows traditional DLM updating and forecasting via simulation as for 1-step ahead, where the synthesis parameters are simulated forward from time $t$ to $t + k$. The latter, customized synthesis involves a trivial modification, in which the model at time $t - 1$ for predicting $y_t$ is modified so that the $k$-step ahead forecast densities made at time $t - k$, replace $x_{tj}$. 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 (2019) find that, compared to the direct approach, the customized synthesis approach significantly improves multi-step ahead forecasts, since the dynamic model parameters, $\{\boldsymbol{\theta}_t, v_t\}$, are now explicitly geared to the $k$-step horizon.

## C Competing predictive strategies

In this section we describe in more detail the competing predictive strategies implemented in the main empirical analysis.

### C.1 Ridge regression

The ridge regression prior implies a closed-form penalised least squares representation of the form

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{z}'\boldsymbol{z} + \frac{1}{\tau^2}\mathrm{I}_p\right)^{-1} \boldsymbol{z}'\boldsymbol{y}$$

where $\boldsymbol{z}$ and $\boldsymbol{y}$ are the matrix of predictors and the vector of response variable, respectively. The posterior mean estimates correspond to the estimates obtained by using a standard $L_2$ penalty term, i.e., $\phi(\boldsymbol{\beta};\cdot) = \tau^{-2}\sum_{j=1}^{p}\beta_j^2$. The parameter $\tau$ determines the amount of global shrinkage, with smaller values resulting in more shrinkage and with $\tau \to \infty$ obtaining the OLS estimator.

### C.2 Lasso

In addition to the ridge, we implement a sparse regression method in the form of Bayesian lasso, as originally outlined by Tibshirani (1996). The intuition is that the lasso penalty is equivalent to the posterior mode estimate under a Laplace prior of the form,

$$p\left(\boldsymbol{\beta}|\sigma^2,\ldots\right) \propto \exp\left(-\frac{\tau^2}{\sigma}\sum_{i=1}^{n}|\beta_i|\right). \tag{C.1}$$

Park and Casella (2008) build on this intuition and suggest that a consistent hierarchical prior can be defined as a scale mixture of Normals, with an exponential mixing density, i.e.,[28]

$$\begin{aligned}
\boldsymbol{\beta}|\sigma^2,\boldsymbol{D}_\lambda &\sim N\left(\boldsymbol{0},\sigma^2\boldsymbol{D}_\lambda\right), \qquad \boldsymbol{D}_\lambda = \mathrm{diag}\left(\lambda_1^2,\ldots,\lambda_p^2\right)\\
\lambda_1^2,\ldots,\lambda_p^2|\tau^2 &\sim \prod_{j=1}^{p}\frac{\tau^2}{2}\exp\left(-\tau^2\lambda_j^2/2\right)d\lambda_j^2, \qquad \lambda_1,\ldots,\lambda_p > 0\\
\sigma^2 &\sim \pi\left(\sigma^2\right)d\sigma^2, \qquad \sigma^2 > 0
\end{aligned} \tag{C.2}$$

With such prior, the posterior mode estimates are similar to the estimates under the penalty term $\phi(\boldsymbol{\beta};\cdot) = \tau^{-2}\sum_{j=1}^{p}|\beta_j|$. This translates in a relatively standard Gibbs sampler which exploits the conjugacy of the inverse Gaussian distribution. More precisely, the conditional distribution of the regression parameters is given by

$$\boldsymbol{\beta}|\sigma^2,\lambda_1^2,\ldots,\lambda_p^2,\boldsymbol{z},\boldsymbol{y} \sim N\left(\boldsymbol{A}^{-1}\boldsymbol{z}'\tilde{\boldsymbol{y}},\sigma^2\boldsymbol{A}^{-1}\right) \tag{C.3}$$

with $\boldsymbol{A} = \left(\boldsymbol{z}'\boldsymbol{z} + \boldsymbol{D}_\lambda^{-1}\right)$. Similarly, the posterior distribution for $\sigma^2$ is conjugate and takes the form of a conditional inverse gamma, i.e.,

$$\sigma^2|\boldsymbol{\beta},\lambda_1^2,\ldots,\lambda_p^2,\boldsymbol{z},\boldsymbol{y} \sim IG\left(\frac{n-1+p}{2},\frac{1}{2}\left(\tilde{\boldsymbol{y}}-\boldsymbol{z}\boldsymbol{\beta}\right)'\left(\tilde{\boldsymbol{y}}-\boldsymbol{z}\boldsymbol{\beta}\right)+\frac{\tau}{2}\boldsymbol{\beta}'\boldsymbol{D}_\lambda^{-1}\boldsymbol{\beta}\right) \tag{C.4}$$

[28] As highlighted in Bianchi and Tamoni (2019), on of the key advantages of the Bayesian lasso with respect to a frequentist approach is that the penalization is done via both a global shrinkage parameter $\tau^2$ and local, i.e., regressor specific, shrinkage parameters. Hence, the Bayesian lasso expresses the penalised regression as a global-local shrinkage estimator. This, in turn, increases the flexibility of the model.

Finally, the conditional posterior distribution of the local shrinkage parameter $1/\lambda_j^2$ is the inverse Gaussian distribution with local and scale parameters as follows,

$$1/\lambda_j^2 = \gamma_j|\boldsymbol{\beta}, \sigma^2, \boldsymbol{z}, \boldsymbol{y} \sim \text{Inverse-Gaussian}\left(\frac{\tau^2\sigma^2}{|\beta_j|}, \tau^2\right)\mathbb{I}\left(\gamma_j > 0\right) \tag{C.5}$$

### C.3 Elastic net

Under certain conditions, Polson and Scott (2010) show that the lasso penalty can lead to over-shrinkage, namely it does not enjoy the so-called "oracle property"; the model selected by the lasso is not necessarily the true data generating process. Originally proposed by Zou and Hastie (2005), the elastic net mitigates the lasso concerns related to $n > T$ and the presence of group-wise correlated regressors. Li et al. (2010) show that the shrinkage prior for the elastic net can be expressed as

$$p\left(\boldsymbol{\beta}|\sigma^2, \dots\right) \propto \exp\left(-\frac{\tau_1^2}{\sigma}\sum_{i=1}^{n}|\beta_i| - \frac{\tau_2^2}{2\sigma}\sum_{i=1}^{n}\beta_i^2\right), \tag{C.6}$$

where the two parameters $\tau_1^2$ and $\tau_2^2$ determine the relative effect of the lasso and the ridge penalty, respectively. Li et al. (2010) show that the scale mixture of normals prior (C.6) implies a set of full conditional distributions as follows,[29]

$$\begin{aligned}
&\boldsymbol{\beta}|\sigma^2, \lambda_1^2, \dots, \lambda_p^2, \boldsymbol{z}, \boldsymbol{y} \sim N\left(\boldsymbol{A}^{*-1}\boldsymbol{z}'\tilde{\boldsymbol{y}}, \sigma^2\boldsymbol{A}^{*,-1}\right), \\
&\sigma^2|\boldsymbol{\beta}, \lambda_1^2, \dots, \lambda_p^2, \boldsymbol{z}, \boldsymbol{y} \sim IG\left(\frac{n-1+p}{2}, \frac{1}{2}\left(\tilde{\boldsymbol{y}} - \boldsymbol{z}\boldsymbol{\beta}\right)'\left(\tilde{\boldsymbol{y}} - \boldsymbol{z}\boldsymbol{\beta}\right) + \frac{\tau}{2}\boldsymbol{\beta}'\boldsymbol{D}_\lambda^{*,-1}\boldsymbol{\beta}\right), \\
&1/\lambda_j^2 = \gamma_j|\boldsymbol{\beta}, \sigma^2, \boldsymbol{z}, \boldsymbol{y} \sim \text{Inverse-Gaussian}\left(\frac{\tau_1^2\sigma^2}{\beta_j^2}, \tau_1^2\right)\mathbb{I}\left(\gamma_j > 0\right)
\end{aligned} \tag{C.7}$$

where $\boldsymbol{A}^* = \left(\boldsymbol{z}'\boldsymbol{z} + \boldsymbol{D}_\lambda^{*-1}\right)$, and $\boldsymbol{D}^*$ is a diagonal matrix with diagonal elements $\left(\lambda_j^{-2} + \tau_2^2\right)^{-1}, j = 1, \dots, p$. It is easy to see that in their functional form the conditional distribution of the elastic net parameters is very similar to the lasso, with only minor exceptions. As far as the shrinkage parameters are concerned, assuming two gamma priors $G\left(r_h, \delta_h\right)$, for $h = 1, 2$, the full conditional distribution of the shrinkage parameters is as follows,

$$\begin{aligned}
&\tau_1^2|\boldsymbol{\beta}, \sigma^2, \lambda_1^2, \dots, \lambda_p^2, \boldsymbol{z}, \boldsymbol{y} \sim G\left(p + r_1, \frac{1}{2}\sum_{j=1}^{p}\lambda_j^2 + \delta_1\right), \\
&\tau_2^2|\boldsymbol{\beta}, \sigma^2, \lambda_1^2, \dots, \lambda_p^2, \boldsymbol{z}, \boldsymbol{y} \sim G\left(p/2 + r_2, \frac{1}{2\sigma^2}\sum_{j=1}^{p}\beta_j^2 + \delta_2\right)
\end{aligned} \tag{C.8}$$

[29] The advantage of the Bayesian approach is that the parameters $\tau_1^2$ and $\tau_2^2$ can be estimated jointly, while in the standard frequentist approach, one needs to implement a sequential cross-validation procedure. However, the latter has been shown to result in over-shrinkage of the coefficients (see, Zou and Hastie, 2005).

### C.4 Random forest

Random forest is a supervised learning algorithm that uses ensemble learning methods by averaging over multiple regression trees (see Breiman, 2001). Similar to model averaging, ensemble learning combines predictions from a set of submodels to make a more accurate prediction. Each regression tree is a nonparametric model that approximates an unknown nonlinear function with piece-wise constant functions, using recursive partitioning of the covariate space (see Breiman, 1996). While standard regression trees are flexible in approximating nonlinear functions, such methods suffer from over-fitting, leading to low bias, high variance predictions. To mitigate this, random forests average over multiple regression trees based on bootstrapped samples, effectively reducing variance. Given the industry portfolio returns, $y$, the financial ratios, $\boldsymbol{C}$, and a number of terminal nodes, $K$, the partitioning of the covariates space is determined by minimizing the sum of squared errors of the following regression model

$$y = \sum_{k=1}^{K} c_k \mathbb{I}_k(\boldsymbol{C}; \boldsymbol{\theta}_k),$$

where $c_k$ is the regression coefficient and $\mathbb{I}_k(\boldsymbol{C}; \boldsymbol{\theta}_k)$ is an indicator function

$$\mathbb{I}_k(\boldsymbol{C}; \boldsymbol{\theta}_k) = \begin{cases} 1 & \text{if } \boldsymbol{C} \in R_k(\boldsymbol{\theta}_k) \\ 0 & \text{otherwise,} \end{cases}$$

such that $\boldsymbol{\theta}_k$ is the set of parameters that define the $k$th partition state $R_k(\boldsymbol{\theta}_k)$ in the covariate space. As a result, a random foreast is equivalent to a regression with $K$ indicator functions that condition the covariates. The model is implemented by using the `Statistics and Machine Learning Toolbox` in `Matlab 2020a`; in particular, we used the default options for the `RegressionBaggedEnsemble` object created by `fitrensemble` for regression. In order to isolate the effect of non-linearities vs weighting of each of the regression tree, the aggregation of each sub-model in the random forest is done by equal weighting each of the regression trees.[30]

### C.5 Neural network

In this exercise, we use a popular neural network structure for time series forecasting, called long short-term memory (LSTM; Hochreiter and Schmidhuber, 1997). LSTM is a type of recurrent neural network (RNN), which belongs a broader class of fully connected neural networks (NN) developed for time series data. Compared to regular NNs, RNNs consists of a NN that is repeated for the number of data lags in the input. RNN has two important implications for time series predictions. The first is that the network is explicitly informed of the temporal relation, in particular how $y_{t-1}$ comes after $y_{t-2}$, thus reflecting how more recent data are more relevant for prediction.

[30] Alternatively, one could weight each regression tree based on past measures of accuracy in the submodels forecasts, i.e., boosted regression tree.

