Received: 2 February 2018 | Revised: 24 May 2019

DOI: 10.1002/jae.2725

RESEARCH ARTICLE

WILEY Journal of APPLIED ECONOMETRICS

Macroeconomic forecast accuracy in a data-rich environment

Rachidi Kotchoni[1] | Maxime Leroux[2] | Dalibor Stevanovic[2]

[1]Economix-CNRS, Université Paris Nanterre, Paris, France

[2]Département des sciences économiques, Université du Québec à Montréal, Montréal, Quebec, Canada

Correspondence
Département des sciences économiques, Université du Québec à Montréal, 315 Ste-Catherine Est, Montréal, QC H2X 3X2, Canada.
Email: dstevanovic.econ@gmail.com

Summary

The performance of six classes of models in forecasting different types of economic series is evaluated in an extensive pseudo out-of-sample exercise. One of these forecasting models, regularized data-rich model averaging (RDRMA), is new in the literature. The findings can be summarized in four points. First, RDRMA is difficult to beat in general and generates the best forecasts for real variables. This performance is attributed to the combination of regularization and model averaging, and it confirms that a smart handling of large data sets can lead to substantial improvements over univariate approaches. Second, the ARMA(1, 1) model emerges as the best to forecast inflation changes in the short run, while RDRMA dominates at longer horizons. Third, the returns on the S&P 500 index are predictable by RDRMA at short horizons. Finally, the forecast accuracy and the optimal structure of the forecasting equations are quite unstable over time.

1 | INTRODUCTION

Many economic data sets have now reached tremendous sizes, both in terms of the number of variables and the number of observations. As all of these series may not be relevant for a particular forecasting exercise, one will have to preselect the most important candidate predictors according to economic theories, the relevant empirical literature, and own heuristic arguments. In a data-rich environment, the econometrician may still be left with a few hundreds of candidate predictors after the preselection process. Unfortunately, the performance of standard econometric models tends to deteriorate as the dimensionality of the data increases. This is the well-known curse of dimensionality. In this context, the challenge faced by empirical researchers is to design computationally efficient methods capable of turning big data sets into concise information.[1]

When confronted with a large number of variables, econometricians often resort to sparse modeling, regularization, or dense modeling. Sparse models involve a variable selection procedure that discards the least relevant predictors. In regularized models, a large number of variables are accommodated but a shrinkage technique is used to discipline the behavior of the parameters (e.g., Ridge). Lasso regularization leads to sparse models ex post as it constrains the coefficients of the least relevant variables to be null. In factor models—an example of dense modeling—the dynamics of a large number of variables is assumed to be governed by a small number of common components. All three approaches entail an implicit or explicit dimensionality reduction that is intended to control the overfitting risk and maximize the out-of-sample forecasting performance.

[1]Bayesian techniques developed in the recent years to handle larger than usual vector autoregression (VAR) models can be viewed as an effort towards this objective. See Banbura, Giannone, and Reichlin (2010), Koop (2013), Carriero, Clark, and Marcellino (2015), and Giannone, Lenza, and Primiceri (2015), among others

Giannone, Lenza, and Primiceri (2017) considered a Bayesian framework that balanced the quest for sparsity with the desire to accommodate a large number of relevant predictors. They found that the posterior distribution of parameters was spread over all types of models rather than being concentrated on a single sparse model or a single dense model. This suggests that a well-designed model averaging technique can outperform any sparse model. We build on this intuition and put forward a new class of regularized data-rich models that combines regularization and model averaging techniques.

Given the growing popularity of models that address big data issues, there is a need for an extensive study that compares their performance. This paper contributes to filling this gap by comparing the performance of six classes of model in forecasting industrial production growth, employment growth, consumer price index (CPI) acceleration (i.e., variations of inflation), and the S&P 500 returns.[2] Only few studies have done such a large-scale comparison exercise: see Boivin and Ng (2005), Stock and Watson (2006), Kim and Swanson (2014), Cheng and Hansen (2015), Carrasco and Rossi (2016), and Groen and Kapetanios (2016).

The first class of forecasting models considered consists of standard and univariate specifications, namely the autoregressive direct (ARD), the autoregressive iterative (ARI), the autoregressive moving average (ARMA(1, 1)), and the autoregressive distributed lag (ADL) models. The second class of model consists of autoregressions that are augmented with factors that are extracted from a set of predictors beforehand: the diffusion indices (DI) of Stock and Watson (2002b), the targeted DI of Bai and Ng (2008), the DI with dynamic factors of Forni, Hallin, Lippi, and Reichlin (2005), and, to some extent, the three-pass regression filter (3PRF) of Kelly and Pruitt (2015). In the third type of model, one jointly specifies the dynamics for the variable of interest (to be forecast) and the factors. In the latter category, we have the factor-augmented VAR (FAVAR) of Boivin and Ng (2005), the factor-augmented VARMA (FAVARMA) of Dufour and Stevanovic (2013), and the dynamic factor model (DFM) of Forni et al. (2005).

The fourth class of model consists of data-rich model averaging techniques, which are known as complete subset regressions (CSRs) (see Elliott, Gargano, & Timmermann, 2013). The fifth class of models, which we term regularized data-rich model averaging (RDRMA), consists of penalized versions of the CSR (i.e., CSR combined either with preselection of variables or with Ridge regularization). This combination of sparsity/regularization and model averaging is quite new in the forecasting literature. Finally, the sixth class of model consists of methods that average all available forecasts. We consider the naive average (AVRG), the median (MED), the trimmed average (T-AVRG), and the inversely proportional average of all forecasts (IP-AVRG), as in Stock and Watson (2004).

The data employed for this study are monthly macroeconomic series from McCracken and Ng (2016). Comparison of the forecasting models is based on their pseudo out-of-sample performance along three metrics: the root mean square prediction error (RMSPE) and the ratio of correctly signed forecasts (RCSF). The results based on the RMSPE are presented in the main text, while the Appendix summarizes the findings for RCSF. Additional results for the core CPI inflation, exchange rates, and interest rates are deferred to the Supporting Information. For each series, horizon, and out-of-sample period, the hyperparameters of the models are recalibrated using the Bayesian information criterion (BIC). Variations in the optimal hyperparameters over time allow us to gauge the stability of our forecast equations.

To the best of our knowledge, our paper is a rare attempt to put so many different models together and compare their predictive performance on several types of data in a pseudo out-of-sample forecasting experiment. Disentangling which type of models have significant forecasting power for real activity, prices, and stock market is valuable for practitioners and policymakers. The pseudo out-of-sample exercise generates a huge volume of empirical results. The presentation that follows focuses on highlights that convey the most important messages.

Irrespective of the forecast horizon and performance evaluation metrics, RDRMA and forecast combinations emerge as the best to forecast real variables. factor-structure-based and factor-augmented models are dominated in terms of RMSPE, but they are good benchmarks when the RCSF is considered. This is attributable to the fact that data-rich models involving factors are flexible enough to accommodate instabilities in the dynamics of the target, as suggested by Carrasco and Rossi (2016) and Pettenuzzo and Timmermann (2017). For the same reason, factor-structure-based and factor-augmented models emerge among the best to forecast real variables during recessions. Our RDRMA improves the RMSPE for industrial production by up to 24%, which supports the finding from Stock and Watson (2006). Kim and Swanson (2014) found that the combination of factor modeling and shrinkage worked best in terms of MSPE, whereas model averaging performed poorly. Our results suggest that data-rich model averaging combined with regularization outperforms the other methods in general.

[2]These variables are selected for their popularity in the forecasting literature. Results for the core CPI, interest rate, and exchange rate variations are available in the supplementary material (Supporting Information).

ARMA(1, 1) emerges as an excellent parsimonious model to forecast the variations of inflation as short horizons. This is in line with Stock and Watson (2007) and Faust and Wright (2013). RDRMA dominates at horizons 9 and 12 months. During recessions, ARMA(1, 1) delivers its best performance 3 months ahead only, while model averaging and forecast combinations dominate at the other horizons. The presence of an MA component in inflation time series has been suggested in the literature but the predictive performance of the ARMA(1, 1) model has not been highlighted in a large-scale model comparison exercise as done here. One possible explanation for this good performance of ARMA(1, 1) is that inflation anticipations are so well anchored that inflation variations are exogenous with respect to the conditioning information set.

In general, the best approaches to forecast the S&P 500 returns are data-rich model averaging (regularized or not) and forecast combinations. Factor structure models have significant predictive power for the sign of the S&P 500 returns and even at long horizons. During recessions, data-rich model averaging and forecast combinations dominate at short horizons, whereas factor-structure-based models dominate at longer horizons. RDRMA and forecast combinations deliver the best performance in terms of correctly signed forecasts in the short run, whereas the FAVAR specifications produce the best RCSF for longer horizons. If we abstract from long horizon during recessions, RW models (with or without drift) are dominated with respect to all metrics and at all horizons. This suggests that stock returns are predictable to some extent.

Overall, our results show that sparsity and regularization can be smartly combined with model averaging to obtain forecasting models that dominate state-of-the-art benchmarks. Our paper therefore provides a frequentist support for the conclusions found by Giannone et al. (2017) in their Bayesian framework. Another important finding is that the performance of models is unstable, as we find overwhelming evidence of structural changes in all aspects of the forecasting equations. However, a combination of regularization and data-rich model averaging gives a very robust and flexible model that is likely to continue performing well in those changing economic environments.

In the remainder of the paper, we first present forecasting models in Section 2. Section 3 presents the design of the pseudo out-of-sample exercise. Section 4 reports the main empirical results. Section 5 analyzes the stability of the forecast accuracy and Section 6 concludes. Additional results are available in the Appendix and in the Supporting Information.

2 | PREDICTIVE MODELING

This section presents the predictive models considered in the paper. We consider the following general framework:[3]

$$\arg\min_{\theta} \sum_{t} L(y_{t+h} - f(X_t;\theta)) + \lambda \text{Pen}(\theta), \quad t = 1, \ldots, T, \tag{1}$$

where y_{t+h} is the variable to be predicted h periods ahead (target) and X_t is the N-dimensional vector of predictors available at time t. L is a loss function that is in most occasions assumed quadratic. The function $f()$ models the predictors' space in a (non)linear and/or (non)parametric way; Pen() represents a regularization or penalization scheme associated with $f()$, while λ is a hyperparameter that allows us to fine tune the regularization strength.

In this paper, our forecasting models assume a quadratic loss function in-sample (i.e., for model estimation). Hence the optimal forecast is the conditional expectation $\mathrm{E}(y_{t+h}|X_t)$. The regularization, when needed, will consist of soft and hard thresholding, as well as of dimensionality reduction by principal component analysis.

2.1 | Forecasting targets

Let Y_t denote an economic time series of interest. If $\ln Y_t$ is a stationary process, we will consider forecasting its average over the period $[t+1, t+h]$ given by

$$y_{t+h}^{(h)} = (\text{freq}/h) \sum_{k=1}^{h} y_{t+k}, \tag{2}$$

where $y_t \equiv \ln Y_t$ and freq depends on the frequency of the data (e.g., 1200 if Y_t is monthly).

Most of the time, we are confronted with $I(1)$ series in macroeconomics. For such series, our goal will be to forecast the average annualized growth rate over the period $[t+1, t+h]$, as in Stock and Watson (2002b) and McCracken and Ng (2016).

[3]See Mullainathan and Spiess (2017) and Frank Diebold's blog https://fxdiebold.blogspot.com/2017/01/all-of-machine-learning-in-one.html.

We shall therefore define $y_{t+h}^{(h)}$ as

$$y_{t+h}^{(h)} = (\text{freq}/h)\sum_{k=1}^{h} y_{t+k} = (\text{freq}/h)\ln(Y_{t+h}/Y_t), \tag{3}$$

where $y_t \equiv \ln Y_t - \ln Y_{t-1}$. In cases where $\ln Y_t$ is better described by an $I(2)$ process, we define $y_{t+h}^{(h)}$ as

$$y_{t+h}^{(h)} = (\text{freq}/h)\sum_{k=1}^{h} y_{t+k} = (\text{freq}/h)\left[\ln(Y_{t+h}/Y_{t+h-1}) - \ln(Y_t/Y_{t-1})\right], \tag{4}$$

where $y_t \equiv \ln Y_t - 2\ln Y_{t-1} + \ln Y_{t-2}$.