The second, compared to more traditional auto-regressive-type models, is that it retains memory of distant input lags. LSTMs, on top of RNNs, has an internal structure of what are called gates. These gates allow the network to automatically decide what part of the network state to remember or forget, thus allowing for the model to remember the relevant input lags, while forgetting the irrelevant ones. For this study, the amount of information remembered between time steps, is set to 50 time lags, considering the persistence of some of the financial ratios. The model is implemented using the default options in the `Deep Learning Toolbox` in `Matlab 2020a`.[31]

# D Further results

In this section, we report additional results on the economic significance of our large-scale dynamic predictive regression strategy, vis-á-vis competing sparse and dense dynamic and recursive forecasting methods. More specifically, we consider a restriction on the investment decision process, that is an investor is forbidden to go short on the risky asset. This translates in a restriction on the vector of portfolios weights which now must be non-negative and sum to one. Table D.1 shows the results for a mean-variance investor with a moderate level of risk aversion equal to $\gamma = 5$. Similar to the findings in the main paper, our DRS predictive modeling outpeforms the competing strategies for the vast majority of industries as well as the aggregate stock market. Interestingly, when short sales are in place, the overall performance of both model combination and penalised regressions tend to deteriorate. Table D.2 extends the results to a representative investor with more general Constant Relative Risk Aversion (CRRA) preferences. Again, we consider an investor with moderate risk aversion equal to $\gamma = 5$. Similar to the unconstrained case, our DRS model outperforms by and large the competing strategies, both for individual industries and for the aggregate stock market. As a whole, Tables D.1-D.2 show that the economic significance of our large-scale dynamic predictive regression model persists after restricting the optimal portfolio allocation to non-negative weights.

# E Bootstrap test for returns predictability across economic states

We implement the bootstrap approach proposed by Gargano et al. (2019) to test if the differences in the forecasting performance over the business cycle are statistically significant. Let $\Delta_j = e^2_{HA,j} - e^2_{i,j}$, with $e^2_{HA,j}$ and $e_{i,j}$ the squared out-of-sample prediction errors for the simple recursive mean and a given model $i$, respectively. The subscript $j = 0, 1$ refers to the expansion (zero) vs recession (one) period. The null hypothesis is that of equal predictability across economic states, and is imposed by substracting the mean to the sample predictions in both states;

---

**Procedure**

[31] An example of time series forecasting using LSTM can be found here https://uk.mathworks.com/help/deeplearning/ug/timeseries-forecasting-using-deep-learning.html

Compute $\hat{\Delta}_j = \Delta_j - \hat{\mu}(\Delta_j)$, with $\hat{\mu}(\Delta_j)$ the sample average prediction error in the state $j$.

**for** *all bootstrap iterations* $b = 1, ..., B$

- Draw a random sample $\hat{\Delta}_0^b, \hat{\Delta}_1^b$ from $\hat{\Delta}_0, \hat{\Delta}_1$.
- Compute the bootstrap statistic $J^b = \mu\left(\hat{\Delta}_0^b\right) - \mu\left(\hat{\Delta}_1^b\right)$ with $\mu\left(\hat{\Delta}_j^b\right)$ the sample average of $\hat{\Delta}_j^b$.

**end**

Compute the p-value as $p = \frac{1}{B}\sum_{b=1}^{B} \mathbb{I}\left[J > J^b\right]$, where $J = \mu\left(\hat{\Delta}_0^b\right) - \mu\left(\hat{\Delta}_1^b\right)$ is based on the original data.

# F Simulation Study

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

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

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

We simulate data by the following data generating process:

$$y = -2z_1 + 3z_2 + 5z_3 + \epsilon, \quad \epsilon \sim N(0, 0.01), \tag{F.1a}$$

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

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

$$z_3 = \nu_3, \quad \nu_3 \sim N(0, 0.01), \tag{F.1d}$$

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

One comment is in order. This simulation study explicitly explores two key characteristics above and beyond the dimensionality of the set of predictors, that are (1) correlated regressors and (2) omitted variables. Both these approaches are explicitly addressed by our DRS method through a time-varying synthesis parameters and aggregate bias. In this respect, by enlarging the set of regressors in the simulation would only increase complexity without adding anythin to the main message.

We generate $N = 510$ samples, use the first ten to fit the initial model, and forecast 500 data points. We consider eight different strategies that are also considered in the empirical application. A more detailed description of these models will be provided in Section 3 below. The first three models are subset of the possible models with either $\{z_1\}$, $\{z_2\}$, or $\{z_1, z_2\}$ are considered as regressors and the models are estimated using ordinary least squares. We also consider a lasso regression and a dynamic factor model.

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

Finally, we compare the seven competing strategies against a simplified, namely time invariant, version of our proposed "decouple-recouple" predictive strategy. Here, the latent states are, as with the two forecast combination schemes, the forecasts from the two linear regressions with $\{z_1\}$ or

$\{z_2\}$, but the synthesis function is time invariant instead of the dynamic specification. This yields a simpler setup for DRS by removing the dynamics from the equation and following suit with the model and strategies compared. Here, the synthesis parameters are estimated using a simple Bayesian linear regression with non-informative priors (Jeffreys' prior).

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

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

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

## Table 1: **Out-of-sample predictive $R^2_{oos}$**

This table reports the out-of-sample $R^2_{oos}$ of a forecasting model where the dependent variable is the industry stock returns in excess of the risk free rate, and the dependent variables consist of a large set of financial ratios and aggregate macroeconomic variables. To compute the out-of-sample $R^2_{oos}$, we compare the forecasts obtained from each methodology to the prediction based on the historical mean. The table reports the $R^2_{oos}$ values, as well as the p-value for the null hypothesis $R^2_{oos} \leq 0$ calculated as in Clark and West (2007). Notice, we report the p-value only when the out-of-sample $R^2_{oos}$ is non-negative, which means a given model outperform a forecast based on the recursive sample average. The out-of-sample prediction errors are obtained by recursive forecasts starting at February 2002. The sample period is from 1970:01-2018:12, monthly.

### **Panel A:** Out-of-sample $R^2_{oos}$

| | Industries | | | | | | | | | | Mkt | EW Port |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Durables | NonDurables | Manuf | Energy | HiTech | Health | Other | Shops | Telecomm | Utils | | |
| OLS | -23.715 | -10.730 | -19.122 | -16.776 | -35.491 | -16.381 | -15.164 | -11.315 | -16.905 | -74.880 | -21.722 | -23.836 |
| Lasso | -1.837 | -1.126 | -0.389 | -2.987 | -1.434 | -0.599 | -1.097 | -0.769 | -1.141 | -0.840 | -1.257 | -1.225 |
| Ridge | -2.183 | -1.644 | -1.361 | -2.059 | -3.014 | -1.072 | -1.574 | -0.698 | -1.532 | -1.098 | -0.584 | -1.529 |
| E-net | -2.677 | -0.767 | -0.802 | -1.865 | -2.198 | -1.236 | -0.861 | -1.060 | -0.687 | -1.138 | -0.787 | -1.280 |
| Random Forest | -0.143 | -0.188 | -0.183 | -0.205 | -0.223 | -0.193 | -0.109 | -0.207 | -0.168 | -0.298 | -0.169 | -0.189 |
| Neural Net | -0.962 | -0.815 | -0.634 | -0.906 | -0.780 | -0.986 | -0.380 | -0.868 | -0.745 | -0.564 | -0.693 | -0.758 |
| Factor Model | -0.600 | -0.136 | -0.538 | -0.218 | -0.120 | -0.157 | -0.615 | -0.237 | -0.202 | -0.150 | -0.403 | -0.307 |
| Equal-weight | -0.089 | -0.073 | -0.039 | -0.070 | -0.122 | -0.115 | 0.006 | -0.116 | -0.044 | 0.000 | 0.012 | -0.059 |
| BMA | -0.198 | -0.025 | -0.075 | -0.056 | -0.105 | -0.068 | -0.144 | -0.157 | -0.089 | 0.000 | -0.083 | -0.091 |
| Median | -0.087 | -0.039 | 0.008 | -0.010 | -0.084 | -0.086 | 0.009 | -0.075 | -0.037 | 0.012 | 0.009 | -0.035 |
| Trimmed Mean | -0.065 | -0.048 | 0.000 | -0.039 | -0.080 | -0.089 | 0.010 | -0.100 | -0.025 | -0.011 | 0.013 | -0.039 |
| Macro | -0.182 | -0.100 | -0.068 | -0.059 | -0.151 | -0.084 | -0.144 | -0.102 | -0.172 | -0.020 | -0.199 | -0.116 |
| DRS | 0.012 | 0.013 | 0.019 | 0.018 | 0.006 | 0.013 | 0.027 | 0.016 | 0.039 | 0.027 | 0.045 | 0.021 |

### **Panel B:** P-values

| | Industries | | | | | | | | | | Mkt | EW Port |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Durables | NonDurables | Manuf | Energy | HiTech | Health | Other | Shops | Telecomm | Utils | | |
| OLS | - | - | - | - | - | - | - | - | - | - | - | - |
| Lasso | - | - | - | - | - | - | - | - | - | - | - | - |
| Ridge | - | - | - | - | - | - | - | - | - | - | - | - |
| E-net | - | - | - | - | - | - | - | - | - | - | - | - |
| Random Forest | - | - | - | - | - | - | - | - | - | - | - | - |
| Neural Net | - | - | - | - | - | - | - | - | - | - | - | - |
| Factor Model | - | - | - | - | - | - | - | - | - | - | - | - |
| Equal-weight | - | - | - | - | - | - | 0.073 | - | - | - | 0.056 | - |
| BMA | - | - | - | - | - | - | - | - | - | - | - | - |
| Median | - | - | 0.063 | - | - | - | 0.078 | - | - | 0.090 | 0.121 | - |
| Trimmed mean | - | - | - | - | - | - | 0.064 | - | - | - | 0.067 | - |
| Macro | - | - | - | - | - | - | - | - | - | - | - | - |
| DRS | 0.064 | 0.021 | 0.005 | 0.017 | 0.092 | 0.056 | 0.008 | 0.020 | 0.000 | 0.014 | 0.002 | 0.027 |

Table 2: **Industry returns predictability over the business cycle**

This table reports the out-of-sample $R^2_{oos}$ of a forecasting regression across different economic cycles. To test if industry returns predictability varies over the business cycle, we split the data into recession and expansionary periods using the NBER dates of peaks and troughs. To compute the out-of-sample $R^2_{oos}$, we compare the forecasts obtained from each methodology to the prediction based on the historical mean in each economic state. The table reports the $R^2_{oos}$ values in each regime as well as the p-value for the null hypothesis of equal probability in recession vs expansion calculated as in Gargano et al. (2019). Notice, we report in bold the positive $R^2_{oos}$ for which the null hypothesis is rejected at the 5% conventional threshold. The out-of-sample prediction errors are obtained by recursive forecasts starting at February 2002. The sample period is from 1970:01-2018:12, monthly.

### **Panel A:** Out-of-sample $R^2_{oos}$ in expansions

| | Industries | | | | | | | | | | Mkt | EW Port |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Durables | NonDurables | Manuf | Energy | HiTech | Health | Other | Shops | Telecomm | Utils | | |
| OLS | -39.904 | -13.822 | -25.472 | -20.574 | -47.129 | -18.140 | -24.728 | -12.132 | -22.210 | -88.978 | -23.552 | -30.604 |
| Lasso | -2.927 | -1.283 | -0.447 | -3.351 | -1.656 | -0.715 | -1.249 | -0.991 | -1.574 | -1.094 | -1.036 | -1.484 |
| Ridge | -3.145 | -1.653 | -0.572 | -2.326 | -3.134 | -1.181 | -2.211 | -0.732 | -1.302 | -1.501 | -0.477 | -1.658 |
| E-net | -2.718 | -0.766 | -0.714 | -1.723 | -2.424 | -1.255 | -1.266 | -1.065 | -0.994 | -1.531 | -0.911 | -1.397 |
| Random Forest | -0.221 | -0.271 | -0.380 | -0.240 | -0.312 | -0.241 | -0.267 | -0.230 | -0.278 | -0.485 | -0.255 | -0.289 |
| Neural Net | -1.255 | -1.011 | -0.924 | -1.053 | -1.112 | -1.038 | -0.640 | -1.074 | -0.785 | -0.894 | -1.047 | -0.985 |
| Factor Model | -0.511 | -0.121 | -0.570 | -0.205 | -0.070 | -0.119 | -0.527 | -0.376 | -0.201 | -0.092 | -0.545 | -0.303 |
| Equal-weight | -0.102 | -0.104 | -0.126 | -0.075 | -0.205 | -0.112 | -0.051 | -0.109 | -0.044 | -0.056 | -0.032 | -0.092 |
| BMA | -0.320 | -0.146 | -0.106 | -0.043 | -0.185 | -0.065 | -0.182 | -0.135 | -0.086 | -0.056 | -0.132 | -0.132 |
| Median | -0.109 | -0.082 | -0.070 | -0.032 | -0.174 | -0.071 | -0.020 | -0.049 | -0.049 | -0.109 | -0.021 | -0.072 |
| Trimmed mean | -0.084 | -0.097 | -0.099 | -0.052 | -0.165 | -0.083 | -0.036 | -0.086 | -0.029 | -0.126 | 0.002 | -0.078 |
| Macro | -0.312 | -0.143 | -0.127 | -0.044 | -0.150 | -0.071 | -0.181 | -0.114 | -0.178 | -0.030 | -0.188 | -0.140 |
| DRS | 0.014 | 0.007 | 0.014 | 0.009 | 0.007 | 0.015 | 0.035 | 0.017 | 0.042 | 0.010 | 0.029 | 0.018 |

### **Panel B:** Out-of-sample $R^2_{oos}$ in recessions

| | Industries | | | | | | | | | | Mkt | EW Port |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Durables | NonDurables | Manuf | Energy | HiTech | Health | Other | Shops | Telecomm | Utils | | |
| OLS | -2.686 | -2.844 | -8.523 | -3.913 | -8.210 | -9.759 | -1.592 | -9.055 | -4.876 | -44.779 | -18.380 | -10.420 |
| Lasso | -0.422 | -0.727 | -0.293 | -1.752 | -0.912 | -0.163 | -0.881 | -0.155 | -0.159 | -0.297 | -1.548 | -0.664 |
| Ridge | -0.933 | -1.620 | -2.677 | -1.153 | -2.733 | -0.663 | -0.672 | -0.602 | -2.054 | -0.239 | -0.673 | -1.275 |
| E-net | -2.624 | -0.771 | -0.949 | -2.349 | -1.668 | -1.165 | -0.285 | -1.045 | **0.009** | -0.297 | -0.469 | -1.056 |
| Random Forest | -0.041 | **0.024** | **0.015** | -0.090 | -0.013 | -0.011 | **0.115** | -0.144 | **0.008** | **0.085** | -0.020 | -0.007 |
| Neural Net | -0.583 | -0.315 | -0.150 | -0.411 | -0.003 | -0.791 | -0.010 | -0.300 | -0.654 | **0.014** | -0.013 | -0.292 |
| Factor Model | -0.716 | -0.175 | -0.485 | -0.263 | -0.237 | -0.299 | -0.738 | 0.015 | -0.204 | -0.274 | -0.052 | -0.312 |
| Equal-weight | -0.072 | 0.006 | **0.023** | -0.050 | **0.073** | -0.125 | **0.110** | -0.133 | -0.044 | 0.012 | **0.034** | -0.015 |
| BMA | -0.039 | **0.028** | -0.022 | -0.099 | **0.081** | -0.078 | -0.090 | -0.216 | -0.095 | 0.012 | -0.228 | -0.068 |
| Median | -0.058 | **0.034** | **0.020** | **0.066** | **0.013** | -0.141 | **0.146** | -0.148 | -0.009 | **0.027** | **0.012** | -0.003 |
| Trimmed mean | -0.042 | **0.075** | **0.017** | 0.006 | **0.012** | -0.109 | **0.124** | -0.139 | -0.015 | **0.023** | **0.021** | -0.002 |
| Macro | -0.013 | **0.011** | **0.030** | -0.110 | -0.153 | -0.130 | -0.090 | -0.066 | -0.159 | 0.002 | -0.119 | -0.073 |
| DRS | **0.010** | **0.070** | **0.031** | **0.051** | 0.002 | **0.027** | **0.059** | **0.036** | 0.033 | **0.062** | **0.068** | **0.041** |

Table 3: **Out-of-sample economic significance: Mean-variance utility**

This table reports the annualized certainty equivalent values (annualised, in %) for portfolio decisions based on the out-of-sample forecasts of industry returns for an investor with *mean-variance* utility and a coefficient of risk aversion $\gamma = 5$. The table reports three asset allocation exercises. The first case is a single investment in a given industry; the second a multiple investment in different industries; the third is a single investment in the aggregate stock market. The asset allocation decision is made at each time $t$ and is based on the predictions from each of the alternative models outlined in the main text. The models are benchmarked against the signal obtained from a recursive sample mean estimate. The predictions are obtained starting in February 2002, and the sample period is from 1970:01-2018:12, monthly. Statistical significance is based on a one-sided Diebold and Mariano (1995) test as extended by Harvey et al. (1997) to account for correlation. We report only the p-values when the difference in the certainty equivalent is positive, i.e., a model outperforms the sample mean forecast.

**Panel A:** Certainty equivalent

| | Industries | | | | | | | | | | Mkt | EW Port |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Durables | NonDurables | Manuf | Energy | HiTech | Health | Other | Shops | Telecomm | Utils | | |
| OLS | -1.128 | -0.291 | -0.478 | -1.734 | -1.182 | -0.808 | -0.425 | -0.486 | -0.753 | -1.057 | -0.480 | -0.409 |
| Lasso | -0.687 | -0.757 | -0.100 | -1.176 | -0.474 | -0.294 | -0.238 | -0.289 | 0.247 | 0.083 | -0.128 | -0.567 |
| Ridge | -0.019 | -0.225 | -0.416 | -1.332 | -0.146 | 0.099 | -0.562 | -0.495 | -0.169 | -0.786 | -0.306 | -0.455 |
| E-net | -0.453 | -0.644 | -0.465 | -0.145 | -0.539 | 0.124 | -0.160 | 0.002 | -0.599 | 0.293 | 0.117 | -0.490 |
| Random Forest | 0.580 | -0.037 | 0.276 | -0.260 | 0.441 | 0.176 | 0.519 | -0.265 | 0.556 | -0.276 | 0.379 | 0.190 |
| Neural Net | -0.018 | -0.069 | 0.024 | 0.095 | -0.121 | -0.093 | 0.036 | -0.068 | 0.011 | 0.040 | 0.026 | -0.012 |
| Factor Model | 0.413 | 0.285 | 0.195 | 0.098 | 0.016 | -0.050 | 0.338 | 0.286 | 0.446 | 0.147 | 0.357 | 0.001 |
| Equal-weight | 0.114 | 0.081 | 0.435 | 0.200 | -0.207 | -0.115 | 0.158 | -0.046 | 0.309 | -0.242 | 0.312 | 0.091 |
| BMA | -0.161 | 0.437 | 0.263 | -0.114 | 0.055 | -0.177 | -0.254 | -0.279 | 0.181 | -0.242 | -0.073 | -0.033 |
| Median | 0.042 | 0.040 | 0.624 | 0.435 | -0.001 | -0.012 | 0.574 | 0.113 | 0.278 | 0.248 | 0.197 | 0.231 |
| Trimmed mean | 0.191 | 0.090 | 0.561 | 0.294 | -0.010 | -0.068 | 0.701 | -0.029 | 0.325 | 0.176 | 0.276 | 0.228 |
| Macro | 0.769 | 0.021 | 0.261 | -0.130 | -0.096 | -0.233 | -0.254 | 0.161 | -0.333 | -0.033 | 0.109 | 0.022 |
| DRS | 1.008 | 1.215 | 1.102 | 1.090 | 0.405 | 0.114 | 0.592 | 1.007 | 1.137 | 1.120 | 0.832 | 0.782 |

**Panel B:** P-values

| | Industries | | | | | | | | | | Mkt | EW Port |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Durables | NonDurables | Manuf | Energy | HiTech | Health | Other | Shops | Telecomm | Utils | | |
| OLS | - | - | - | - | - | - | - | - | - | - | - | - |
| Lasso | - | - | - | - | - | - | - | - | 0.265 | 0.035 | - | - |
| Ridge | - | - | - | - | - | - | - | - | - | - | - | - |
| E-net | - | - | - | - | - | 0.487 | - | - | - | 0.186 | 0.150 | - |
| Random Forest | 0.098 | - | 0.320 | - | 0.071 | 0.269 | 0.045 | - | 0.078 | - | 0.098 | 0.231 |
| Neural Net | 0.068 | 0.371 | 0.233 | 0.366 | 0.236 | - | 0.441 | - | 0.175 | 0.279 | 0.189 | - |
| Factor | 0.120 | 0.403 | 0.389 | 0.223 | 0.383 | - | 0.188 | 0.365 | 0.093 | 0.234 | 0.146 | 0.254 |
| Equal-weight | 0.376 | 0.348 | 0.079 | 0.288 | - | - | 0.316 | - | 0.118 | - | 0.173 | 0.242 |
| BMA | - | 0.091 | 0.125 | - | 0.429 | - | - | - | 0.254 | - | - | - |
| Median | 0.460 | 0.418 | 0.026 | 0.108 | - | - | 0.040 | 0.270 | 0.133 | 0.171 | 0.258 | 0.209 |
| Trimmed mean | 0.315 | 0.330 | 0.043 | 0.190 | - | - | 0.040 | - | 0.101 | 0.243 | 0.207 | 0.184 |
| Macro | 0.039 | 0.466 | 0.166 | - | - | - | - | 0.193 | - | - | 0.404 | 0.254 |
| DRS | 0.023 | 0.048 | 0.034 | 0.011 | 0.098 | 0.113 | 0.035 | 0.014 | 0.045 | 0.027 | 0.033 | 0.038 |

Table 4: **Out-of-sample economic significance: Power utility**

This table reports the annualized certainty equivalent values (annualised, in %) for portfolio decisions based on the out-of-sample forecasts of industry returns for an investor with *power utility* and a coefficient of risk aversion $\gamma = 5$. The table reports three asset allocation exercises. The first case is a single investment in a given industry; the second a multiple investment in different industries; the third is a single investment in the aggregate stock market. The asset allocation decision is made at each time $t$ and is based on the predictions from each of the alternative models outlined in the main text. The models are benchmarked against the signal obtained from a recursive sample mean estimate. The predictions are obtained starting in February 2002, and the sample period is from 1970:01-2018:12, monthly. Statistical significance is based on a one-sided Diebold and Mariano (1995) test as extended by Harvey et al. (1997) to account for correlation. We report only the p-values when the difference in the certainty equivalent is positive, i.e., a model outperforms the sample mean forecast.