2.2 | Regularized data-rich model averaging

Our main workhorse is regularized data-rich model averaging (RDRMA), an approach that combines preselection and regularization with the CSR of Elliott et al. (2013). The idea of CSR is to generate a large number of predictions based on different subsets of X_t and construct the final forecast as the simple average of the individual forecasts:

$$y_{t+h,m}^{(h)} = c + \rho y_t + \beta X_{t,m} + \varepsilon_{t,m}, \tag{5}$$

$$\hat{y}_{T+h|T}^{(h)} = \frac{\sum_{m=1}^{M} \hat{y}_{T+h|T,m}^{(h)}}{M}, \tag{6}$$

where $X_{t,m}$ contains L series for each model $m = 1, \ldots, M$.[4]

We modify the CSR by following the intuition of Giannone et al. (2017), who found in a Bayesian forecasting exercise that posterior predictive distributions are a combination of many different models rather than being concentrated on a single sparse model or a single dense model. This finding suggests that a well-designed model averaging technique can outperform any sparse model. As not all the predictors in X_t will be relevant to forecast y_{t+h}, we propose either to preselect those that have enough predicting power or regularize each predictive regression ex post. Similar to our strategy, Diebold and Shin (2018) propose a Lasso-based procedure to set some forecast combining weights to zero. Instead, we propose to shrink the space of potential regressors, and therefore the set of possible predictive models.

2.2.1 | Targeted CSR

In the targeted CSR, we preselect a subset of relevant predictors (first step) before applying the CSR algorithm (second step). This first step is intended to discipline the behavior of the CSR algorithm ex ante. We follow Bai and Ng (2008) in this step and consider soft and hard thresholding.

1. Hard or soft thresholding $\rightarrow X_t^* \in X_t$

   (a) *Hard thresholding*. A univariate predictive regression is done for each predictor X_{it}:

$$y_{t+h}^{(h)} = \alpha + \sum_{j=0}^{3} \rho_j y_{t-j} + \beta_i X_{i,t} + \epsilon_t. \tag{7}$$

   The subset X_t^* is obtained by gathering those series whose coefficients β_i have the t-stat larger than the critical value t_c: $X_t^* = \{X_i \in X_t | t_{Xi} > t_c\}$, with $t_c = 1.65$.

[4] L is usually set to 1, 10 or 20 and M is the total number of models considered (up to 5,000 in this paper).

(b) *Soft thresholding*. A predictive Lasso regression is performed for all predictors X_t:

$$\hat{\beta}^{\text{lasso}} = arg\min_{\beta}\left[\sum_{t=1}^{T}\left(y_{t+h}^{(h)} - \alpha + \sum_{j=0}^{3}\rho_j y_{t-j} + \beta X_t\right)^2 + \lambda\sum_{i=1}^{N}|\beta_i|\right]. \tag{8}$$

Here, we let the Lasso regularizer select the subset of relevant predictors $X_t^* = \{X_i \in X_t | \beta_i^{\text{lasso}} \neq 0\}$. The hyperparameter λ is selected to target approximately 30 series, which was used in Bai and Ng (2008) and in Giannone et al. (2017).

2. CSR of Equations (5)–(6) on the subset of relevant predictors X_t^*.

We consider four specifications of targeted CSR: soft and hard thresholding, with 10 and 20 regressors, labeled T-CSR-soft,10, T-CSR-soft,20, T-CSR-hard,1.65,10, and T-CSR-hard,1.65,20 respectively later in the tables. In terms of the general predictive setup in Equation (1), the first step of this model uses two types of the regularization: subset selection and Lasso.

2.2.2 | Ridge CSR

Alternatively, one may choose to use the entire set of predictors X_t but discipline the CSR algorithm ex post using a Ridge penalization. Each predictive regression (Equation (6)) of the CSR algorithm is estimated as follows:

$$\hat{\beta}^{\text{ridge}} = \arg\min_{\beta}\left[\sum_{t=1}^{T}(y_{t+h,m}^{(h)} - c - \rho y_t - \beta X_{t,m})^2 + \lambda\sum_{i=1}^{N}\beta_i^2\right]. \tag{9}$$

The final forecast is constructed as usual:

$$\hat{y}_{T+h|T}^{(h)} = \frac{\sum_{m=1}^{M}\hat{y}_{T+h|T,m}^{(h)}}{M}.$$

The intuition here is rather simple. As the CSR consists of combining a large number of forecasts obtained from randomly selected subsets of predictors, some subsets of predictors will likely be subject to multicollinearity problems. This issue is important in macroeconomic applications where many series are known to be highly correlated. A Ridge penalization allows us to elude this problem and produces a well-behaved forecast from every subsample. We consider two specifications of Ridge CSR based on 10 and 20 regressors, labeled R-CSR,10 and R-CSR,20, respectively.

2.3 | Benchmark models

We consider several benchmark models that have been extensively used in the literature. Table 1 lists all the models grouped in six categories. A detailed description is deferred to the Supporting Information Appendix.

The first category of models consists of standard time series models (which use a limited number of predictors), such as autoregressive predictive models with direct and iterative ways of constructing the forecast, ARMA(1, 1), and autoregressive distributed lag models.

The second and third categories of model exploit large data sets in two different ways. The second category gathers factor-augmented regressions that are instances of the diffusion indices model of Stock and Watson (2002a). The main feature of these models is that they treat the factors as exogenous predictors (i.e., factors are extracted separately and plugged into the forecasting equation). By contrast, the joint dynamics of the factors is endogenous in the third category of models, meaning that it is intertwined with the dynamics of the variable that we seek to forecast.

CSRs are gathered in a fourth category called "data-rich model averaging," whereas their regularized and sparse versions are gathered in the fifth category. The sixth category of forecasting methods simply consists of alternative ways of averaging all available forecasts. In total, we have 31 different forecasting approaches to evaluate in the horse race.

3 | EMPIRICAL EVALUATION OF THE FORECASTING MODELS

This section presents the data and the design of the pseudo out-of-sample experiment.

TABLE 1 List of all forecasting models

Standard time series models	
ARD	Autoregressive direct
ARI	Autoregressive iterative
ARMA(1, 1)	Autoregressive moving average
ADL	Autoregressive distributed lag
Factor-augmented regressions	
ARDI	Autoregressive diffusion indices, Stock and Watson (2002a)
ARDIT	Targeted diffusion indices, Bai and Ng (2008)
ARDI-DU	ARDI with dynamic factors, Forni et al. (2005)
3PRF	Three-pass regression filter, Kelly and Pruitt (2015)
Factor-structure-based models	
FAVAR	Factor-augmented VAR, Boivin and Ng (2005)
FAVARMA	Factor-augmented VARMA, Dufour et al. (2013)
DFM	Dynamic factor model, Forni et al. (2005)
Data-rich model averaging	
CSR	Complete subset regressions, Elliott et al. (2013)
Regularized data-rich model averaging	
T-CSR	Targeted CSR
R-CSR	Ridge CSR
Lasso	Least absolute shrinkage and selection operator
Forecast combinations	
AVRG	Equal-weighted forecasts average
Median	Median forecast
T-AVRG	Trimmed average
IP-AVRG	Inversely proportional average

3.1 | Data

We use historical data to evaluate and compare the performance of all the forecasting models described previously. The data set employed is an updated version of Stock and Watson's macroeconomic panel. It consists of 134 monthly macroeconomic and financial time series that are observed from 1960:M01 to 2014:M12 and it can be accessed via the Federal Reserve of St. Louis's website (FRED). Details on the construction of these series can be found in McCracken and Ng (2016).

The empirical exercise is easier when the data set is balanced. In practice, there is usually a tradeoff between the relevance of a time series and its availability (and frequency). Not all series are available from the 1960:M01 starting date in the McCracken and Ng (2016) database. This is accommodated in the rolling window setup by expanding the information set used for the prediction as the window moves forward.

Our models all assume that the variables y_t and X_t are stationary. However, most macroeconomic and financial indicators must undergo some transformation in order to achieve stationarity. This suggests that unit root tests must be performed before knowing the exact transformation to use for a particular series. The unit root literature provides much evidence on the lack of power of unit root test procedures in finite samples, especially with highly persistent series. Therefore, we simply follow McCracken and Ng (2016) and Stock and Watson (2002b) and assume that price indices are all $I(2)$ while interest and unemployment rates are $I(1)$.[5]

3.2 | Pseudo out-of-sample experimental design

The pseudo out-of-sample period is 1970:M01 to 2014:M12. The forecasting horizons considered are 1–12 months. There are 540 evaluation periods for each horizon. All models are estimated on rolling windows. We have compared the forecast accuracy of rolling versus expanding (or recursive) windows and the results are similar. For each model, the optimal hyperparameters (number of factors, number of lags, etc.) are specifically selected for each evaluation period and forecasting horizon. The size of the rolling window is $120 - h$ months.

[5]Bernanke et al. (2005) keep inflation, interest, and unemployment rates in levels. Choosing Stock and Watson (SW) or Bernanke, Boivin, and Eliasz (BBE) transformations has effects on correlation patterns in X_t. Under BBE, the group of interest rates is highly correlated as well as the inflation rates. As pointed out by Boivin and Giannoni (2006), the presence of these clusters may alter the estimation of *common* factors. Under SW, these clusters are less important. Recently, Banerjee, Marcellino and Masten (2014) and Barigozzi, Lippi, and Luciani (2016) proposed to deal with the unit root instead of differentiating the data.

3.3 | Variables of interest

We focus on four variables in the subsequent presentation: industrial production (INDPRO), employment (EMP), consumer price index (CPI), and SP& 500 index. INDPRO and EMP are real variables, CPI is a nominal variable, while S&P 500 represents the stock market. Additional results are available in the Supporting Information for core CPI, 10-year Treasury constant maturity rate (GS10), and the US–UK and US–Canada bilateral exchange rates. The logarithm of INDPRO, EMP and S&P 500 are treated as $I(1)$, whereas the logarithm of the CPI is assumed to be $I(2)$, as in Stock and Watson (2002b) and McCracken and Ng (2016).

3.4 | Forecast evaluation metrics

Following standard practice in the forecasting literature, we evaluate the quality of our point forecasts by using the root mean square prediction error (RMSPE). A standard Diebold–Mariano test procedure is used to compare the predictive accuracy of each model against the autoregressive direct model.

For the sake of generality, we also implement the model confidence set (MCS) introduced in Hansen, Lunde, and Nason (2011). The MCS allows us to select the subset of best models at a given confidence level. It is constructed by first finding the best forecasting model, and then selecting the subset of models that are not significantly different from the best model at a desired confidence level. We construct each MCS based on the quadratic loss function and 4,000 bootstrap replications. As expected, we find that the $(1 - \alpha)$ MCS contains more models when α is smaller. The empirical results for 75% are presented in the main text, while the Supporting Information contain the results for $\alpha = 10\%, 50\%$.

In the Appendix, we consider an alternative metric to evaluate our point forecasts: the ratio of correctly signed forecasts (RCSF). This metric captures some aspects of the distribution of the forecasts that the RMSPE may miss. For instance, a model that is dominated in terms of RMSPE can still have superior performance at generating forecasts that have the same signs as the target.

4 | MAIN RESULTS

This section presents our main empirical results for industrial production, employment growth, variations of inflation, and returns on the S&P 500 index. The analysis is done for the full out-of-sample period as well as for NBER recessions taken separately (i.e., when the target belongs to a recession episode). Indeed, knowledge of the models that have performed best historically during recessions is of interest for policymakers, practitioners, and real-time forecasters. If the probability of recession is high enough at a given period, our results can provide ex ante guidance on which model is likely to perform best in such circumstances.

4.1 | Industrial production growth

We now examine the performance of the various models at forecasting the industrial production growth. Table 2 presents the ratio of the RMSPE of each model and that of the ARD model (henceforth, relative RMSPE), both for the full out-of-sample period (1970–2014) and NBER recessions (i.e., target observation belongs to a recession episode). In the main text, the results are shown only for horizons 1, 3, 6, 9, and 12 months. Bold characters identify the models that are selected into the 75% MCS. The best model in terms of relative RMSPE (i.e., the minimum relative RMSPE) for each horizon is underlined, and the significance levels for Diebold–Mariano tests are displayed using the conventional notation with three, two, and one asterisks.

When the full out-of-sample period is considered, the best approach to forecast industrial production growth belongs to either forecast combinations or RDRMA. Note that the MCS contains models that belong to factor-augmented regressions, factor-structure-based models, and data rich model averaging, but not to standard time series models. Note that actual magnitudes of forecasts errors are in line with Stock and Watson (2002b).

During recessions, the best model to forecast industrial production growth belongs to either factor-augmented regressions or factor-structure-based models. This may be explained by the fact that these models are flexible enough to accommodate the faster than usual changes in economic variables during recession. Here too, the MCS contains forecasting models that pertain to other categories, notably data-rich model averaging (regularized or not) and forecast combinations. Interestingly, Lasso is present in the MCS at most horizons during recessions. As expected, the magnitude of forecast errors increases during recessions; see RMSPE for the ARD model.

TABLE 2 Industrial production: relative RMSPE

Model	Full out-of-sample					NBER recessions periods				
	$h = 1$	$h = 3$	$h = 6$	$h = 9$	$h = 12$	$h = 1$	$h = 3$	$h = 6$	$h = 9$	$h = 12$
Standard time series models										
ARD (RMSPE)	0.0856	0.0625	0.0582	0.0541	0.0506	0.1409	0.1108	0.1053	0.0943	0.085
ARI	1	1.03	1.01	1	1.02	1	1.08**	0.99	0.98	0.99
ARMA(1, 1)	0.97*	0.99	1.01	1.04	1.10*	0.92**	0.99	0.97*	0.98	1.01
ADL	1	1.01*	1.01	1.03	1	0.98	1.03*	1.02	1.01	0.99
Factor-augmented regressions										
ARDI	0.93**	0.90**	**0.84****	0.83**	0.81***	0.83***	**0.81*****	**0.72*****	0.82**	0.85**
ARDI-soft	0.94*	0.90**	**0.84****	**0.80****	0.83**	**0.75*****	**0.81*****	**0.70*****	**0.77*****	0.79***
ARDI-hard,1.28	**0.92****	**0.86*****	0.85**	**0.77*****	**0.79*****	0.81***	**0.79*****	**0.71*****	**0.77*****	0.80***
ARDI-hard,1.65	0.94**	**0.89*****	**0.83****	**0.78*****	**0.77*****	**0.82*****	**0.79*****	**0.69*****	**0.74*****	**0.74*****
ARDI-tstat,1.96	0.98	0.89**	0.87**	0.84**	0.81***	0.89**	0.85**	**0.74*****	0.85**	0.83***
ARDI-DU	0.92**	**0.88*****	0.84**	0.82**	0.82***	0.82***	**0.82*****	**0.72*****	0.85**	0.85**
3PRF	0.93**	0.93**	0.94**	0.92**	0.94**	0.86***	0.89***	0.91**	0.94**	0.95
Factor-structure-based models										
FAVARI	**0.92****	**0.88*****	0.85***	0.86***	0.86***	**0.80*****	**0.82*****	**0.78*****	0.85***	0.84***
FAVARD	**0.91****	**0.90****	0.87**	0.87**	0.84**	**0.79*****	**0.82*****	**0.74*****	0.84**	0.83**
FAVARMA-FMA	0.94**	0.91**	0.85***	0.84***	**0.82*****	0.81***	**0.82*****	**0.76*****	**0.80*****	**0.78*****
FAVARMA-FAR	0.98	0.97	0.94	0.96	1	0.83***	**0.78*****	**0.72*****	**0.75*****	0.84***
DFM	**0.92*****	**0.90*****	0.85***	0.85***	0.86***	0.84***	0.88***	0.82***	0.86***	0.88***
Data-rich model averaging										
CSR,1	0.98**	0.99	0.96**	0.96***	0.96***	0.98	1.05	0.99	0.98***	0.97***
CSR,10	0.92***	**0.88*****	**0.83*****	0.81***	0.81***	0.86***	0.88***	0.81***	0.84***	0.84***
CSR,20	**0.91*****	**0.86*****	**0.80*****	**0.78*****	**0.77*****	0.84**	0.83***	**0.74*****	**0.78*****	0.79***
Regularized data-rich model averaging										
T-CSR-soft,10	0.93**	**0.86*****	**0.81*****	**0.78*****	**0.78*****	0.82***	**0.82*****	**0.75*****	**0.78*****	0.79***
T-CSR-soft,20	0.99	0.92*	**0.83****	**0.80****	0.83**	0.83***	**0.79*****	**0.71*****	**0.75*****	**0.76*****
T-CSR-hard,1.65,10	**0.91****	**0.85*****	**0.80*****	**0.78*****	**0.76*****	0.82***	**0.81*****	**0.73*****	**0.79*****	**0.76*****
T-CSR-hard,1.65,20	0.94*	0.89***	0.84**	0.82**	0.82**	0.83**	**0.83*****	**0.74*****	0.82**	**0.79*****
R-CSR,10	**0.92*****	**0.88*****	0.84***	0.82***	0.80***	0.86***	0.87***	0.80***	0.82***	0.82***
R-CSR,20	**0.90*****	**0.85*****	**0.80*****	**0.77*****	**0.76*****	**0.81*****	**0.81*****	**0.74*****	**0.77*****	**0.76*****
Lasso	1.08*	1.04	0.94	0.88	0.93	0.88*	**0.82****	**0.74*****	**0.77****	**0.79****
Forecast combinations										
AVRG	**0.90*****	**0.85*****	**0.80*****	**0.78*****	**0.77*****	0.81***	**0.82*****	**0.74*****	**0.78*****	0.80***
Median	**0.90*****	**0.85*****	**0.80*****	**0.78*****	**0.77*****	0.81***	0.82***	**0.74*****	0.80***	0.80***
T-AVRG	**0.90*****	**0.85*****	**0.80*****	**0.78*****	**0.77*****	0.82***	0.82***	0.75***	**0.80*****	0.80***
IP-AVRG,1	**0.90*****	**0.85*****	**0.80*****	**0.77*****	**0.76*****	0.82***	**0.82*****	**0.73*****	**0.78*****	0.79***
IP-AVRG,0.95	**0.90*****	**0.86*****	**0.80*****	**0.78*****	**0.77*****	0.81***	**0.82*****	**0.74*****	**0.79*****	0.80***

Note. The numbers in the table are the relative RMSPE of each model with respect to the ARD model. The RMSPE of the ARD model is also indicated to assess the importance of errors. Models that are retained in the MCS are indicated in bold. The best models (with minimum relative MSPE) are underlined. Asterisks indicate ***1%, **5%, and *10% significance levels for the Diebold–Mariano test.

Two messages emerge from these results. First, data-rich models and forecast combinations dominate standard time series models when it comes to predicting industrial production growth. Second, the fact that several models belonging to different categories are jointly present in the MCS naturally explains why forecast combinations perform so well.

4.2 | Employment growth

We now examine the results for employment growth, presented in Table 3. The results are quite similar to what is obtained for industrial production growth. As previously, standard time series model are dominated and are never selected in the MCS.

Over the full out-of-sample period, the best models to predict employment growth often belong to RDRMA, while the MCS contains many versions of forecast combinations. Models involving factors are much less present in the MCS than previously. During recessions, the best models and the MCS are almost evenly distributed between factor-augmented regressions and RDRMA. Factor-structure-based models emerge as the best at short horizons during recession.

TABLE 3 Employment: relative MSPE

Model	Full out-of-sample h = 1	h = 3	h = 6	h = 9	h = 12	NBER recessions periods h = 1	h = 3	h = 6	h = 9	h = 12
Standard time series models										
ARD (RMSPE)	0.0184	0.0152	0.0158	0.0163	0.0167	0.0245	0.0242	0.0261	0.0265	0.0256
ARI	1	0.99	0.98*	0.98	1.01	1	1.01	0.98	0.99	1
ARMA(1, 1)	1	0.99	1	1.04	1.07	1.01	1.05**	0.99	0.99	1
ADL	0.99	1.01	1.03*	1.02	1.01	0.98	1.03	1.05**	1.02	1
Factor-augmented regressions										
ARDI	0.94*	0.91*	0,93	0,94	0.88**	**0.84****	**0.84****	0.86**	0.91*	**0.84****
ARDI-soft	1.02	**0.88***	0.95	**0.79*****	**0.83*****	0.94	**0.85**	**0.83****	**0.75*****	**0.83****
ARDI-hard,1.28	0.95	**0.88****	**0.87****	0.84***	0.87**	**0.84****	**0.84***	**0.83****	**0.82****	0.87*
ARDI-hard,1.65	**0.93***	0.89**	**0.87****	0.85**	0.87**	**0.82****	0.84*	0.86*	**0.84****	0.86**
ARDI-tstat,1.96	0.96	**0.88****	0.89**	0.86***	0.85***	**0.90***	**0.82****	**0.78*****	0.85***	0.85**
ARDI-DU	0.94*	0.91*	0.92	0.89*	0.89*	**0.87****	0.87*	0.87*	0.88*	0.84**
3PRF	1	0.97	0.98	0.99	1	**0.91***	0.98	1.02	1.05*	1.05
Factor-structure-based models										
FAVARI	0.94	0.92	0.94	0.97	0.98	**0.82****	0.97	1.03	1.06	1.04
FAVARD	**0.93***	0.89*	0.89**	0.90**	0.89**	**0.81****	0.89*	0.89*	0.96	0.94
FAVARMA-FMA	**0.93***	0.91*	0.92*	0.95	0.93*	**0.83****	0.96	0.97	1.01	0.98
FAVARMA-FAR	0.95	0.92	0.96	0.99	1	0.91	0.97	1.04	1.08*	1.11**
DFM	0.96	0.91*	**0.88*****	0.87***	0.88***	0.96	0.98	0.97	0.93**	0.90***
Data-rich model averaging										
CSR,1	1.06***	1.03	0.99	0.99	0.99	1.13***	1.11***	1.04**	1.01	0.98
CSR,10	0.97	0.90**	0.87***	0.86***	0.86***	0.98	0.95	0.91**	0.90**	0.87**
CSR,20	0.95*	**0.85*****	**0.84*****	0.83***	0.84***	0.90*	0.87**	0.85***	0.87**	0.84**
Regularized data-rich model averaging										
T-CSR-soft,10	0.96	**0.85*****	**0.85*****	**0.80*****	0.83***	0.91	0.87**	0.86***	0.84***	0.83***
T-CSR-soft,20	0.98	**0.85****	**0.85****	**0.81*****	0.85**	**0.89**	**0.82****	**0.83*****	**0.81*****	**0.77*****
T-CSR-hard,1.65,10	0.93*	**0.85*****	**0.83*****	0.83***	0.85***	0.88**	0.86**	0.85**	0.88**	0.87**
T-CSR-hard,1.65,20	0.95	**0.83****	**0.85****	0.86**	0.90*	**0.85****	**0.81****	0.84**	0.89*	0.91
R-CSR,10	**0.93*****	**0.86*****	**0.85*****	0.84***	0.83***	**0.88*****	0.85***	0.84***	0.85***	0.83***
R-CSR,20	0.93**	**0.83*****	**0.82*****	**0.80*****	**0.79*****	**0.85****	**0.81*****	**0.80*****	**0.82*****	**0.79*****
Lasso	1.07*	0.92	0.91	0.89*	0.96	1	**0.83***	0.87**	0.85**	**0.79****
forecast combinations										
AVRG	**0.91*****	**0.84*****	**0.83*****	**0.82*****	**0.82*****	**0.85****	0.86**	0.87***	0.86***	0.85***
Median	**0.91****	**0.83*****	**0.83*****	**0.82*****	0.83***	**0.86****	0.86**	0.86***	0.87***	0.85***
T-AVRG	**0.91****	**0.84*****	**0.84*****	0.82***	0.83***	**0.85****	0.87**	0.87***	0.87***	0.86***
IP-AVRG,1	**0.91****	**0.83*****	**0.83*****	**0.81*****	**0.82*****	**0.85****	0.85**	0.85***	0.85***	0.85***
IP-AVRG,0.95	**0.91*****	**0.83*****	**0.83*****	**0.82*****	0.82***	**0.85****	0.85**	0.86***	0.86***	0.85***

Note. see note to Table 2.

In summary, RDRMA is a robust approach to forecasting real series irrespective of whether we are in recession or not. The actual magnitudes of those improvements can be inferred from the root MSPE that is reported for the reference model ARD. For example, using the Ridge CSR,20 model to predict industrial growth 1 year ahead increases the forecast accuracy by 120 basis points (3.85%) over the benchmark (5.05%), which is an economically significant improvement. In case of the employment growth, the same model decreases the RMSPE by 35 basis point (1.32% against 1.67%).

Forecast combinations perform quite well on average but they may be outperformed by factor-augmented or factor-structure-based models during recessions. A researcher who only cares about the average performance of his model at forecasting a real series should consider using either RDRMA or forecast combination. By contrast, a researcher who cares more about the performance of his forecasting model during recession (i.e., when uncertainty and instabilities are higher than usual) should rather use factor-augmented regressions.

4.3 | CPI inflation

We now examine the performance of the various models at forecasting variations of CPI inflation. The target of interest here is therefore the second difference of the logarithm of CPI (i.e., CPI acceleration). Table 4 shows the results.

Over the whole out-of-sample period, ARMA(1, 1) dominates all individual data-rich forecasting models at short horizons. At 9 months horizon and beyond, RDRMA emerges as the best forecasting model but its performance is not significantly different from ARMA(1, 1). During recessions, the ARMA(1, 1) model still performs well at short horizons but the targeted ARDI performs better at longer horizons. In terms of actual magnitudes, the predictive accuracy for CPI inflation change is very good. For the full sample and 1-year horizon, the R-CSR,20 model improves the forecast precision by 29 basis points (1.94%) over the ARD model (2.23%). Beyond the statistical significance, this amelioration is particularly valuable for monetary policy authorities that require accurate inflation forecasts (anticipations).