**Panel A:** Certainty equivalent

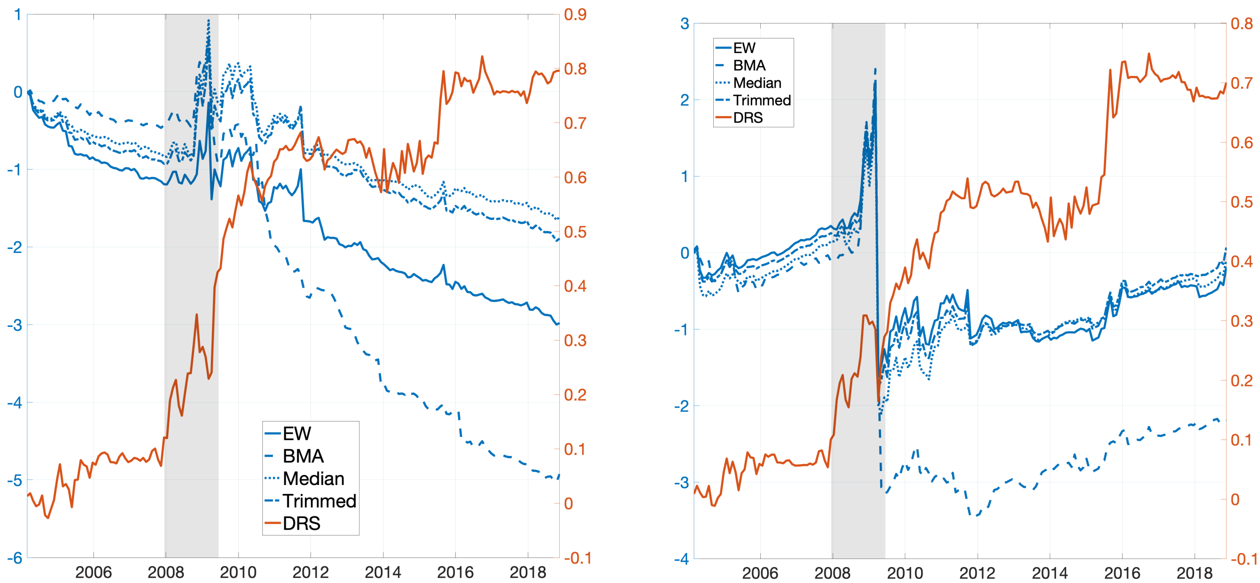
| | Industries | | | | | | | | | | Mkt | EW Port |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Durables | NonDurables | Manuf | Energy | HiTech | Health | Other | Shops | Telecomm | Utils | | |
| OLS | -2.004 | -1.745 | -1.207 | -3.308 | -4.770 | -3.643 | -2.530 | -2.556 | -3.771 | -6.575 | -3.833 | -3.267 |
| Lasso | -1.341 | -0.580 | -0.685 | -2.581 | -1.192 | -0.388 | -0.083 | -0.895 | -0.390 | -1.773 | -1.043 | -0.996 |
| Ridge | -1.489 | -2.314 | -1.940 | -1.510 | -2.028 | -0.907 | -0.683 | -0.855 | -1.812 | -1.441 | -0.389 | -1.397 |
| E-net | -0.911 | -0.994 | -0.971 | -1.861 | -2.282 | -0.880 | -0.668 | -1.091 | 0.569 | -1.594 | 0.207 | -0.952 |
| Random Forest | -0.146 | -0.222 | -0.082 | 0.123 | -0.070 | -0.158 | 0.036 | -0.185 | -0.075 | -0.047 | -0.044 | -0.079 |
| Neural Net | -0.776 | -0.457 | -0.154 | -0.913 | 0.120 | -0.814 | 0.400 | -0.205 | 0.139 | 0.178 | 0.209 | -0.207 |
| Factor Model | 0.239 | -0.244 | -0.137 | 0.037 | 0.331 | 0.295 | -0.231 | 0.537 | 0.244 | 0.660 | 0.035 | 0.161 |
| Equal-weight | 0.261 | -0.010 | 0.510 | 0.408 | 0.187 | -0.131 | 0.515 | 0.335 | 0.550 | -1.036 | 0.338 | 0.175 |
| BMA | -0.962 | 0.244 | 0.440 | -0.111 | 0.078 | -0.233 | -0.258 | 0.029 | 0.128 | -1.036 | -0.031 | -0.156 |
| Median | 0.099 | -0.065 | 0.676 | 0.646 | 0.440 | 0.075 | 0.579 | 0.565 | 0.493 | 0.146 | 0.335 | 0.381 |
| Trimmed mean | 0.298 | -0.017 | 0.683 | 0.508 | 0.463 | -0.035 | 0.609 | 0.370 | 0.544 | 0.027 | 0.345 | 0.345 |
| Macro | -0.143 | -0.146 | 0.456 | -0.133 | 0.081 | -0.262 | -0.260 | 0.194 | -0.725 | -0.033 | 0.143 | -0.075 |
| DRS | 0.854 | 0.458 | 0.809 | 0.937 | 0.779 | 0.490 | 0.915 | 0.553 | 1.255 | 0.720 | 0.872 | 0.785 |

**Panel B:** P-values

| | Industries | | | | | | | | | | Mkt | EW Port |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Durables | NonDurables | Manuf | Energy | HiTech | Health | Other | Shops | Telecomm | Utils | | |
| OLS | - | - | - | - | - | - | - | - | - | - | - | - |
| Lasso | - | - | - | - | - | - | - | - | - | - | - | - |
| Ridge | - | - | - | - | - | - | - | - | - | - | - | - |
| E-net | - | - | - | - | - | - | - | - | 0.200 | - | 0.183 | - |
| Random Forest | - | - | - | 0.300 | - | - | 0.139 | - | - | - | - | - |
| Neural Net | - | - | - | - | 0.451 | - | 0.103 | - | 0.234 | 0.320 | 0.386 | - |
| Factor Model | 0.215 | - | - | 0.389 | 0.092 | 0.153 | - | 0.081 | 0.288 | 0.062 | 0.337 | 0.202 |
| Equal-weight | 0.314 | - | 0.087 | 0.080 | 0.314 | - | 0.018 | 0.125 | 0.067 | - | 0.222 | 0.153 |
| BMA | - | 0.257 | 0.111 | - | 0.433 | - | - | 0.464 | 0.374 | - | - | - |
| Median | 0.424 | - | 0.049 | 0.069 | 0.110 | 0.384 | 0.014 | 0.062 | 0.075 | 0.267 | 0.232 | 0.167 |
| Trimmed mean | 0.282 | - | 0.033 | 0.076 | 0.105 | - | 0.018 | 0.080 | 0.060 | 0.456 | 0.224 | 0.148 |
| Macro | - | - | 0.168 | - | 0.434 | - | - | 0.251 | - | - | 0.407 | - |
| DRS | 0.021 | 0.012 | 0.018 | 0.011 | 0.032 | 0.091 | 0.013 | 0.032 | 0.011 | 0.028 | 0.032 | 0.027 |

Figure 1: **Cumulative sum of squared forecasting error differentials**

This figure reports the differences between the cumulative square prediction error for the historical average benchmark and for the forecasts based on some of the most successful competing predictive strategies. The predictions are obtained starting in February 2004, and the sample period is from 1970:01-2018:12, monthly.

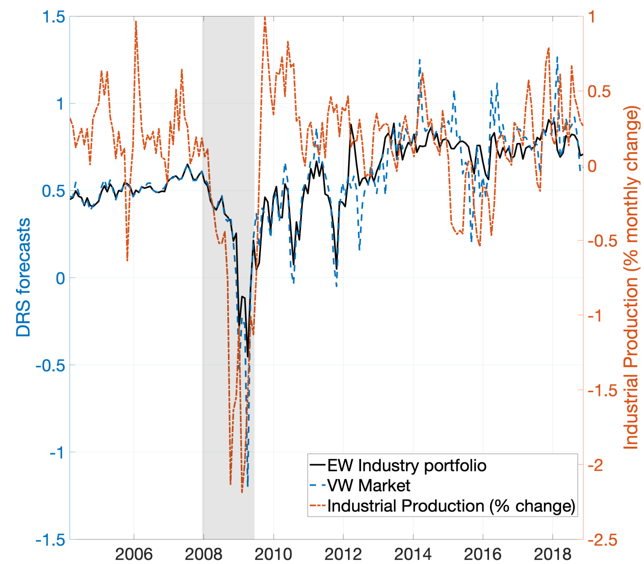



(a) Industries

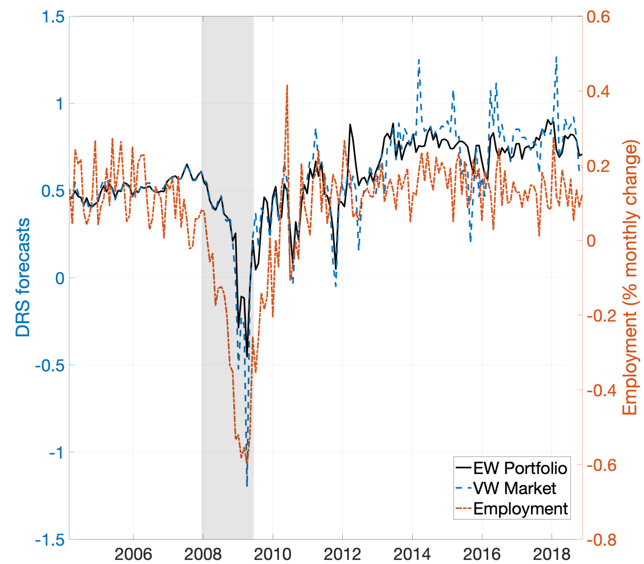
(b) Market

Figure 2: **DRS predictions vs macroeconomic variables**

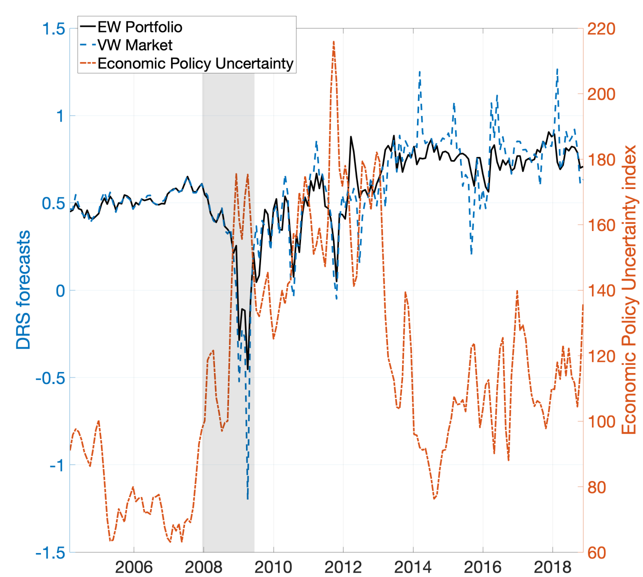
This figure reports the average DRS forecast across industries and the aggregate value-weighted market portfolio against a set of macroeconomic variables such as industrial production (top-left panel), employment (top-right panel), a measure of economic policy uncertainty (bottom-left panel), and the VIX (bottom-right panel). The predictions are obtained starting in February 2004, and the sample period is from 1970:01-2018:12, monthly.



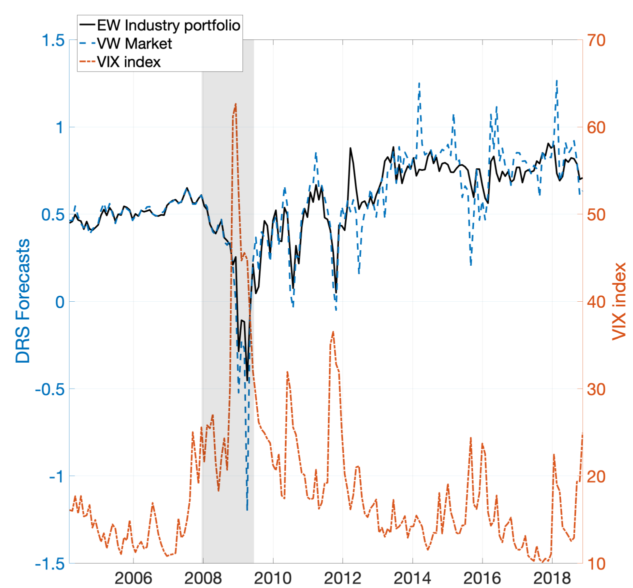

(a) Industrial Production




(b) Employment



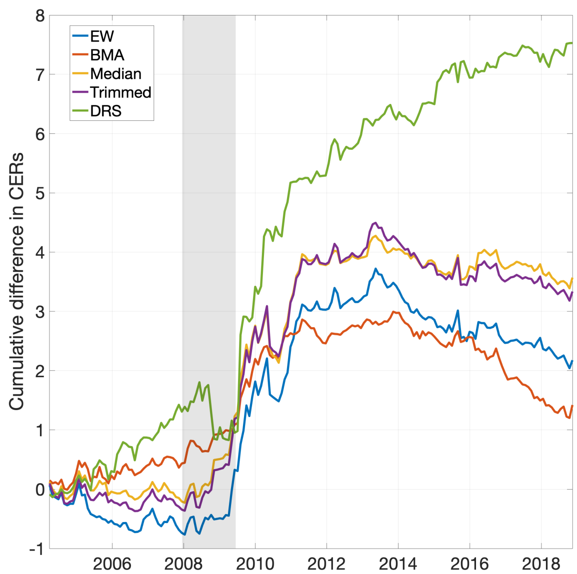

(c) Economic Policy Uncertainty



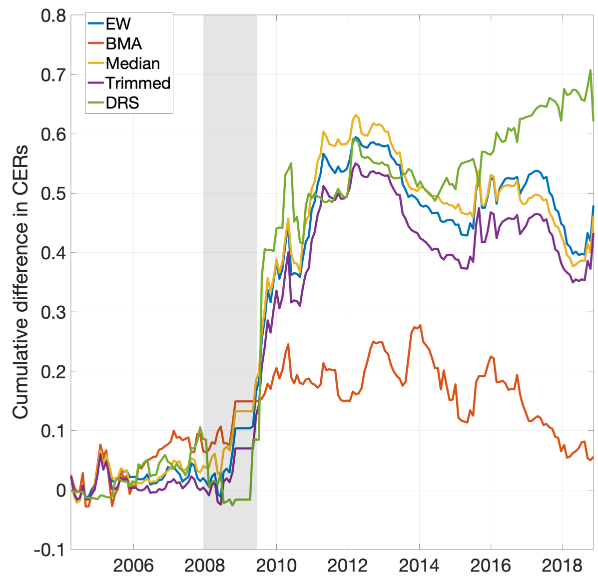

(d) VIX

Figure 3: **Cumulative sum of certainty equivalent returns differentials**

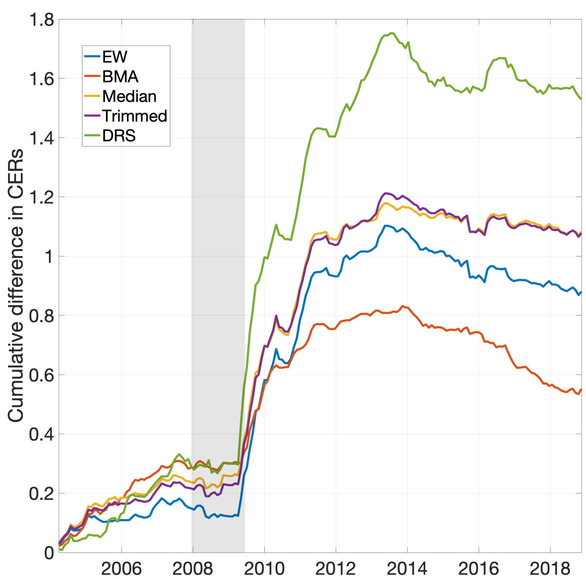
This figure reports the cumulative average Certainty Equivalent Return (CER) of a given predictive strategy against the historical average. The top panels show the results for a mean-variance investor whereas the bottom panels show the results for an investor with Constant Relative Risk Aversion (CRRA) utility function. The left panels show the results for the equal-weight portfolio of industries, whereas the right panel show the results for the aggregate value-weighted market portfolio. The sample period is from 1970:01-2018:12, monthly.