Few studies document the performance of ARMA models at predicting inflation. Stock and Watson (2007) suggested that the MA component of the inflation process had increased since 1984. Ng and Perron (Ng and Perron (1996), Ng and Perron (2001) also documented similar evidence. Foroni et al. (2019) found that the presence of an MA component improved the forecasting power of mixed-frequency models when predicting US inflation.

One plausible explanation for the good performance of ARMA(1, 1) is that inflation is generally well anticipated, so that its variations behave like an exogenous noise. Consequently, data-rich models tend to be overparametrized and have poor predictive performance for this series. During recessions, economic variables are subject to unusually large shocks,

TABLE 4 CPI inflation: relative RMSPE

	Full out-of-Sample					NBER recessions periods				
Model	$h = 1$	$h = 3$	$h = 6$	$h = 9$	$h = 12$	$h = 1$	$h = 3$	$h = 6$	$h = 9$	$h = 12$
Standard time series models										
ARD (RMSPE)	0.0318	0.0279	0.0232	0.0217	0.0223	0.0493	0.0473	0.035	0.0294	0.0277
ARI	1.00	1.05*	1.16***	1.19***	1.18**	1	1.09	1.27**	1.04*	1.01
ARMA(1, 1)	0.94**	0.89**	0.93	0.95	0.93**	0.94	0.87**	0.99	0.98	1.01
ADL	1.02	1.06	1.21	1.06	0.99	1.05	1.06	0.88*	0.88	0.98
Factor-augmented regressions										
ARDI	1	1.08	1.12	1.01	0.90**	0.95	1.12	0.92**	0.91*	0.91*
ARDI-soft	0.96	1.13*	1.1	1.01	0.94	0.89*	1.14	0.95	0.91	0.86**
ARDI-hard,1.28	1.01	1.06	1.11	1.02	0.91*	0.93	1.1	0.98	0.85*	0.80**
ARDI-hard,1.65	1.02	1.07	1.06	1.05	0.94	0.96	1.14	0.89*	0.84*	0.77***
ARDI-tstat,1.96	1.01	1.02	1.02	0.98	0.92**	1.00	1.03	0.96	0.95	0.92
ARDI-DU	1.00	1.06	1.16	1.03	0.93*	0.96	1.10	0.95	0.93	0.9
3PRF	1.07**	1.14***	1.21***	1.18***	1.14***	1.03	1.14	1.27***	1.07*	1.08
Factor-structure-based models										
FAVARI	1.06*	1.20***	1.50***	1.65***	1.70***	0.97	1.16	1.74**	1.43**	1.49*
FAVARD	1.06*	1.18***	1.47***	1.62***	1.73***	0.96	1.14	1.73**	1.33**	1.35*
FAVARMA-FMA	1.05	1.17***	1.47***	1.62***	1.64***	0.96	1.13	1.69**	1.40**	1.43*
FAVARMA-FAR	1.21***	1.75***	2.73***	3.57***	3.99***	1.02	1.56**	2.72***	2.68***	3.05***
DFM	0.98	1.03	1.16*	1.26*	1.29*	0.94	1.04	1.36*	1.03	1.03
Data-rich model averaging										
CSR,1	1.03	1.11**	1.25***	1.27***	1.24***	0.95	1.05	1.40**	1.12	1.11*
CSR,10	1.01	1.11**	1.23***	1.22***	1.18**	0.92	1.08	1.36**	1.07	1.05
CSR,20	1.02	1.11**	1.26***	1.20***	1.16**	0.93	1.09	1.44**	1.07	1.03
Regularized data-rich model averaging										
T-CSR-soft,10	0.97	1.07*	1.16***	1.16**	1.11*	0.87	1.01	1.22**	1.02	1
T-CSR-soft,20	1.00	1.11**	1.20***	1.19***	1.13**	0.87	1.00	1.13***	1.03	1.04
T-CSR-hard,1.65,10	1.01	1.09**	1.17***	1.16**	1.12**	0.92	1.03	1.24**	1	0.95
T-CSR-hard,1.65,20	1.03	1.10**	1.17***	1.12**	1.11**	0.92	0.96	1.16**	1.01	0.93
R-CSR,10	0.96**	0.97	1.00	0.94**	0.88***	0.91**	0.98	0.95	0.91**	0.90*
R-CSR,20	0.95*	0.97	1.03	0.94**	0.87***	0.88*	0.98	0.96	0.88*	0.89*
Lasso	1.09*	1.15***	1.19***	1.10*	1.06	0.90	1.02	0.96	0.92	1.04
Forecast combinations										
AVRG	0.94**	0.96	1.01	1.01	0.98	0.87**	0.95	1.09	0.95	0.94
Median	0.95*	0.96	0.99	0.96	0.93*	0.87**	0.98	1.08	0.92*	0.91*
T-AVRG	0.94**	0.96	0.99	0.98	0.93*	0.87**	0.96	1.06	0.93*	0.91*
IP-AVRG,1	0.94**	0.96*	0.99	0.96	0.91**	0.87**	0.95	1.02	0.92*	0.90*
IP-AVRG,0.95	0.94**	0.96*	0.98	0.95**	0.89***	0.87**	0.96	1.01	0.91*	0.88**

Note. see note to Table 2.

and the stability of the relationship that bound variables is not warranted. As a result, the ARMA(1, 1) model loses its predictive power and data-rich models become favored.

4.4 | Stock market index

We now examine the results for the S&P 500 returns. In principle, a forecasting model for stock market returns should include the real-time vintages of the predictors. Unfortunately, these vintages are not available for a large number of predictors. Our models are therefore based on the latest information available on all predictors. Table 5 shows the results.

Under the assumption of market efficiency, random walk models have become the standard benchmark in the literature on return predictability. Indeed, stock market returns are said to be predictable if one can find a model that forecasts them

TABLE 5 S&P 500: relative RMSPE wrt RW

	Full out-of-Sample					**NBER recession periods**				
Model	$h=1$	$h=3$	$h=6$	$h=9$	$h=12$	$h=1$	$h=3$	$h=6$	$h=9$	$h=12$
Random walks										
RW (RMSPE)	0.451	**0.3069**	**0.2370**	**0.2007**	**0.1785**	0.7126	**0.4780**	<u>**0.3313**</u>	**0.2719**	<u>**0.2329**</u>
RWD	0.99	**0.99**	**0.98**	**0.98**	**0.98**	1.01**	1.05***	1.11***	1.18***	1.22***
Standard time series models										
ARD	**0.97*****	**0.98**	**0.98**	**0.99**	**0.98**	0.98	1.05***	1.10***	1.15***	1.19***
ARI	**0.97*****	**0.98**	**0.98**	**0.98**	**0.97**	0.98	1.04**	1.09***	1.14***	1.19***
ARMA(1, 1)	0.98*	**0.99**	**0.99**	**0.98**	**0.98**	0.99	1.05***	1.10***	1.15***	1.19***
ADL	**0.98**	1.01	**1.02**	**0.98**	**0.99**	**0.95**	1.06	1.10***	1.14***	1.18***
Factor-augmented regressions										
ARDI	**0.96****	**0.99**	1.04	1.06	1.03	<u>**0.94***</u>	**1.01**	1.09	1.09	1.13
ARDI-soft	1	1.03	1.12	1.04	1.07	1	1.06	1.07	1.06	1.21**
ARDI-hard,1.28	1	**1.00**	1.09	1.07	1.01	1	**1.02**	1.1	1.12	1.20*
ARDI-hard,1.65	0.99	**1.00**	1.11	1.07	1.01	0.98	**1.02**	1.09	1.09	1.19*
ARDI-tstat,1.96	0.99	**1.00**	1.07	1.06	**1.00**	**0.96**	**1.02**	1.11*	1.12	1.16*
ARDI-DU	**0.96****	**0.99**	**1.04**	1.05	1.01	**0.95***	1.03	1.1	1.12	**1.12**
3PRF	**0.97***	**0.99**	**1.03**	**1.03**	1.02	**0.96**	1.04	**1.03**	1.02	**1.12***
Factor-structure-based models										
FAVARI	**0.98**	**0.99**	**1.01**	**1.04**	1.04	**0.98**	**1.01**	1.05	1.01	**1.04**
FAVARD	**0.98**	**1.00**	1.06	1.1	1.07	**0.97**	**1.02**	1.05	**0.96**	**1.04**
FAVARMA-FMA	**0.98**	**0.98**	**1.02**	**1.05**	1.05	**0.97**	**1.01**	**1.04**	1.01	**1.07**
FAVARMA-FAR	0.99	1.05	1.11*	1.16*	1.18*	0.99	1.14**	1.24*	1.11*	**1.10***
DFM	**0.96****	**0.98**	**0.99**	**0.99**	**0.98**	**0.96***	**1.02**	1.05	1.07	1.12**
Data-rich model averaging										
CSR,1	**0.96*****	**0.98***	<u>**0.98**</u>	**0.97**	**0.97**	**0.97***	**1.03***	1.08***	1.14***	1.18***
CSR,10	**0.96****	**0.97**	**1.00**	**0.99**	**0.97**	0.97	**1.00**	1.05	1.10*	1.16**
CSR,20	0.99	1.01	1.07	1.05	1	1	1.04	1.1	1.19*	1.19**
Regularized data-rich model averaging										
T-CSR-soft,10	1	**1.00**	**1.05**	**1.04**	1.03	1.02	**1.00**	**1.01**	1.06	1.17**
T-CSR-soft,20	1.10***	1.11*	1.23	1.18*	1.14*	1.13**	1.05	**1.02**	**1.01**	1.15
T-CSR-hard,1.65,10	0.98	1	**1.06**	**1.04**	1	0.99	**1.00**	**1.02**	1.04	1.17*
T-CSR-hard,1.65,20	1.01	1.06	1.16*	1.16*	1.10*	1.02	**1.02**	**1.02**	1.01	1.15*
R-CSR,10	<u>**0.95*****</u>	<u>**0.97***</u>	**0.98**	<u>**0.96**</u>	<u>**0.95**</u>	**0.96***	**1.00**	1.04	1.06	1.13**
R-CSR,20	**0.96****	**0.98**	**1.02**	**0.98**	**0.97**	0.97	**1.00**	**1.02**	1.04	1.13*
Lasso	1.26***	1.32***	1.45**	1.46**	1.33***	1.29***	1.24*	1.15	<u>**0.94**</u>	**1.11**
Forecast combinations										
AVRG	**0.96****	**0.97**	**1.00**	**0.98**	**0.96**	0.96	<u>**0.99**</u>	**1.02**	1.03	1.12*
Median	**0.96****	**0.97**	**1.00**	**0.99**	**0.96**	**0.96**	**1.00**	1.03	1.05	1.12*
T-AVRG	**0.96****	**0.97**	**1.00**	**0.98**	**0.96**	**0.96**	**1.00**	1.03	1.04	1.12*
IP-AVRG,1	**0.96****	**0.97**	**1.00**	**0.99**	**0.97**	**0.96**	**0.99**	**1.02**	1.04	1.13*
IP-AVRG,0.95	**0.96****	**0.97**	**1.00**	**1.00**	**0.97**	**0.96**	**1.00**	**1.02**	1.04	1.13*

Note. The numbers in the table are the relative RMSPE of each model with respect to the RW model. The RMSPE of the RW model is also indicated to assess the importance of errors. Models that are retained in the MCS are indicated in bold. The best models (with minimum relative RMSPE) are underlined. Asterisks indicate ***1%, **5%, and *10% significance levels for the Diebold–Mariano test.

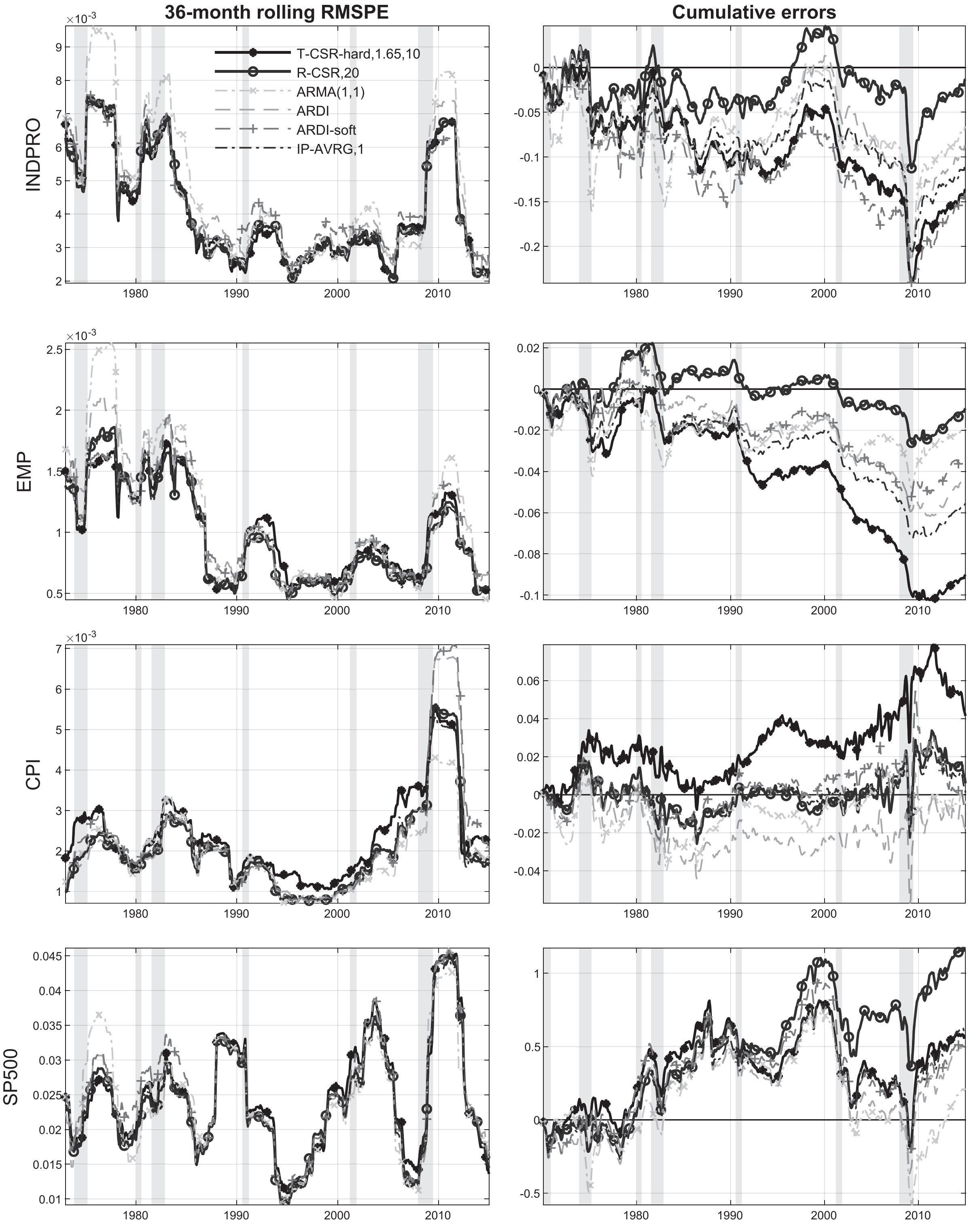

FIGURE 1 RMSPE over time. The figure shows the 3-year moving average of the RMSPE of selected models for $h = 3$, and the cumulated forecast errors

better than random walk models. Therefore, we need to consider the random walk model with or without drift (RWD and RW) as the benchmarks for the SP& 500 returns.

Over the full out-of-sample period, our RDRMA generates the best point forecasts at most horizons. Table 5 shows that the R-CSR,10 specification improves up to 5% and 3% with respect to RW at 1 and 3 months forecasting horizon, respectively. It also dominates at longer horizons, but the forecasts are not statistically different according to the

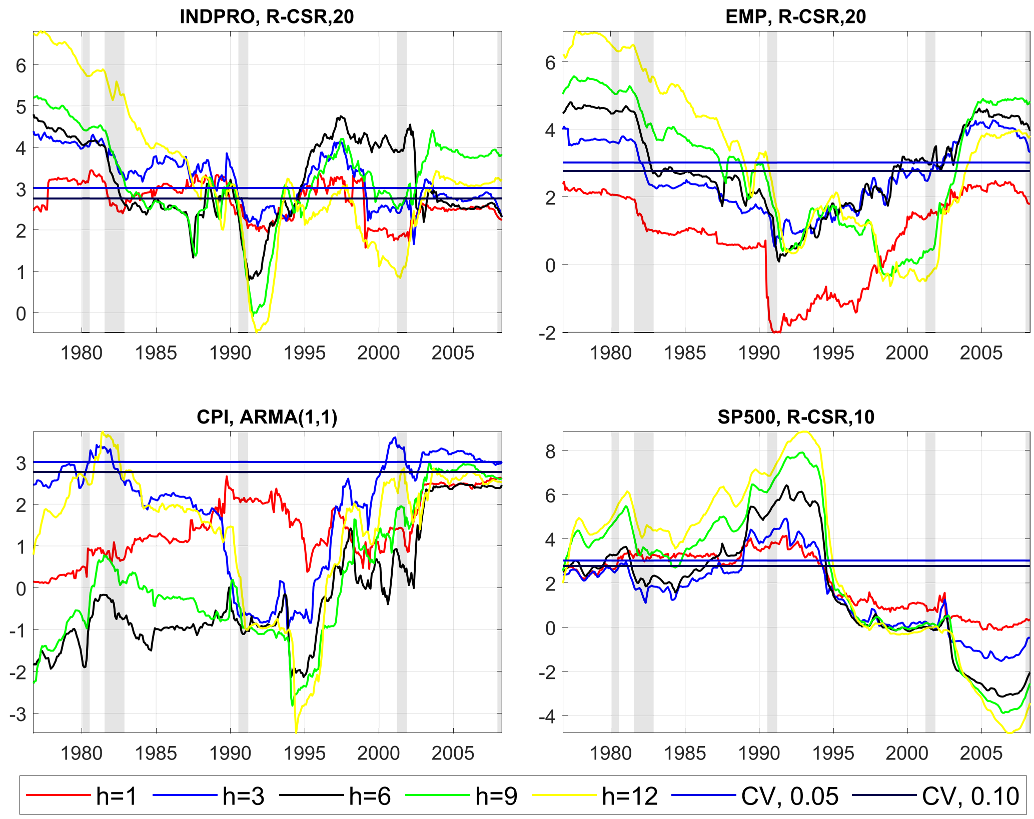

FIGURE 2 Giacomini–Rossi fluctuation test for best RMSPE models against the ARD benchmark, except for S&P 500, where the reference model is RW. CV, 0.05 and CV, 0.10 correspond to 5% and 10% critical values, respectively [Colour figure can be viewed at wileyonlinelibrary.com]

Diebold–Mariano test. During recessions, a factor-augmented regression does better than random walk models at the shortest horizon ($h = 1$). Finally, RWD outperforms RW in general, but the latter model dominates significantly during recessions.

The results above support the claim that stock returns are predictable only at short horizons. Further results presented in the Appendix suggest that the RCSF is higher for most models than for the RW at most horizons. This implies that a nonlinear predictability of stock returns is still possible at longer horizons.

5 | STABILITY OF FORECAST ACCURACY

In this section we examine the stability of the forecast accuracy and of the optimal structure of the forecasting equations over time.

5.1 | Stability of forecast performance

Here we examine the stability of the forecast accuracy.[6] Figure 1 plots the 3-year moving average of the RMSPE of selected models for $h = 3$-month-ahead forecasts, as well as the cumulated forecast errors. The selected models are two of our RDRMA techniques, and the alternative models showed the best overall performance in the horse race. In the left-hand column of the figure we see a significant downturn in the level of RMSPE for real activity series from the mid-1980s, which coincides with the Great Moderation period. There are also systematic shifts during recessions, those around the oil price shocks, Great inflation, and Great Recession being by far larger compared to 1991 and 2001 downturns. These changes in volatility are in line with macroeconomic uncertainty dynamics in Jurado et al. (2015). In the case of CPI inflation change, we note a slow downward trend since 1982 that vanished at the beginning of the 1990s, which coincides with the inflation

[6]See Giacomini and Rossi (2009) and Rossi and Sekhposyan (2010, 2011), among others, for examples of time-varying forecast performance.

targeting regime. As suggested by Boivin and Ng (2006), monetary policy became more aggressive in stabilizing economic activity, which also resulted in more anchored inflation expectations. Hence the volatility of inflation predictions shrunk during that period. However, it started rising from 2000 and skyrocketed to historical peaks during the Great Recession. It has dropped back to the normal level since then. The dynamics of S&P 500 returns RMSPE is closely related to NBER cycles.

Despite the large swings in the absolute measure of forecasting performance, it turns out that the RMSPE trajectories have rather parallel trends, meaning that the relative performance of any two models is quite stable over time (exceptions may be observed during recession episodes). For real variables, at least one of our RDRMA models regularly produces lower RMSPEs compared to the alternatives. In the case of inflation, our R-CSR model is close to ARMA(1, 1), while all models have similar performance when predicting S&P 500 returns.

The right-hand column plots the cumulated forecast errors across the out-of-sample period. The R-CSR model is undoubtedly the least biased when predicting industrial production and employment growths, and has similar performance to ARMA for CPI inflation change. All models underestimate the level of stock returns during the Great Moderation.

Giacomini and Rossi (2010) proposed a test to compare the out-of-sample forecasting performance of two competing models in the presence of instabilities. Figure 2 shows the results for several horizons and two critical values. We report the comparison between the overall best RMSPE model for each series and the ARD alternative, except for the S&P 500, where the reference model is RW. Following the Monte Carlo results in Giacomini and Rossi, the moving average of the standardized difference of MSPEs is produced with a 162-month window, which corresponds to 30% of the out-of-sample period. The results point to some instability in the forecast accuracy, but the relative performance of our models is still very good most of the time.

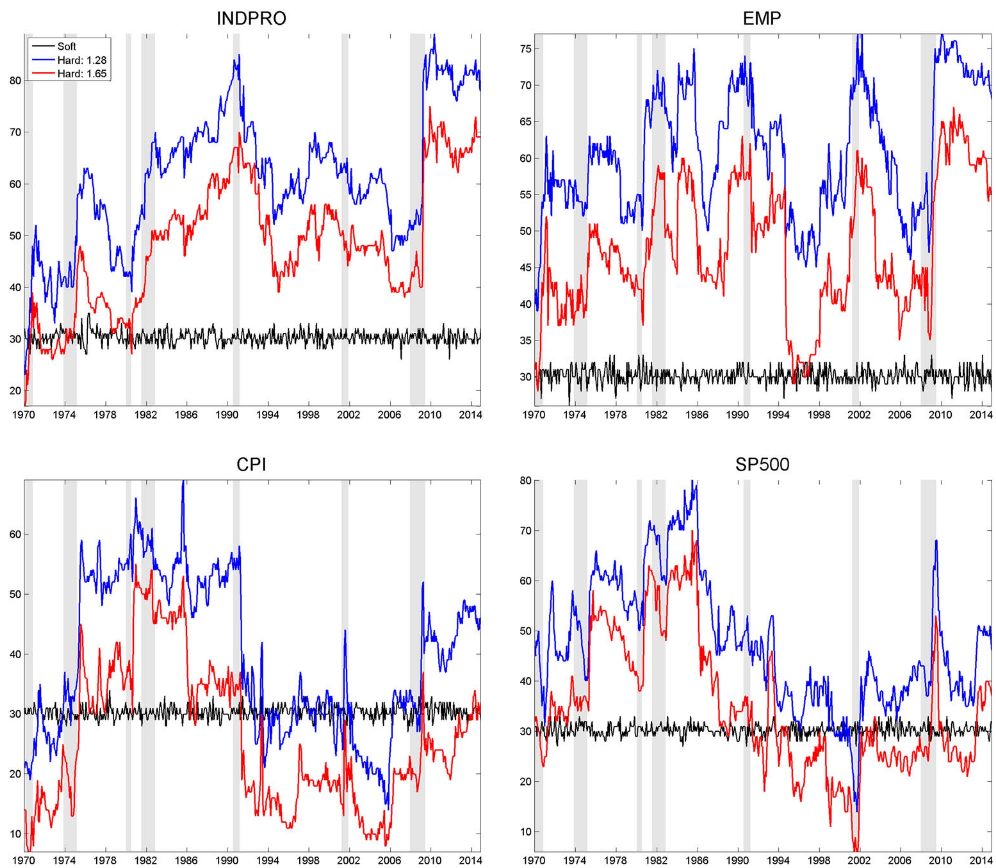

FIGURE 3 Number of series preselected by hard and soft thresholding when predicting at the 3-month horizon [Colour figure can be viewed at wileyonlinelibrary.com]