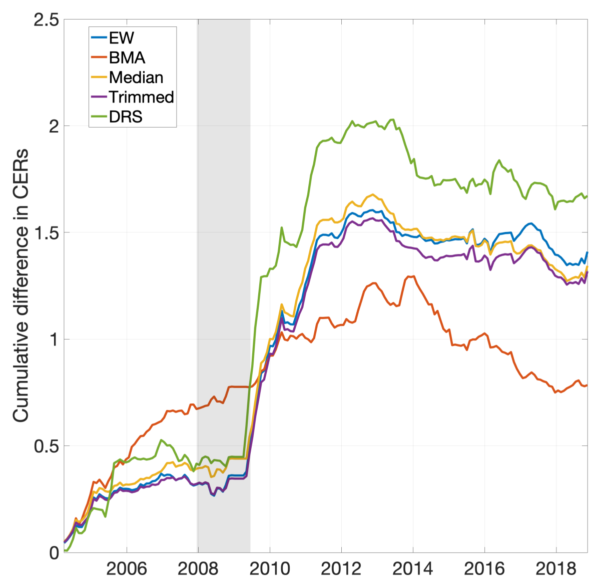

(a) MV utility - Industries




(b) MV utility - Market



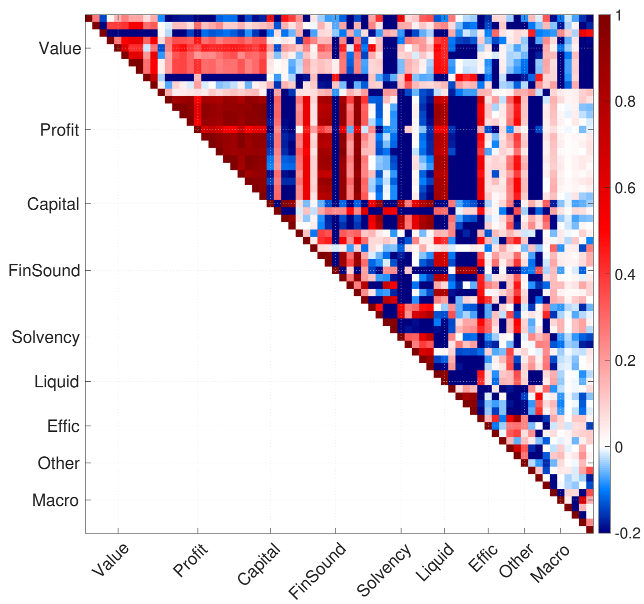

(c) CRRA utility - Industries



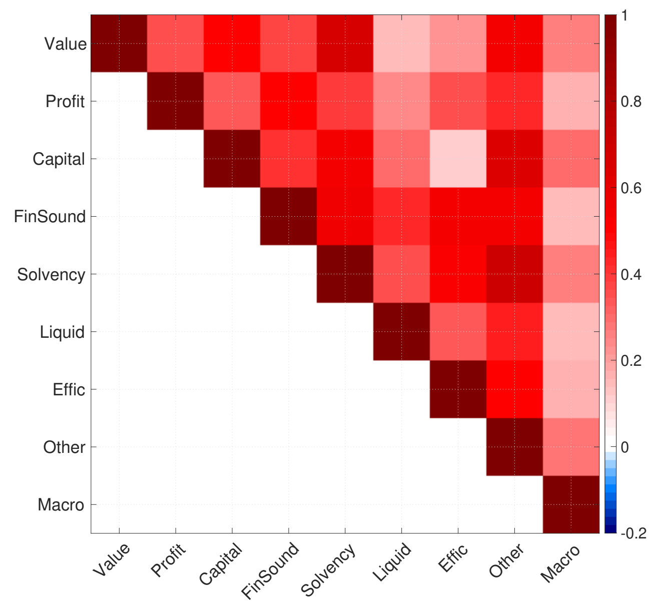

(d) CRRA utility - Market

Figure 4: **Correlation among financial ratios and sub-model forecasts**

This figure reports the correlation among the set of predictors for the average industry (left panel) and the correlation across the forecasts from each sub-regression model (right panel). The red (blue) areas represent positive (negative) correlation. The sample period is from 1970:01-2018:12, monthly.




(a) Financial ratios




(b) Sub-models

Figure 5: **Covariance between sub-model predictions: Aggregate market**

This figure reports the model-implied covariance of the sub-models in the DRS strategy, one for each group of financial ratios, *i.e.*, $Cov(\theta_{it}, \theta_{jt})$. The figure reports the results for the aggregate stock market. We report the covariances for four different time period, that are August 2006, November 2008, January 2011 and December 2015.

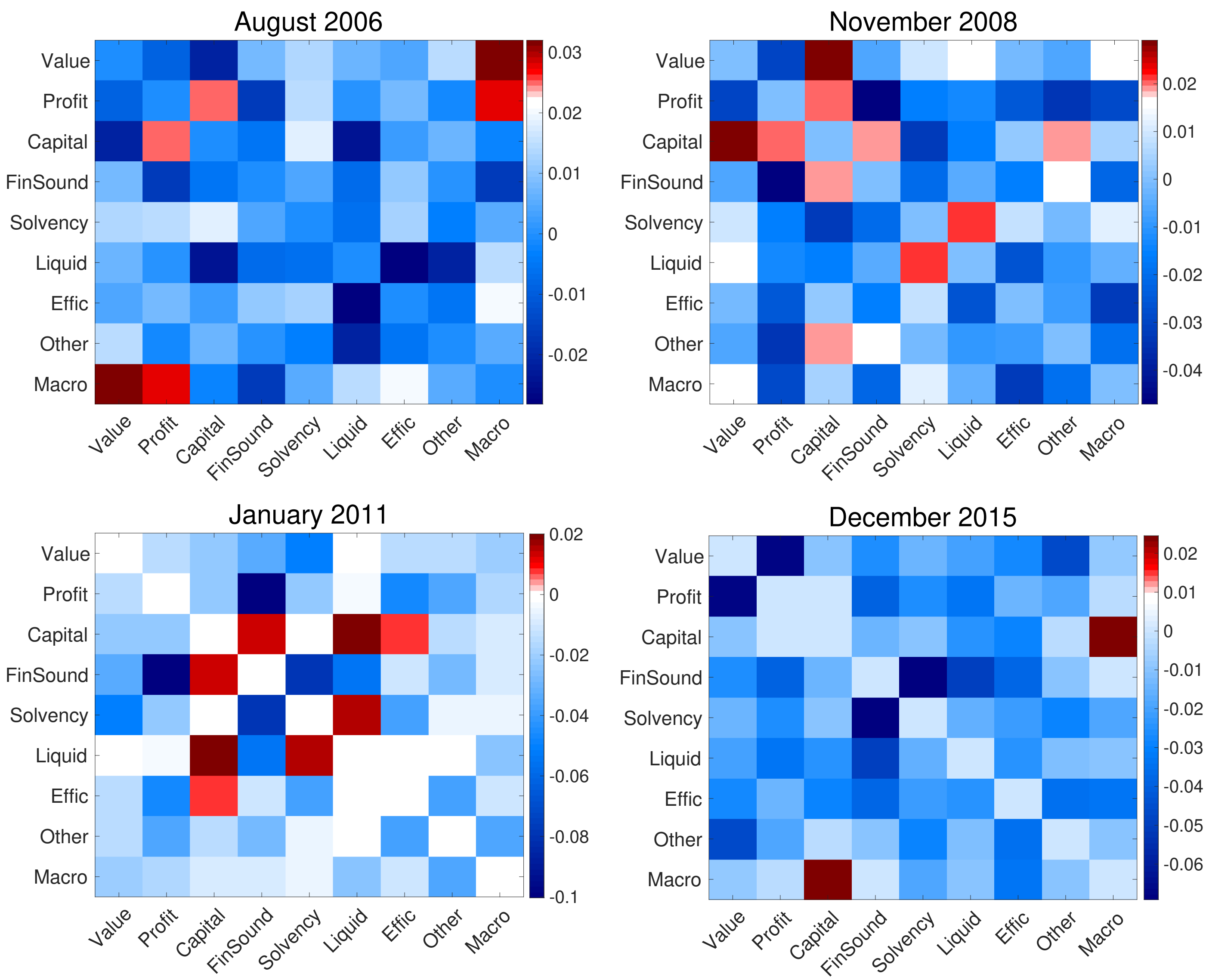


Figure 6: **Covariance between sub-model predictions: Average industry**

This figure reports the model-implied covariance of the sub-models in the DRS strategy, one for each group of financial ratios, *i.e.*, $Cov(\theta_{it}, \theta_{jt})$. The figure reports the results for the average industry portfolio. We report the covariances for four different time period, that are August 2006, November 2008, January 2011 and December 2015.