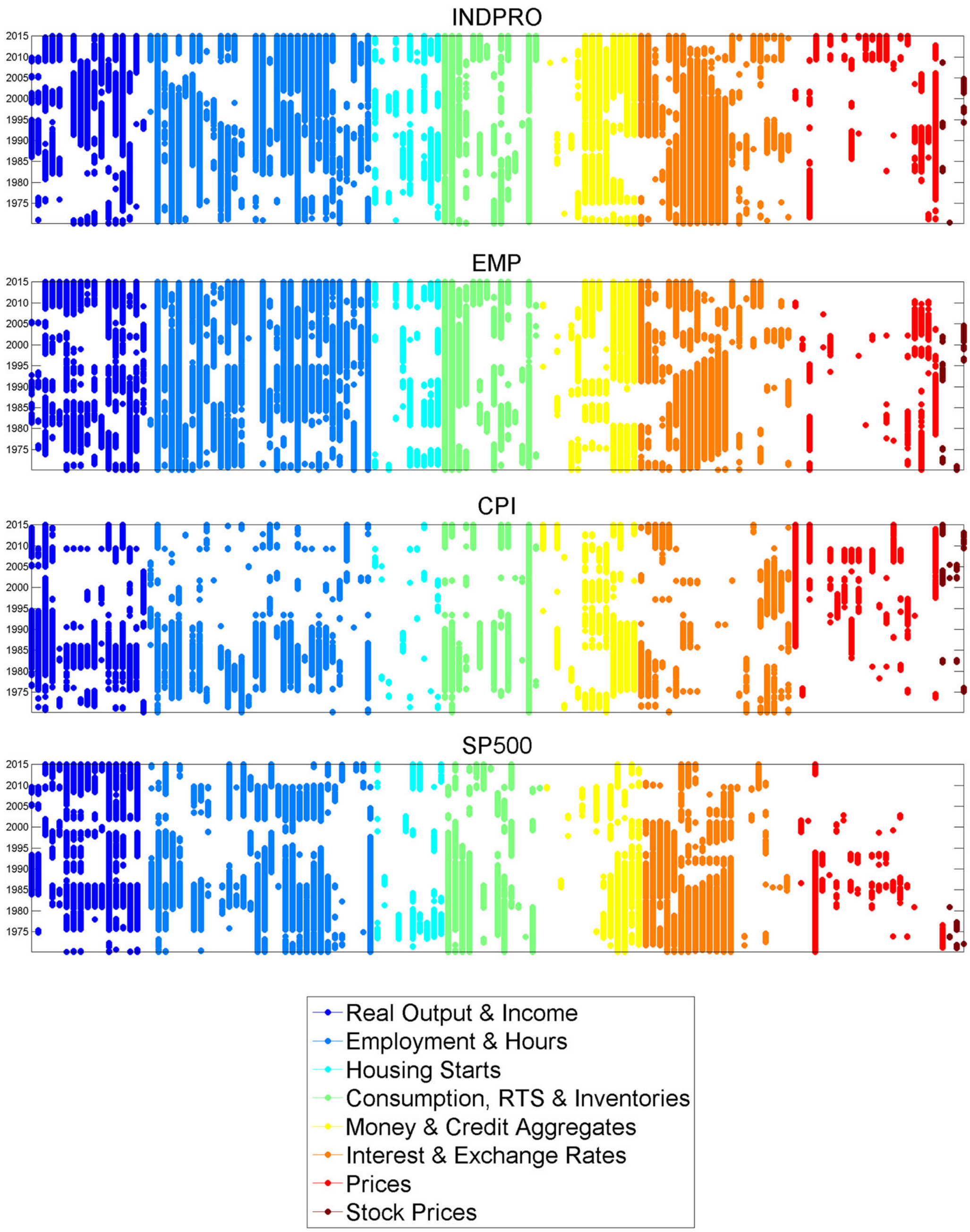

FIGURE 4 Series preselected by hard thresholding with $t_c = 1.65$ when predicting at the 3-month horizon. The content of each group is described in McCracken and Ng (2016) [Colour figure can be viewed at wileyonlinelibrary.com]

5.2 | Stability of forecast relationships

Several recent studies have suggested that factor loadings and the number of factors are likely to change over time.[7] The results presented here point towards the same direction. The number of principal components retained in factor-augmented models varies considerably across the out-of-sample period, forecasting horizons, and the series of interest. In general, real variables require more factors in the forecasting equations than inflation or stock market returns.[8]

Figure 3 plots the number of series selected by soft (Lasso) and hard thresholds for all series at the 3-month horizon. Recall that this is the first step in ARDIT models as well as in our targeted CSR model. The results are similar for the two

[7]See, among others, Breitung and Eickmeier (2011), D'Agostino et al. (2013), Eickmeier et al. (2015) and Carriero et al. (2015), Cheng et al. (2016), Mao Takongmo and Stevanovic (2015), and Guerin et al. (2016).
[8]Owing to space constraints, the related figures are presented in the Supporting Information.

real activity series. The number of candidate predictors is generally lower when predicting CPI inflation growth. In the case of stock returns, the number of selected series is declining until the Great Recession.

Figure 4 shows the type of series selected by hard thresholding with $t_c = 1.65$ for 3-month-ahead predicting. We group the data as in McCracken and Ng (2016) and show whether a series has been selected or not over the whole out-of-sample period. The probability that a particular predictor will be consistently selected is higher for some groups and depends on the series being predicted. For instance, indicators in Employment & Hours, Consumption, and Money & Credit groups are often present when predicting industrial production and employment. There is a lot of instability in predictor selection for CPI where only a small number of candidates are systematically present. A similar pattern is observed in the case of S&P 500. However, even if a single predictor may appear to be randomly selected, we note that categories of predictors are in general well represented over time, as in De Mol et al. (2008). Those variations support our RDRMA technique, which relies on regularization to smartly combine the relevant information that is likely to vary over time.

Given this historical evidence on structural instability in forecasting models and predictive accuracy, we believe our RDMRA models are likely to continue to perform well because of two important features. First, they rely on model averaging, which is known to improve forecasting performance.[9] Second, regularization makes the set of all models, to be averaged, more robust to structural changes. In the targeted CSR, the targeting in the first step provides a more efficient and less restrictive framework than keeping the set of predictors fixed for every variable and horizon Bai and Ng (2008). This preselection works in a similar fashion as model weighting where Del Negro et al. (2016) and Elliott and Timmermann (2005) show that allowing weights to change improves the forecasting performance. The ex post regularization in the second model, the Ridge-CSR, shrinks the coefficients of uninformative predictors towards zero to avoid overfitting, which in turn reduces the instability in model predictions Fan and Li (2001). This implicit weighting (ex ante or ex post) is exactly the source of improvement upon the original CSR model. A combination with data-rich model averaging provides a very robust and flexible model that is likely to continue performing well in the future, despite the changing economic environments.

6 | CONCLUSION

This paper adds the RDRMA technique to the list of predictive models in the context of a data-rich environment. We compare its performance to five classes of forecasting models on different macroeconomic series in an extensive out-of-sample exercise. The series considered are industrial production growth, employment growth, inflation growth, and S&P 500 returns. The comparison of the models is based on their pseudo out-of-sample performance. For each series, horizon, and out-of-sample period, the hyperparameters of our models (number of lags, number of factors, etc.) are recalibrated using the BIC.

Considering the growth rate of real series, we find that RDRMA and forecast combinations deliver the best forecasts in terms of RMSPE over the full out-of-sample period. During recessions, factor structure-based and factor-augmented models deliver the best performance, although forecast combinations and RDRMA are still often selected into the MCS during recessions. Univariate models are largely dominated. We therefore conclude that RDRMA and forecast combinations are robust approaches to predict real series.

The ARMA(1, 1) model delivers incredibly good forecasts in terms of RMSPE for inflation growth at a quite moderate cost. The best data-rich or forecast combination approach does not outperform the ARMA(1, 1) model. During recessions, factor-augmented regression may outperform the ARMA(1, 1) model at horizons beyond 3 months.

Considering the S&P 500 returns, RDRMA delivers the best forecasts 1 month ahead in terms of RMSPE. At longer horizons, data-rich model averaging (regularized or not) delivers the lost RMSPE but the MCS encompasses random walk models. During recessions, factor-augmented models outperform random walk models only at the 1 month horizon. At longer horizons, random walk models are selected again into the MCS. Random walk models are dominated at all horizons in terms of RCSF.

Overall, RDRMA and forecast combinations emerge as robust forecast approaches when the performance evaluation metric is the RMSPE. Thus our results suggest that sparsity and regularization can be smartly combined with model averaging to obtain forecasting models that dominate state-of-the-art benchmarks.

[9]See Bates and Granger (1969), Hendry and Clements (2004), and Elliott et al. (2015) for theoretical and empirical demonstrations, and Boot and Nibbering (2019) for a theoretical derivation of expected gains of the CSR.

Finally, we examine the stability of the forecasting equations and their performance over time. The results suggest significant time instability in the forecast accuracy as well as in the structure of the optimal forecasting equations over time. However, our RDRMA models are flexible enough to adapt to those structural changes and maintain very good relative predictive performance.

ACKNOWLEDGMENTS

We thank Mehmet Caner, Todd Clark, Marco Del Negro, John Galbraith, Eric Ghysels, Domenico Giannone, Serena Ng, Barbara Rossi, Frank Schorfheide, participants of the Penn Big Data Conference and anonymous referees for valuable comments. Stevanovic acknowledges financial support from the Fonds de recherche sur la société et la culture (Qué bec) and the Social Sciences and Humanities Research Council.

OPEN RESEARCH BADGES

This article has earned an Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at [http://qed.econ.queensu.ca/jae/datasets/kotchoni001/].

REFERENCES

Bai, J., & Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, *146*, 304–317.

Banbura, M., Giannone, D., & Reichlin, L. (2010). Large Bayesian vector autoregressions. *Journal of Applied Econometrics*, *25*, 71–92.

Banerjee, A., Marcellino, M., & Masten, I. (2014). Forecasting with factor-augmented error correction models. *International Journal of Forecasting*, *30*(3), 589–612.

Barigozzi, M., Lippi, M., & Luciani, M. (2016). *Non-stationary dynamic factor models for large-datasets*. Washington, DC: Board of Governors of the of Federal Reserve System. (FEDS Working Paper No. 2016-024).

Bates, J., & Granger, C. W. J. (1969). The combination of forecasts. *the Operational Research Society*, *20*(4), 451–468.

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

Boivin, J., & Giannoni, M. (2006). Has monetary policy become more effective? *The Review of Economics and Statistics*, *88*(3), 445–462.

Boivin, J., & Ng, S. (2005). Understanding and comparing factor-based forecasts. *International Journal of Central Banking*, *1*, 117–151.

Boivin, J., & Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, *132*, 169–194.

Boot, T., & Nibbering, D. (2019). Forecasting using random subspace methods. *Journal of Econometrics*, *209*(2), 391–406.

Breitung, J., & Eickmeier, S. (2011). Testing for structural breaks in dynamic factor models. *Journal of Econometrics*, *163*(1), 71–84.

Carrasco, M., & Rossi, B. (2016). In-sample inference and forecasting in misspecified factor models. *Journal of Business & Economic Statistics*, *34*(3), 313–338.

Carriero, A., Clark, T., & Marcellino, M. (2015). Bayesian VARs: Specification choices and forecast accuracy. *Journal of Applied Econometrics*, *30*, 46–73.

Cheng, X., & Hansen, B. E. (2015). Forecasting with factor-augmented regression: A frequentist model averaging approach. *Journal of Econometrics*, *186*(2), 280–293.

Cheng, X., Liao, Z., & Schorfheide, F. (2016). Shrinkage estimation of high-dimensional factor models with structural instabilities. *The Review of Economic Studies*, *83*(4), 1511–1543.

D'Agostino, A., Gambetti, L., & Giannone, D. (2013). Macroeconomic forecasting and structural change. *Journal of Applied Econometrics*, *28*(1), 82–101.

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

Del Negro, M., Hasegawa, R. B., & Schorfheide, F. (2016). Dynamic prediction pools: An investigation of financial frictions and forecasting performance. *Journal of Econometrics*, *192*(2), 391–405.

Diebold, F. X., & Shin, M. (2018). Machine learning for regularized survey forecast combination: Partially-egalitarian lasso and its derivatives. *International Journal of Forecasting*. doi:10.1016/j.ijforecast.2018.09.006

Dufour, J.-M., & Stevanovic, D. (2013). Factor-augmented VARMA models with macroeconomic applications. *Journal of Business & Economic Statistics*, *31*(4), 491–506.

Eickmeier, S., Lemke, W., & Marcellino, M. (2015). Classical time varying factor-augmented vector auto-regressive models estimation, forecasting and structural analysis. *Journal of the Royal Statistical Society, Series A*, *178*(3), 493–533.

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

Elliott, G., Gargano, A., & Timmermann, A. (2015). Complete subset regressions with large-dimensional sets of predictors. *Journal of Economic Dynamics and Control*, *54*, 86–110.

Elliott, G., & Timmermann, A. (2005). Optimal forecast combination under regime switching. *International Economic Review*, *46*(4), 1081–1102.

Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, *96*, 1348–1360.

Faust, J., & Wright, J. (2013). Forecasting inflation. In: G. Elliott, & A. Timmermann (Eds.). *Handbook of economic Forecasting 2A*, Netherlands: Elsevier, Amsterdam.

Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2005). The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association*, *100*, 830–839.

Foroni, C., Marcellino, M., & Stevanovic, D. (2019). Mixed-frequency models with moving-average components. *Journal of Applied Econometrics*. doi:10.1002/jae.2701

Giacomini, R., & Rossi, B. (2009). Detecting and predicting forecast breakdowns. *The Review of Economic Studies*, *76*(2), 669–705.

Giacomini, R., & Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, *25*(4), 595–620.

Giannone, D., Lenza, M., & Primiceri, G. (2015). Prior selection for vector autoregressions. *The Review of Economics and Statistics*, *97*(2), 436–451.

Giannone, D., Lenza, M., & Primiceri, G. (2017). *Macroeconomic prediction with big data: The illusion of sparsity*. Federal Reserve Bank of New York.

Groen, J. J., & Kapetanios, G. (2016). Revisiting useful approaches to data-rich macroeconomic forecasting. *Computational Statistics and Data Analysis*, *100*, 221–239.

Guerin, P., Leiva-Leon, D., & Marcellino, M. (2016). Markov-switching three-pass regression filter. (*Technical report*). Milan, Italy.

Hansen, P., Lunde, A., & Nason, J. (2011). The model confidence set. *Econometrica*, *79*(2), 453–497.

Hendry, D. F., & Clements, M. P. (2004). Pooling of forecasts. *The Econometrics Journal*, *7*(1), 1–31.

Jurado, K., Ludvigson, S., & Ng, S. (2015). Measuring uncertainty. *The American Economic Review*, *105*(3), 1177–1216.

Kelly, B., & Pruitt, S. (2015). The three-pass regression filter: A new approach to forecasting using many predictors. *Journal of Econometrics*, *186*(2), 294–316.

Kim, H. H., & Swanson, N. R. (2014). Forecasting financial and macroeconomic variables using data reduction methods: New empirical evidence. *Journal of Econometrics*, *178*(2), 352–367.

Koop, G. (2013). Forecasting with medium and large Bayesian VARs. *Journal of Applied Econometrics*, *28*, 177–203.

Mao Takongmo, C., & Stevanovic, D. (2015). Selection of the number of factors in presence of structural instability: A Monte Carlo study. *Actualité Economique*, *91*, 177–233.

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

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *The Journal of Economic Perspectives*, *31*(2), 574–589.

Ng, S., & Perron, P. (1996). Useful modifications to some unit root tests in with dependent errors and their local asymptotic properties. *The Review of Economic Studies*, *63*, 435–463.

Ng, S., & Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power. *Econometrica*, *69*(6), 1519–1554.

Pesaran, H., & Timmermann, A. (1992). A simple nonparametric test of predictive performance. *Journal of Business & Economic Statistics*, *10*(4), 461–465.

Pettenuzzo, D., & Timmermann, A. (2017). Forecasting macroeconomic variables under model instability. *Journal of Business & Economic Statistics*, *35*(2), 183–201.

Rossi, B., & Sekhposyan, T. (2010). Have economic models' forecasting performance for US output growth and inflation changed over time, and when? *International Journal of Forecasting*, *26*(4), 808–835.

Rossi, B., & Sekhposyan, T. (2011). Understanding models' forecasting performance. *Journal of Econometrics*, *164*(1), 158–172.

Satchell, S., & Timmermann, A. (1995). An assessment of the economic value of non-linear foreign exchange rate forecasts. *Journal of Forecasting*, *14*(6), 477–497.

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

Stock, J. H., & Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, *20*(2), 147–162.

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

Stock, J. H., & Watson, M. W. (2006). Macroeconomic forecasting using many predictors. In: G. Elliott, C. W. J. Granger, & A. Timmermann (Eds.). *Handbook of economic forecasting 1*, (pp. 516–550); North-Holland: Amsterdam, Netherlands.

Stock, J. H., & Watson, M. W. (2007). Why has U.S. inflation become harder to forecast? *Journal of Money, Credit, and Banking*, *39*(s1), 3–33.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Kotchoni R, Leroux M, Stevanovic D. Macroeconomic forecast accuracy in a data-rich environment. *J Appl Econ*. 2019;34:1050–1072. https://doi.org/10.1002/jae.2725

APPENDIX A: OTHER FORECAST EVALUATION METRICS

A.1 | Ratio of correctly signed forecasts

Here, we compare the forecasting methods in terms of their ability to generate forecasts that are correctly signed. Indeed, a forecasting model that is outperformed in terms of the MSPE can still have significant predictive power for the sign of the target variable (see Satchell and Timmermann (1995). This possibility can be assessed by means of the Pesaran and Timmermann (1992) sign forecast test. The test statistic is given by

$$S_n = \frac{\hat{p} - \hat{p}^*}{\sqrt{\mathrm{var}(\hat{p}) - \mathrm{var}(\hat{p}^*)}},$$

where $\hat{p}$ is the sample ratio of correctly signed forecasts (RCSF) and $\hat{p}^*$ is the estimate of its expectation. This test statistic is not influenced by the distance between the realization and the forecast, as is the case for MSPE. Under the null hypothesis that the signs of the forecasts are independent of the signs of the target, we have $S_n \rightarrow \mathrm{N}(0, 1)$.[10] Tables A1–A4 present the success ratio with the test significance. The highest values are in bold. Implicitly, the benchmark model here is the random walk without drift.

[10]Let q denote the proportion of positive realizations in the actual data and $\hat{q}$ the proportion of positive forecasts. Under H_0, the estimated theoretical number of correctly signed forecasts is $\hat{p}^* = q\hat{q} + (1-q)(1-\hat{q})$.

TABLE A1 RCSF for industrial production growth

Model	*Full out-of-sample* $h=1$	$h=3$	$h=6$	$h=9$	$h=12$	NBER recessions periods $h=1$	$h=3$	$h=6$	$h=9$	$h=12$
Standard time series models										
ARD	0.68***	0.77***	0.75***	0.75	0.74	0.54	0.49**	0.29	0.22	0.24
ARI	0.68***	0.76***	0.76***	0.76**	0.76	0.54	0.41*	0.26	0.26	0.25
ARMA(1, 1)	0.70***	0.78***	0.76***	0.74	0.74	0.65**	0.48**	0.27	0.22	0.25
ADL	0.70***	0.77***	0.75***	0.74	0.76	0.61*	0.47**	0.31*	0.25	0.26
Factor-augmented regressions										
ARDI	0.69***	0.81***	0.81***	0.82***	0.82***	0.71	0.76**	0.55	0.51	0.42
ARDI-soft	0.70***	0.81***	0.81***	**0.84*****	0.82***	0.75	0.78	0.66	0.64	0.56*
ARDI-hard,1.28	0.72***	0.82***	0.79***	0.81***	0.84***	0.72	0.72	0.58	0.51	0.64***
ARDI-hard,1.65	0.71***	0.81***	0.79***	0.82***	0.84***	0.71	0.78	0.61	0.56	**0.69*****
ARDI-tstat,1.96	0.69***	0.80***	0.80***	0.81***	0.83***	0.62	0.62	0.53	0.51	0.51**
ARDI-DU	0.72***	0.81***	0.80***	0.82***	0.83***	0.73	0.72	0.53	0.47	0.48
3PRF	0.71***	0.79***	0.77***	0.76***	0.75**	0.67	0.55	0.36	0.31	0.32
Factor-structure-based models										
FAVARI	0.74***	0.82***	0.80***	0.82***	0.82***	0.73	**0.79***	0.59	0.44	0.42
FAVARD	0.74***	**0.83*****	0.81***	0.82***	0.83***	0.78	0.79*	0.61	0.52	0.52
FAVARMA-FMA	0.73***	0.81***	0.82***	0.82***	0.83***	0.71	0.78	0.66	0.52	0.52*
FAVARMA-FAR	0.69***	0.79***	0.80***	0.79***	0.76***	0.78	0.76	0.64	0.55***	0.44**
DFM	0.71***	0.80***	0.78***	0.79***	0.80***	0.72	0.67	0.41	0.35	0.35
Data-rich model averaging										
CSR,1	0.68***	0.77***	0.76**	0.76	0.75	0.52*	0.41*	0.24	0.21	0.24
CSR,10	0.73***	0.82***	0.81***	0.81***	0.82***	0.72	0.65	0.49	0.4	0.41*
CSR,20	**0.75*****	0.83***	0.80***	0.83***	0.83***	0.76	0.69	0.58	0.55	0.51*
Regularized data-rich model averaging										
T-CSR-soft,10	0.73***	0.83***	**0.82*****	0.83***	0.83***	0.78	0.71	0.64	0.55	0.59***
T-CSR-soft,20	0.72***	0.81***	0.80***	0.80***	0.82***	0.79	0.71	**0.68**	0.62	0.64**
T-CSR-hard,1.65,10	0.74***	0.82***	0.81***	0.83***	**0.85*****	0.76	0.69	0.58	0.58	0.60***
T-CSR-hard,1.65,20	0.73***	0.82***	0.80***	0.80***	0.82***	0.74	0.75*	0.66	0.56	0.60*
R-CSR.10	0.72***	0.82***	0.81***	0.82***	0.82***	0.72	0.69	0.51	0.45	0.44**
R-CSR,20	0.74***	0.83***	0.81***	0.82***	0.84***	**0.81**	0.71	0.59	0.55	0.54***
Lasso	0.69***	0.76***	0.76***	0.79***	0.80***	0.76**	0.66	0.65	**0.65**	0.61**
Forecast combinations										
AVRG	0.73***	0.83***	0.81***	0.82***	0.83***	0.75	0.73	0.55	0.51	0.46**
Median	0.73***	0.82***	0.81***	0.83***	0.84***	0.73	0.73	0.53	0.51	0.48**
T-AVRG	0.73***	0.82***	0.81***	0.82***	0.84***	0.73	0.73**	0.53	0.48	0.47**
IP-AVRG,1	0.73***	0.83***	0.81***	0.83***	0.83***	0.75	0.73	0.55	0.51	0.48**
IP-AVRG,0.95	0.73***	0.82***	0.81***	0.83***	0.83***	0.74	0.72	0.55	0.48	0.45*

Note. This table shows the success ratio with the Pesaran and Timmermann (1992) sign forecast test significance. Asterisks indicate ***1%, **5%, and *10% significance levels. The highest values are in bold.