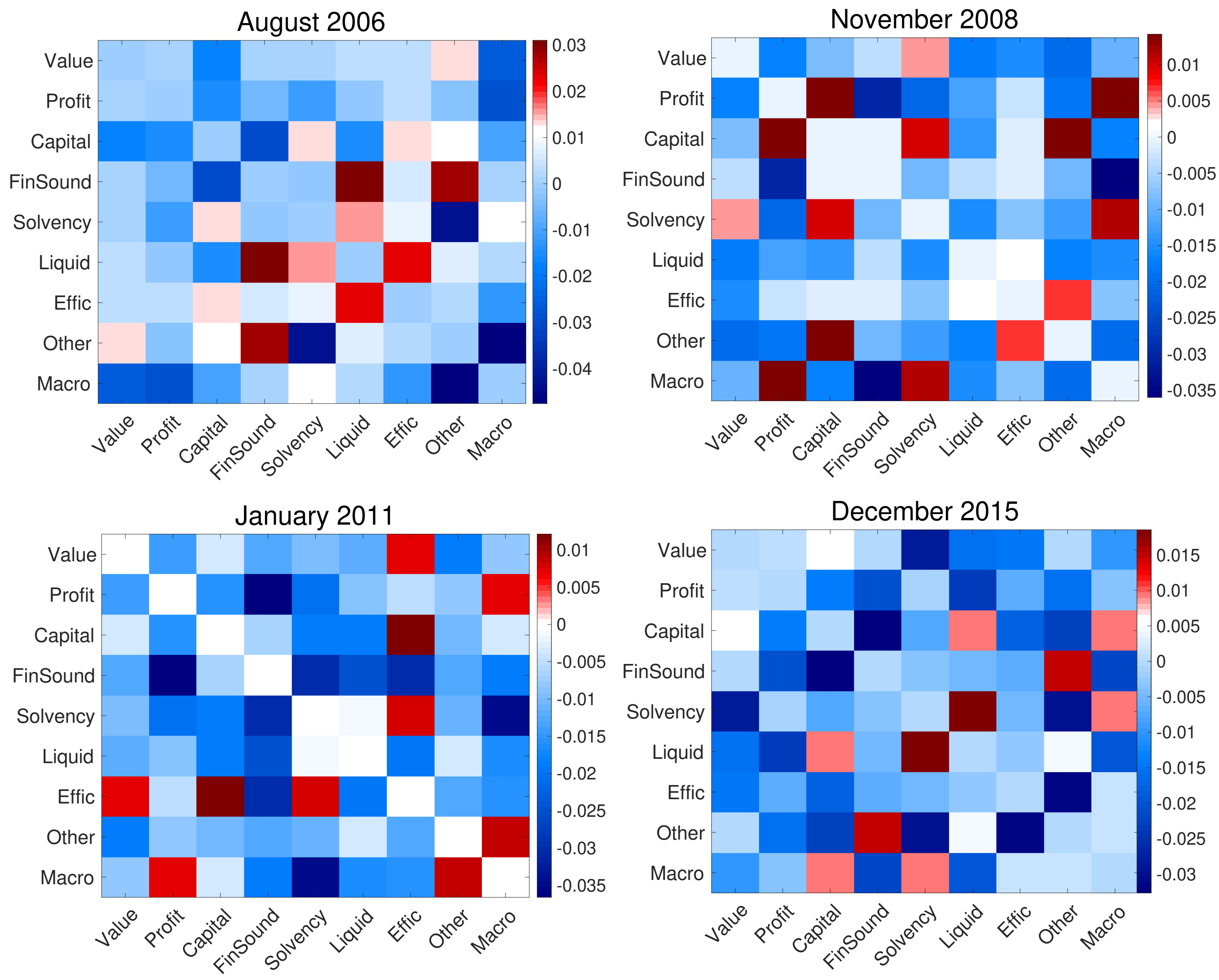
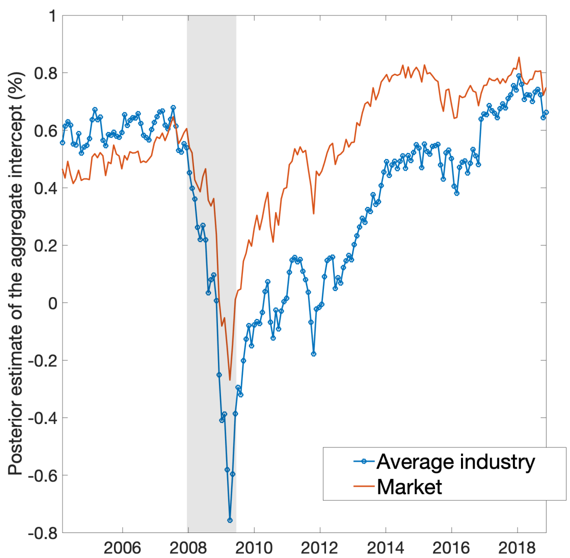
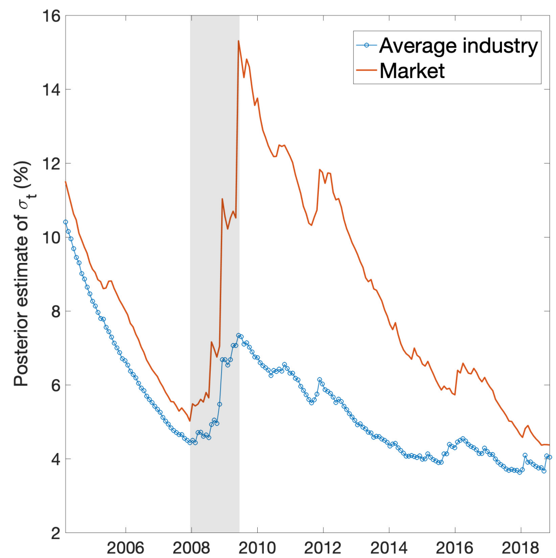


Figure 7: **Aggregate bias and observation uncertainty**

This figure reports the posterior estimates of the aggregate intercept, $\theta_0$, (left panel) and the standard deviation of the measurement error, $\sigma_t$. We report the results for both the average industry (blue line with markers) and for the aggregate stock market (orange solid line). The predictions are obtained starting in January 2001, and the sample period is from 1970:01-2018:12, monthly.




(a) Bias




(b) Observation uncertainty

Figure 8: **Marginal effects of individual forecasts: Average industry**

This figure reports the posterior estimates of the latent parameters $\theta_{jt}$, $j = 1, \ldots, J$. We report the estimates for the average industry from January 2004 to December 2018. Top panels report the estimates for capital and valuation groups (left panel) and profitability, financial soundness and solvency (right panel). Bottom panels report the estimates for the liquidity and efficiency groups (left panel) and the "other" group of financial ratios and aggregate macroeconomic variables.

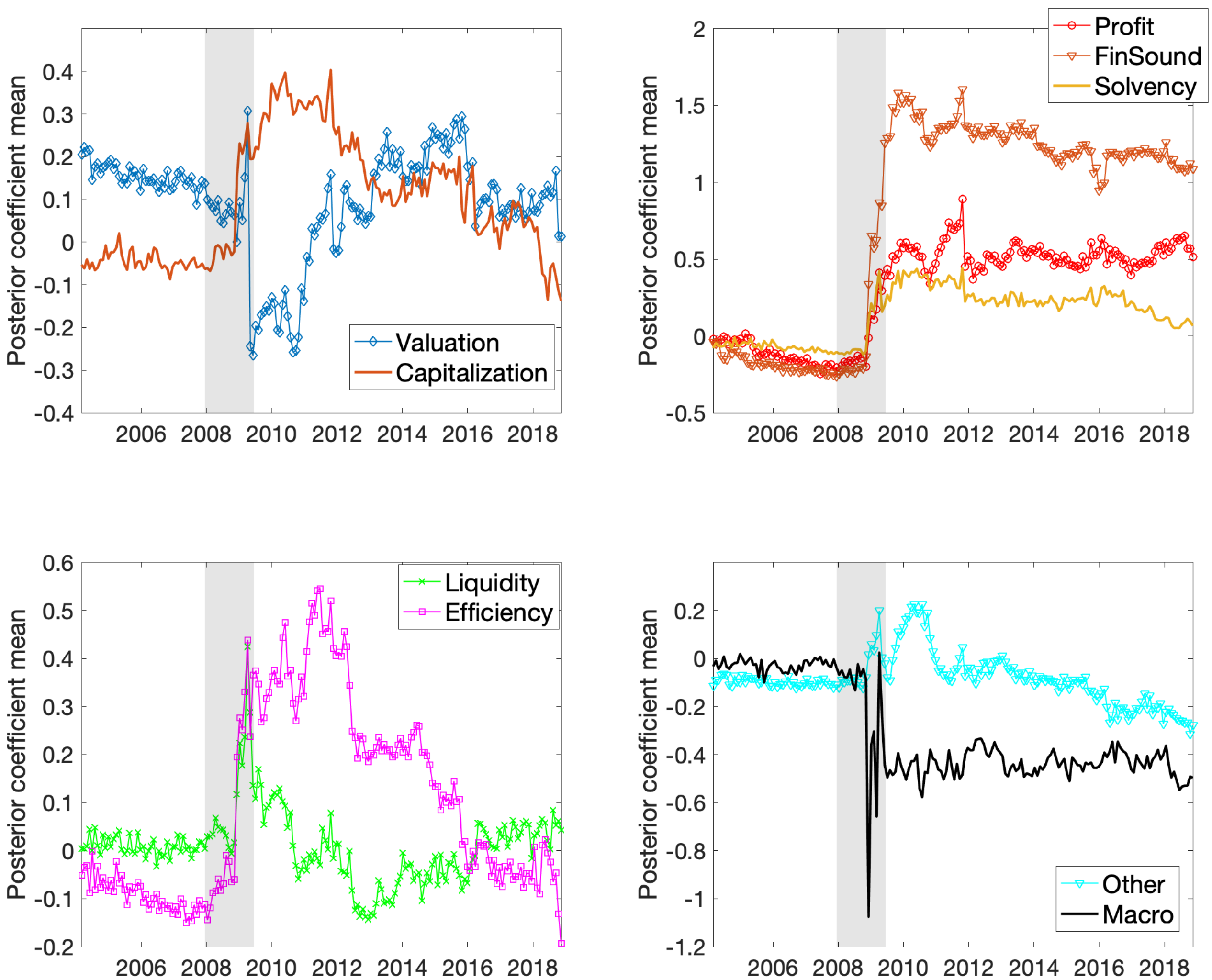


Figure 9: **Marginal effects of individual forecasts: Aggregate market**

This figure reports the posterior estimates of the latent parameters $\theta_{jt}$, $j = 1, \ldots, J$. We report the estimates for the aggregate value-weighted market portfolio from January 2004 to December 2018. Top panels report the estimates for capital and valuation groups (left panel) and profitability, financial soundness and solvency (right panel). Bottom panels report the estimates for the liquidity and efficiency groups (left panel) and the "other" group of financial ratios and aggregate macroeconomic variables.

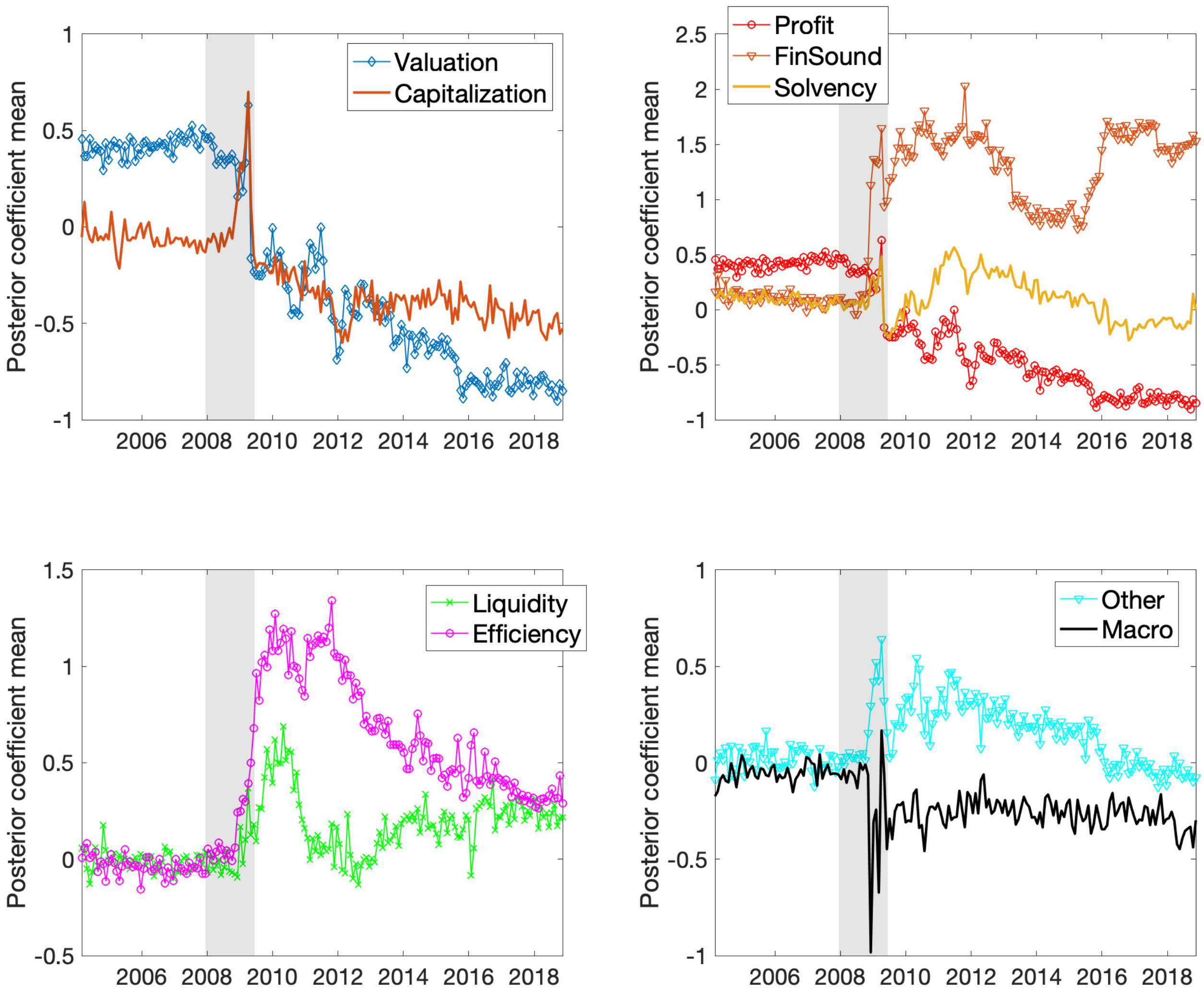
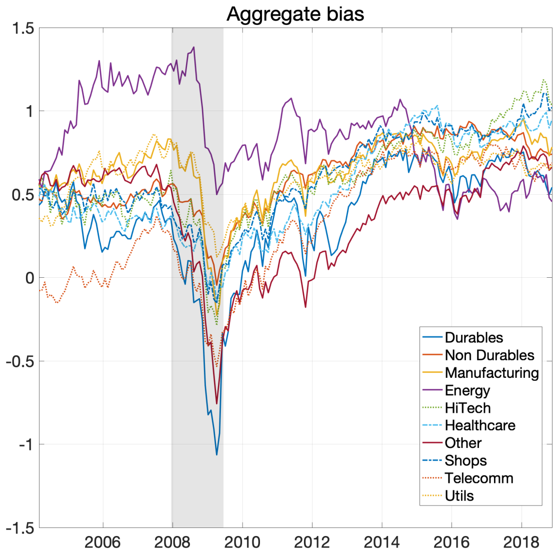
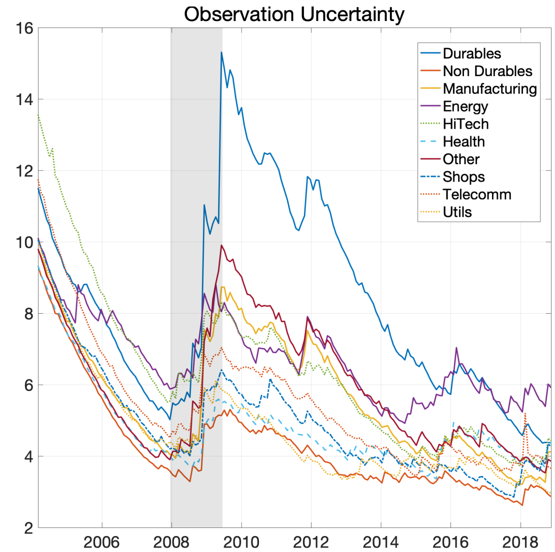


Figure 10: **Industry-specific parameters**

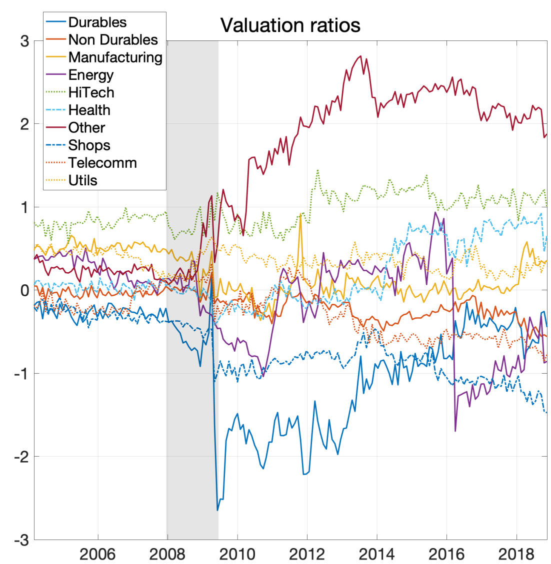
This figure reports the posterior estimates of the bias (top-left panel) and observation uncertainty (top-right panel), as well as two representative marginal effects on the aggregate forecasting from two groups of financial ratios, i.e., $\theta_{jt}$, $j = 1, \ldots, J$, namely value and profitability. We report the estimates for from January 2004 to December 2018.




(a) Bias




(b) Observation uncertainty




(c) Valuation ratios

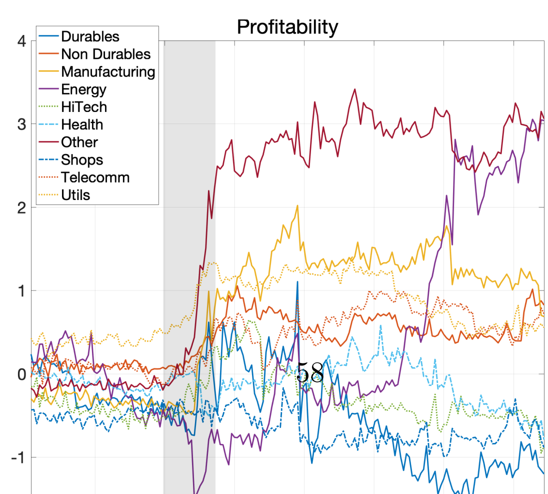


Table A.1: **Financial ratios**

This table reports description of the financial ratios used in the main empirical analysis. All original accounting variables are obtained from Compustat Quarterly and Annual file, whereas pricing related data are obtained both from CRSP and Compustat. Earnings-related variables are obtained from IBES. We also consider a market-wide aggregation where we consider a value-weighted aggregation of the financial ratios and returns. The sample size is from 1970:01 to 2018:12, monthly.

| Financial Ratio | Category | Description |
|---|---|---|
| Capitalization Ratio | Capitalization | Total Long-term Debt as a fraction of the sum of Total Long-term Debt, Common/Ordinary Equity and Preferred Stock |
| Common Equity/Invested Capital | Capitalization | Common equity as a fraction of invested capital |
| Long-term Debt/Invested Capital | Capitalization | Long-term debt as a fraction of the investment capital |
| Total Debt/Invested Capital | Capitalization | Total Debt (Long-term and Current) as a fraction of Invested Capital |
| Asset Turnover | Efficiency | Sales as a fraction of the average Total Assets based on the most recent two periods |
| Inventory Turnover | Efficiency | COGS as a fraction of the average Inventories based on the most recent two periods |
| Payables Turnover | Efficiency | COGS and change in Inventories as a fraction of the average of Accounts Payable based on the most recent two periods |
| Receivables Turnover | Efficiency | Sales as a fraction of the average of Accounts Receivables based on the most recent two periods |
| Sales/Stockholders Equity | Efficiency | Sales per dollar of total Stockholders' Equity |
| Sales/Invested Capital | Efficiency | Sales per dollar of Invested Capital |
| Sales/Working Capital | Efficiency | Sales per dollar of Working Capital, defined as difference between Current Assets and Current Liabilities |
| Inventory/Current Assets | Financial Soundness | Inventories as a fraction of Current Assets |
| Receivables/Current Assets | Financial Soundness | Accounts Receivables as a fraction of Current Assets |
| Free Cash Flow/Operating Cash Flow | Financial Soundness | Free Cash Flow as a fraction of Operating Cash Flow, where Free Cash Flow is defined as the difference between Operating Cash Flow and Capital Expenditures |
| Operating CF/Current Liabilities | Financial Soundness | Operating Cash Flow as a fraction of Current Liabilities |
| Cash Flow/Total Debt | Financial Soundness | Operating Cash Flow as a fraction of Total Debt |
| Cash Balance/Total Liabilities | Financial Soundness | Cash Balance as a fraction of Total Liabilities |
| Cash Flow Margin | Financial Soundness | Income before Extraordinary Items and Depreciation as a fraction of Sales |
| Short-term Debt/Total Debt | Financial Soundness | Short-term Debt as a fraction of Total Debt |
| Profit Before Depreciation/Current Liabilities | Financial Soundness | Operating Income before D&A as a fraction of Current Liabilities |
| Current Liabilities/Total Liabilities | Financial Soundness | Current Liabilities as a fraction of Total Liabilities |
| Total Debt/EBITDA | Financial Soundness | Gross Debt as a fraction of EBITDA |
| Long-term Debt/Book Equity | Financial Soundness | Long-term debt as a fraction of book equity |
| Interest/Average Long-term debt | Financial Soundness | Interest as a fraction of average Long-term debt based on most recent two periods |
| Interest/Average Total Debt | Financial Soundness | Interest as a fraction of average Total Debt based on most recent two periods |
| Long-term Debt/Total liabilities | Financial Soundness | Long-term Debt as a fraction of Total Liabilities |
| Total liabilities/Total Tangible Assets | Financial Soundness | Total liabilities to total tangible assets |
| Cash Conversion Cycle (Days) | Liquidity | Inventories per daily COGS plus Account Receivables per daily Sales minus Account Payables per daily COGS |
| Cash ratio | Liquidity | Cash and Short-term Investments as a fraction of Current Liabilities |
| Current ratio | Liquidity | Current Assets as a fraction of Current Liabilities |
| Quick Ratio (Acid Test) | Liquidity | Quick Ratio: Current Assets net of Inventories as a fraction of Current Liabilities |
| Accruals/Average Assets | Other | Accruals as a fraction of average Total Assets based on most recent two periods |
| Research and Development/Sales | Other | R&D expenses as a fraction of Sales |
| Advertising Expenses/Sales | Other | Advertising Expenses as a fraction of Sales |
| Labor Expenses/Sales | Other | Labor Expenses as a fraction of Sales |
| Effective Tax Rate | Profitability | Income Tax as a fraction of Pretax Income |

Table A.1: **Financial ratios (Cont'd)**

This table reports description of the financial ratios used in the main empirical analysis. All original accounting variables are obtained from Compustat Quarterly and Annual file, whereas pricing related data are obtained both from CRSP and Compustat. Earnings-related variables are obtained from IBES. We also consider a market-wide aggregation where we consider a value-weighted aggregation of the financial ratios and returns. The sample size is from 1970:01 to 2018:12, monthly.

| Financial Ratio | Category | Description |
|---|---|---|
| Gross Profit/Total Assets | Profitability | Gross Profitability as a fraction of Total Assets |
| After-tax Return on Average Common Equity | Profitability | Net Income as a fraction of average of Common Equity based on most recent two periods |
| After-tax Return on Total Stockholders' Equity | Profitability | Net Income as a fraction of average of Total Shareholders' Equity based on most recent two periods |
| After-tax Return on Invested Capital | Profitability | Net Income plus Interest Expenses as a fraction of Invested Capital |
| Gross Profit Margin | Profitability | Gross Profit as a fraction of Sales |
| Net Profit Margin | Profitability | Net Income as a fraction of Sales |
| Operating Profit Margin After Depreciation | Profitability | Operating Income After Depreciation as a fraction of Sales |
| Operating Profit Margin Before Depreciation | Profitability | Operating Income Before Depreciation as a fraction of Sales |
| Pre-tax Return on Total Earning Assets | Profitability | Operating Income After Depreciation as a fraction of average Total Earnings Assets (TEA) based on most recent two periods. |
| Pre-tax return on Net Operating Assets | Profitability | Operating Income After Depreciation as a fraction of average Net Operating Assets (NOA) based on most recent two periods. |
| Pre-tax Profit Margin | Profitability | Pretax Income as a fraction of Sales |
| Return on Assets | Profitability | Operating Income Before Depreciation as a fraction of average Total Assets based on most recent two periods |
| Return on Capital Employed | Profitability | Earnings Before Interest and Taxes as a fraction of average Capital Employed based on most recent two periods |
| Return on Equity | Profitability | Net Income as a fraction of average Book Equity based on most recent two periods. |
| Total Debt/Equity | Solvency | Total Liabilities to Shareholders' Equity (common and preferred) |
| Total Debt/Total Assets | Solvency | Total Debt as a fraction of Total Assets |
| Total Debt/Capital | Solvency | Total Debt as a fraction of Total Capital |
| After-tax Interest Coverage | Solvency | Multiple of After-tax Income to Interest and Related Expenses |
| Interest Coverage Ratio | Solvency | Multiple of Earnings Before Interest and Taxes to Interest and Related Expenses |
| Dividend Payout Ratio | Valuation | Dividends as a fraction of Income Before Extra. Items |
| Forward P/E to 1-year Growth (PEG) ratio | Valuation | Price-to-Earnings, excl. Extraordinary Items (diluted) to 1-Year EPS Growth rate |
| Forward P/E to Long-term Growth (PEG) ratio | Valuation | Price-to-Earnings, excl. Extraordinary Items (diluted) to Long-term EPS Growth rate |
| Trailing P/E to Growth (PEG) ratio | Valuation | Price-to-Earnings, excl. Extraordinary Items (diluted) to 3-Year past EPS Growth |
| Book/Market | Valuation | Book Value of Equity as a fraction of Market Value of Equity |
| Shillers Cyclically Adjusted P/E Ratio | Valuation | Multiple of Market Value of Equity to 5-year moving average of Net Income |
| Dividend Yield | Valuation | Indicated Dividend Rate as a fraction of Price |
| Enterprise Value Multiple | Valuation | Multiple of Enterprise Value to EBITDA |
| Price/Cash flow | Valuation | Multiple of Market Value of Equity to Net Cash Flow from Operating Activities |
| P/E (Diluted, Excl. EI) | Valuation | Price-to-Earnings, excl. Extraordinary Items (diluted) |
| Price/Operating Earnings (Basic, Excl. EI) | Valuation | Price to Operating EPS, excl. Extraordinary Items (Basic) |
| Price/Sales | Valuation | Multiple of Market Value of Equity to Sales |
| Price/Book | Valuation | Multiple of Market Value of Equity to Book Value of Equity |

Table D.1: **OOS economic significance w/o short sales: Mean-variance utility**

This table reports the annualized certainty equivalent values (annualised, in %) for portfolio decisions based on the out-of-sample forecasts of industry returns for an investor with *mean-variance* utility, a coefficient of risk aversion $\gamma = 5$, and *with short sales forbidden*, i.e., $\boldsymbol{w}_i \in (0, 1)$. The table reports three asset allocation exercises. The first case is a single investment in a given industry; the second a multiple investment in different industries; the third is a single investment in the aggregate stock market. The asset allocation decision is made at each time $t$ and is based on the predictions from each of the alternative models outlined in the main text. The models are benchmarked against the signal obtained from a recursive sample mean estimate. The predictions are obtained starting in February 2002, and the sample period is from 1970:01-2018:12, monthly. Statistical significance is based on a one-sided Diebold and Mariano (1995) test as extended by Harvey et al. (1997) to account for correlation. We report only the p-values when the difference in the certainty equivalent is positive, i.e., a model outperforms the sample mean forecast.