TABLE A2 RCSF for employment growth

	Full out-of-sample					NBER recessions periods				
Model	$h = 1$	$h = 3$	$h = 6$	$h = 9$	$h = 12$	$h = 1$	$h = 3$	$h = 6$	$h = 9$	$h = 12$
Standard time series models										
ARD	0.89***	0.90***	0.84***	0.82***	0.80***	0.76***	0.71***	0.49***	0.51**	0.52*
ARI	0.89***	0.90***	0.85***	0.81***	0.81***	0.76***	0.68***	0.51***	0.51**	0.52*
ARMA(1, 1)	0.89***	0.88***	0.83***	0.81***	0.79***	0.79***	0.64***	0.52***	0.51**	0.52*
ADL	0.89***	0.89***	0.84***	0.81***	0.81***	**0.84*****	0.69***	0.51***	0.52***	0.53**
Factor-augmented regressions										
ARDI	0.89***	0.91***	0.85***	0.83***	0.83***	0.82***	0.76***	0.48	0.48	0.60***
ARDI-soft	0.87***	0.90***	0.88***	0.85***	0.83***	0.81***	0.74***	0.59***	0.56**	**0.65*****
ARDI-hard,1.28	0.87***	0.90***	**0.88*****	0.86***	**0.84*****	0.79***	0.73***	**0.64*****	**0.58****	0.65***
ARDI-hard,1.65	0.87***	0.90***	0.88***	**0.87*****	0.84***	0.80***	0.75***	0.59***	0.55**	0.64***
ARDI-tstat,1.96	0.87***	0.91***	0.87***	0.84***	0.83***	0.76***	0.76***	0.61***	0.54*	0.61***
ARDI-DU	0.89***	0.90***	0.87***	0.83***	0.84***	0.82***	0.72***	0.58***	0.53	0.65***
3PRF	0.86***	0.88***	0.84***	0.82***	0.81***	0.74***	0.67***	0.49***	0.49*	0.52
Factor-structure-based models										
FAVARI	0.88***	0.90***	0.85***	0.83***	0.81***	0.80***	0.68***	0.52***	0.48	0.51*
FAVARD	0.89***	0.91***	0.86***	0.84***	0.83***	0.81***	0.75***	0.58***	0.54**	0.58**
FAVARMA-FMA	0.89***	0.90***	0.85***	0.83***	0.82***	0.81***	0.72***	0.55***	0.49*	0.55***
FAVARMA-FAR	**0.89*****	0.90***	0.86***	0.82***	0.81***	0.81***	0.66***	0.49***	0.49**	0.51*
DFM	0.88***	0.90***	0.85***	0.82***	0.81***	0.76***	0.67***	0.47**	0.48	0.53**
Data-rich model averaging										
CSR,1	0.87***	0.89***	0.84***	0.81***	0.79***	0.69***	0.62***	0.49***	0.51**	0.52*
CSR,10	0.87***	0.89***	0.85***	0.80***	0.81***	0.74***	0.62***	0.48***	0.47	0.54**
CSR,20	0.87***	0.89***	0.86***	0.82***	0.82***	0.76***	0.64***	0.55***	0.54**	0.59***
Regularized data-rich model averaging										
T-CSR-soft,10	0.87***	0.90***	0.86***	0.84***	0.82***	0.78***	0.68***	0.54***	0.55***	0.56***
T-CSR-soft,20	0.87***	0.91***	0.87***	0.84***	0.83***	0.76***	0.74***	0.64***	0.54	0.59**
T-CSR-hard,1.65,10	0.88***	0.90***	0.87***	0.84***	0.82***	0.79***	0.68***	0.60***	0.56***	0.59***
T-CSR-hard,1.65,20	0.86***	0.90***	0.86***	0.83***	0.81***	0.75***	0.71***	0.56***	0.54*	0.56**
R-CSR,10	0.89***	0.91***	0.87***	0.83***	0.83***	0.80***	0.75***	0.53***	0.49*	0.56***
R-CSR,20	0.89***	**0.92*****	0.87***	0.84***	0.84***	0.81***	**0.78*****	0.59***	0.55***	0.61***
Lasso	0.85***	0.90***	0.87***	0.82***	0.82***	0.72**	0.72***	0.60***	0.54	0.59**
Forecast combinations										
AVRG	0.88***	0.91***	0.85***	0.82***	0.82***	0.80***	0.71***	0.49***	0.49*	0.55***
Median	0.88***	0.91***	0.85***	0.82***	0.82***	0.80***	0.72***	0.51***	0.53**	0.56***
T-AVRG	0.88***	0.91***	0.85***	0.82***	0.82***	0.80***	0.72***	0.49***	0.51*	0.55***
IP-AVRG,1	0.88***	0.91***	0.85***	0.83***	0.83***	0.80***	0.72***	0.51***	0.54**	0.59***
IP-AVRG,0.95	0.88***	0.91***	0.86***	0.83***	0.82***	0.80***	0.71***	0.51***	0.53**	0.56***

Note. This table shows the success ratio with the Pesaran and Timmermann (1992) sign forecast test significance. Asterisks indicate ***1%, **5%, and *10% significance levels. The highest values are in bold.

TABLE A3 RCSF for the CPI inflation growth

Model	Full out-of-sample					NBER recessions periods				
	$h = 1$	$h = 3$	$h = 6$	$h = 9$	$h = 12$	$h = 1$	$h = 3$	$h = 6$	$h = 9$	$h = 12$
Standard time series models										
ARD	0.62***	0.71***	0.74***	0.75***	0.72***	0.60**	0.69***	0.76***	0.69***	0.64**
ARI	0.62***	0.65***	0.67***	0.69***	0.68***	0.60**	0.64**	0.75***	0.73***	0.60*
ARMA(1,1)	**0.67*****	0.71***	0.71***	0.71***	0.72***	0.60**	0.72***	0.75***	0.73***	0.65**
ADL	0.63***	0.69***	0.74***	0.75***	0.73***	0.60*	0.66***	0.76***	0.73***	0.66***
Factor-augmented regressions										
ARDI	0.63***	0.70***	0.74***	0.74***	0.74***	0.64**	0.67***	0.74***	0.71***	0.66***
ARDI-soft	0.62***	0.70***	0.73***	0.74***	0.72***	0.64**	0.68***	0.73***	0.72***	0.69***
ARDI-hard,1.28	0.60***	0.69***	0.73***	0.75***	0.74***	0.62**	0.67***	0.75***	0.75***	0.69***
ARDI-hard,1.65	0.63***	0.71***	0.73***	0.75***	0.74***	0.65***	0.71***	0.73***	0.76***	**0.73*****
ARDI-tstat,1.96	0.62***	0.71***	0.73***	0.73***	0.75***	0.65***	0.72***	0.79***	0.72***	0.73***
ARDI-DU	0.63***	**0.73*****	0.73***	0.75***	**0.75*****	0.62**	0.71***	0.73***	0.71***	0.67***
3PRF	0.57***	0.63***	0.66***	0.66***	0.66***	0.55	0.68***	0.69***	0.74***	0.64**
Factor-structure-based models										
FAVARI	0.56**	0.60***	0.56***	0.54*	0.54	0.67***	0.59	0.53	0.53	0.58
FAVARD	0.55*	0.62***	0.62***	0.60***	0.57***	0.66**	0.59	0.56	0.61**	0.64**
FAVARMA-FMA	0.57***	0.62***	0.59***	0.57***	0.57***	0.66***	0.62**	0.61**	0.59	0.61**
FAVARMA-FAR	0.5	0.40***	0.38***	0.35***	0.36***	0.58	0.38***	0.40*	0.42	0.48
DFM	0.63***	0.71***	0.71***	0.71***	0.68***	0.64***	**0.74*****	0.78***	0.67***	0.61**
Data-rich model averaging										
CSR,1	0.58***	0.63***	0.64***	0.65***	0.64***	0.59**	0.64***	0.64**	0.68***	0.58
CSR,10	0.63***	0.65***	0.68***	0.68***	0.67***	0.67***	0.69***	0.69***	0.72***	0.64**
CSR,20	0.62***	0.65***	0.68***	0.69***	0.68***	0.67***	0.68***	0.71***	0.71***	0.64**
Regularized data-rich model averaging										
T-CSR-soft,10	0.64***	0.67***	0.72***	0.69***	0.70***	0.66***	0.69***	0.74***	0.74***	0.68***
T-CSR-soft,20	0.64***	0.65***	0.68***	0.69***	0.68***	0.69***	0.73***	0.73***	0.73***	0.66***
T-CSR-hard,1.65,10	0.63***	0.67***	0.67***	0.71***	0.71***	0.66***	0.72***	0.69***	0.72***	0.68***
T-CSR-hard,1.65,20	0.63***	0.65***	0.68***	0.71***	0.70***	0.68***	0.66***	0.74***	**0.78*****	0.71***
R-CSR,10	0.64***	0.71***	0.74***	0.75***	0.75***	0.65**	0.74***	0.78***	0.74***	0.69***
R-CSR,20	0.66***	0.69***	0.73***	**0.76*****	0.74***	0.67***	0.69***	0.75***	0.74***	0.69***
Lasso	0.65***	0.64***	0.72***	0.70***	0.69***	0.66***	0.72***	0.75***	0.75***	0.67***
Forecast combinations										
AVRG	0.64***	0.71***	0.74***	0.74***	0.74***	0.69***	0.71***	**0.80*****	0.76***	0.71***
Median	0.66***	0.70***	0.72***	0.75***	0.75***	**0.72*****	0.72***	0.74***	0.75***	0.71***
T-AVRG	0.65***	0.71***	0.72***	0.75***	0.73***	0.71***	0.69***	0.75***	0.76***	0.69***
IP-AVRG,1	0.65***	0.71***	0.74***	0.75***	0.75***	0.69***	0.69***	0.79***	0.74***	0.71***
IP-AVRG,0.95	0.64***	0.71***	**0.74*****	0.74***	0.74***	0.69***	0.69***	0.80***	0.72***	0.68***

Note. This table shows the success ratio with the Pesaran and Timmermann (1992) sign forecast test significance. Asterisks indicate ***1%, **5%, and *10% significance levels. The highest values are in bold.

TABLE A4 RCSF for the S&P 500 returns

	Full out-of-sample					NBER recessions periods				
Model	$h = 1$	$h = 3$	$h = 6$	$h = 9$	$h = 12$	$h = 1$	$h = 3$	$h = 6$	$h = 9$	$h = 12$
Standard time series models										
ARD	0.59*	0.63	0.64	0.64**	0.64***	0.54	0.39	0.24**	0.15	0.14**
ARI	0.59*	0.64**	0.64	0.64**	0.63***	0.54	0.41	0.21***	0.14**	0.13***
ARMA(1,1)	0.60**	0.63	0.63	0.63**	0.63***	0.52	0.41	0.21**	0.11***	0.13***
ADL	0.59**	0.62***	0.66***	0.63	0.64	**0.69*****	0.46	0.41	0.26**	0.31
Factor-augmented regressions										
ARDI	0.62***	0.64***	0.67***	0.67***	0.69***	0.65***	0.55	0.46	0.42	0.4
ARDI-soft	0.61***	0.63***	0.67***	0.68***	0.66**	0.62**	0.52	0.45	0.38	0.33
ARDI-hard,1.28	0.60***	0.66***	0.67***	0.67***	0.68***	0.65***	0.58	0.48	0.4	0.36
ARDI-hard,1.65	0.61***	0.64***	0.65***	0.65***	0.67**	0.62**	0.55	0.46	0.38	0.35
ARDI-tstat,1.96	0.61***	0.65***	0.66***	0.67***	0.67**	0.64**	**0.62****	0.44	0.4	0.39
ARDI-DU	0.60***	0.65***	0.67***	0.67***	0.70***	0.61**	0.54	0.46	0.4	0.44
3PRF	0.61***	0.66***	0.69***	0.68***	0.65	0.59*	0.56	0.48	0.45	0.36
Factor-structure-based models										
FAVARI	0.63***	0.66***	**0.69*****	0.71***	**0.70*****	0.64**	0.59	0.54	0.52	0.47
FAVARD	0.62***	0.64***	0.69***	**0.71*****	0.70***	0.61**	0.58	0.54	**0.56**	0.47
FAVARMA-FMA	0.63***	0.66***	0.69***	0.70***	0.68***	0.66***	0.6	**0.55**	0.51	0.42
FAVARMA-FAR	0.60***	0.62***	0.67***	0.67***	0.67***	0.60**	0.47	0.52	0.51	0.48
DFM	0.60**	0.64***	0.68***	0.68***	0.67	0.61**	0.51	0.41	0.33**	0.32
Data-rich model averaging										
CSR,1	0.61**	0.63*	0.65	0.65	0.65***	0.56	0.41	0.29	0.16**	0.16**
CSR,10	0.62***	0.66***	0.68***	0.67***	0.68**	0.67***	0.55	0.42	0.38	0.35
CSR,20	0.61***	0.64***	0.68***	0.66***	0.65	0.62**	0.52	0.47	0.36	0.35
Regularized data-rich model averaging										
T-CSR-soft,10	0.59***	0.64***	0.67***	0.69***	0.66**	0.61**	0.56	0.44	0.4	0.35
T-CSR-soft,20	0.57**	0.62***	0.65***	0.64***	0.65**	0.64**	0.61*	0.52*	0.44	0.46
T-CSR-hard,1.65,10	**0.63*****	0.65***	0.66***	0.67***	0.65	0.62**	0.54	0.46	0.41	0.33
T-CSR-hard,1.65,20	0.63***	0.62***	0.62***	0.63**	0.61	0.64***	0.53	0.42	0.41	0.33
R-CSR,10	0.63***	0.66***	0.68***	0.69***	0.69***	0.67***	0.56	0.46	0.4	0.38
R-CSR,20	0.62***	0.64***	0.67***	0.68***	0.66*	0.65***	0.55	0.47	0.4	0.35
Lasso	0.54	0.61***	0.61***	0.60**	0.6	0.54	0.6	0.55	0.52	**0.49**
Forecast combinations										
AVRG	0.63***	0.67***	0.68***	0.70***	0.67	0.67***	0.55	0.45	0.4	0.35
Median	0.62***	0.66***	0.67***	0.68***	0.68*	0.66***	0.55	0.46	0.38	0.36
T-AVRG	0.62***	0.66***	0.68***	0.67***	0.67	0.68***	0.56	0.45	0.38	0.35
IP-AVRG,1	0.63***	**0.67*****	0.68***	0.69***	0.67*	0.69***	0.55	0.45	0.38	0.35
IP-AVRG,0.95	0.63***	0.67***	0.68***	0.68***	0.68*	0.69***	0.54	0.45	0.36	0.35
Random walks										
RW with drift	0.58	0.60**	0.63	0.63***	0.62***	0.48	0.29**	0.19**	0.07***	0.08***

Note. This table shows the success ratio with the Pesaran and Timmermann (1992) sign forecast test significance. Asterisks indicate ***1%, **5%, and *10% significance levels. The highest values are in bold.