**Panel A:** Certainty equivalent

| | Industries | | | | | | | | | | Mkt | EW Port |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Durables | NonDurables | Manuf | Energy | HiTech | Health | Other | Shops | Telecomm | Utils | | |
| OLS | -0.478 | -0.425 | -0.291 | -0.734 | 0.182 | -0.081 | 0.135 | -0.753 | -1.057 | -0.248 | 0.000 | -0.032 |
| Lasso | 0.151 | 0.256 | -0.138 | -1.027 | -0.138 | 0.328 | 0.187 | -0.097 | -0.188 | -0.202 | -0.718 | -0.144 |
| Ridge | 0.013 | -0.553 | -0.346 | 0.048 | -0.434 | -0.499 | -0.086 | -0.582 | -1.250 | -0.092 | 0.225 | -0.323 |
| E-net | -0.142 | 0.020 | -0.202 | -0.919 | -0.606 | 0.058 | 0.500 | -0.681 | 0.555 | 0.082 | 0.127 | -0.110 |
| Random Forest | 0.409 | 0.158 | 0.284 | -0.211 | 0.467 | 0.337 | 0.417 | 0.018 | 0.335 | 0.126 | 0.249 | 0.235 |
| Neural Net | -0.049 | -0.073 | 0.027 | 0.090 | -0.140 | -0.114 | -0.015 | -0.079 | -0.036 | 0.040 | 0.031 | -0.029 |
| Factor Model | 0.496 | 0.262 | 0.422 | -0.009 | 0.486 | 0.066 | 0.389 | 0.403 | 0.352 | 0.331 | 0.347 | 0.259 |
| Equal-weight | 0.023 | 0.032 | 0.435 | 0.122 | -0.080 | 0.033 | 0.393 | 0.044 | 0.186 | -0.242 | 0.268 | 0.111 |
| BMA | 0.028 | 0.336 | 0.396 | 0.103 | 0.137 | -0.048 | -0.022 | -0.019 | 0.092 | -0.242 | 0.031 | 0.072 |
| Median | -0.051 | 0.004 | 0.432 | 0.367 | 0.204 | 0.102 | 0.240 | 0.172 | 0.150 | 0.117 | 0.258 | 0.181 |
| Trimmed mean | 0.021 | 0.047 | 0.468 | 0.268 | 0.077 | 0.069 | 0.317 | 0.073 | 0.181 | 0.100 | 0.242 | 0.169 |
| Macro | 0.464 | -0.030 | 0.415 | 0.090 | 0.102 | -0.060 | -0.021 | 0.310 | -0.161 | 0.005 | 0.152 | 0.115 |
| DRS | 0.690 | 0.383 | 0.677 | 0.394 | 0.724 | 0.704 | 0.941 | 0.292 | 0.391 | 0.571 | 0.438 | 0.500 |

**Panel B:** P-values

| | Industries | | | | | | | | | | Mkt | EW Port |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Durables | NonDurables | Manuf | Energy | HiTech | Health | Other | Shops | Telecomm | Utils | | |
| OLS | - | - | - | - | 0.304 | - | 0.424 | - | - | - | 0.500 | - |
| Lasso | 0.386 | 0.224 | - | - | - | 0.072 | 0.401 | - | - | - | - | - |
| Ridge | 0.492 | - | - | 0.475 | - | - | - | - | - | - | 0.208 | - |
| E-net | - | 0.481 | - | - | - | 0.435 | 0.067 | - | 0.077 | 0.391 | 0.036 | - |
| Random Forest | 0.109 | 0.198 | 0.159 | - | 0.087 | 0.078 | 0.071 | 0.462 | 0.170 | 0.274 | 0.223 | 0.183 |
| Neural Net | - | - | 0.264 | 0.383 | - | - | - | - | - | 0.145 | 0.241 | - |
| Factor Model | 0.057 | 0.109 | 0.074 | - | 0.003 | 0.308 | 0.020 | 0.061 | 0.070 | 0.035 | 0.065 | 0.080 |
| Equal-weight | 0.454 | 0.409 | 0.012 | 0.340 | - | 0.402 | 0.025 | 0.376 | 0.178 | - | 0.055 | 0.250 |
| BMA | 0.447 | 0.080 | 0.014 | 0.184 | 0.174 | - | - | - | 0.270 | - | 0.410 | 0.226 |
| Median | - | 0.487 | 0.018 | 0.107 | 0.197 | 0.180 | 0.178 | 0.145 | 0.198 | 0.231 | 0.128 | 0.187 |
| Trimmed mean | 0.455 | 0.366 | 0.007 | 0.167 | 0.361 | 0.274 | 0.046 | 0.279 | 0.164 | 0.277 | 0.062 | 0.223 |
| Macro | 0.078 | - | 0.032 | 0.130 | 0.281 | - | - | 0.016 | - | 0.471 | 0.213 | 0.174 |
| DRS | 0.024 | 0.057 | 0.014 | 0.020 | 0.019 | 0.058 | 0.020 | 0.049 | 0.049 | 0.027 | 0.029 | 0.033 |

Table D.2: **OOS economic significance w/o short sales: Power utility**

This table reports the annualized certainty equivalent values (annualised, in %) for portfolio decisions based on the out-of-sample forecasts of industry returns for an investor with *power* utility, a coefficient of risk aversion $\gamma = 5$, and *with short sales forbidden*, i.e., $\boldsymbol{w}_i \in (0, 1)$. The table reports three asset allocation exercises. The first case is a single investment in a given industry; the second a multiple investment in different industries; the third is a single investment in the aggregate stock market. The asset allocation decision is made at each time $t$ and is based on the predictions from each of the alternative models outlined in the main text. The models are benchmarked against the signal obtained from a recursive sample mean estimate. The predictions are obtained starting in February 2002, and the sample period is from 1970:01-2018:12, monthly. Statistical significance is based on a one-sided Diebold and Mariano (1995) test as extended by Harvey et al. (1997) to account for correlation. We report only the p-values when the difference in the certainty equivalent is positive, i.e., a model outperforms the sample mean forecast.

**Panel A:** Certainty equivalent

| | Industries | | | | | | | | | | Mkt | EW Port |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Durables | NonDurables | Manuf | Energy | HiTech | Health | Other | Shops | Telecomm | Utils | | |
| OLS | -0.176 | -0.154 | -1.737 | -0.121 | -0.205 | -0.200 | -0.448 | -0.116 | -0.520 | -0.068 | -0.054 | -0.290 |
| Lasso | -0.949 | -0.724 | -1.447 | -0.573 | 0.304 | 0.289 | 0.566 | -0.057 | 0.172 | 0.145 | 0.133 | 0.157 |
| Ridge | 0.465 | 0.038 | -0.854 | 0.162 | 0.371 | 0.262 | 0.229 | 0.045 | 0.080 | 0.026 | 0.340 | 0.222 |
| E-net | 0.338 | 0.490 | -0.032 | 0.454 | 0.540 | 0.473 | 0.242 | 0.343 | 0.285 | 0.095 | 0.366 | 0.159 |
| Random Forest | 0.271 | 0.242 | 0.490 | 0.454 | 0.540 | 0.473 | 0.218 | 0.285 | 0.309 | 0.366 | 0.338 | 0.362 |
| Neural Net | -0.204 | -0.224 | -0.076 | 0.121 | -0.205 | -0.200 | -0.041 | -0.201 | -0.294 | -0.047 | -0.028 | -0.127 |
| Factor Model | 0.446 | 0.471 | 0.194 | 0.342 | 0.168 | 0.286 | 0.203 | 0.383 | 0.322 | 0.401 | 0.345 | 0.324 |
| Equal-weight | 0.551 | 0.174 | 0.187 | 0.378 | 0.527 | 0.331 | 0.437 | 0.260 | 0.262 | 0.360 | 0.179 | 0.331 |
| BMA | 0.482 | 0.466 | 0.465 | 0.293 | 0.533 | 0.084 | 0.396 | 0.448 | 0.378 | 0.360 | 0.438 | 0.395 |
| Median | 0.600 | 0.070 | 0.271 | 0.309 | 0.520 | 0.435 | 0.379 | 0.277 | 0.359 | 0.205 | 0.506 | 0.357 |
| Trimmed mean | 0.736 | 0.131 | 0.273 | 0.420 | 0.428 | 0.360 | 0.398 | 0.522 | 0.532 | 0.160 | 0.361 | 0.393 |
| Macro | 0.179 | 0.051 | 0.355 | 0.277 | 0.058 | 0.099 | 0.394 | 0.056 | -0.165 | 0.037 | 0.168 | 0.137 |
| DRS | 1.115 | 1.124 | 0.712 | 0.892 | 0.796 | 1.035 | 0.983 | 1.114 | 1.087 | 1.069 | 1.053 | 0.998 |

**Panel B:** P-values

| | Industries | | | | | | | | | | Mkt | EW Port |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Durables | NonDurables | Manuf | Energy | HiTech | Health | Other | Shops | Telecomm | Utils | | |
| OLS | - | - | - | - | - | - | - | - | - | - | - | - |
| Lasso | - | - | - | - | 0.325 | 0.018 | 0.084 | - | 0.094 | 0.443 | 0.388 | 0.225 |
| Ridge | 0.012 | 0.384 | - | 0.098 | 0.175 | 0.403 | 0.110 | 0.403 | 0.357 | 0.284 | 0.053 | 0.228 |
| E-net | 0.001 | 0.360 | - | 0.240 | 0.150 | 0.026 | 0.005 | 0.437 | 0.005 | 0.208 | 0.004 | 0.144 |
| Random Forest | 0.012 | 0.206 | 0.113 | 0.172 | 0.001 | 0.005 | 0.005 | 0.194 | 0.004 | 0.113 | 0.016 | 0.076 |
| Neural Net | - | - | - | 0.001 | - | - | - | - | - | - | - | - |
| Factor | 0.048 | 0.059 | 0.141 | 0.176 | 0.311 | 0.249 | 0.512 | 0.516 | 0.401 | 0.580 | 0.313 | 0.301 |
| Equal-weight | 0.039 | 0.053 | 0.368 | 0.246 | 0.056 | 0.154 | 0.164 | 0.136 | 0.125 | 0.092 | 0.137 | 0.143 |
| BMA | 0.103 | 0.103 | 0.104 | 0.096 | 0.044 | 0.330 | 0.070 | 0.036 | 0.082 | 0.040 | 0.029 | 0.094 |
| Median | 0.067 | 0.364 | 0.089 | 0.059 | 0.012 | 0.069 | 0.148 | 0.158 | 0.262 | 0.134 | 0.025 | 0.126 |
| Trimmed mean | 0.046 | 0.268 | 0.110 | 0.102 | 0.014 | 0.028 | 0.125 | 0.101 | 0.079 | 0.216 | 0.005 | 0.099 |
| Macro | 0.255 | 0.414 | 0.033 | 0.141 | 0.326 | 0.304 | 0.070 | 0.147 | - | 0.362 | 0.127 | 0.218 |
| DRS | 0.011 | 0.001 | 0.020 | 0.012 | 0.031 | 0.038 | 0.039 | 0.020 | 0.004 | 0.028 | 0.038 | 0.022 |

Table E.1: **Simulation results**

This table reports the out-of-sample comparison for different simulated data for our DRS predictive model against a variety of alternative forecasting methods, such as a model with all regressors, the lasso, a latent factor model, an equal-weight linear pooling, and BMA. The performance comparison is based on the Root Mean Squared Error (RMSE).

**Panel A:** Forecasting based on first $n$ samples

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

**Panel B:** Forecasting based on last $l$ samples

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