Economix

http://economix.fr

Forecasting economic activity in data-rich environment

Document de Travail
Working Paper
2017-05

Maxime Leroux
Rachidi Kotchoni
Dalibor Stevanovic

UMR 7235

Université de Paris Ouest Nanterre La Défense
(bâtiment G)
200, Avenue de la République
92001 NANTERRE CEDEX

Tél et Fax : 33.(0)1.40.97.59.07
Email : nasam.zaroualete@u-paris10.fr

université Paris Ouest Nanterre La Défense

Forecasting economic activity in data-rich environment*

Maxime Leroux[†] Rachidi Kotchoni[‡] Dalibor Stevanovic [§]

This version: January 24, 2017

Abstract

This paper compares the performance of five classes of forecasting models in an extensive out-of-sample exercise. The types of models considered are standard univariate models, factor-augmented regressions, dynamic factor models, other data-rich models and forecast combinations. These models are compared using four types of data: real series, nominal series, the stock market index and exchange rates. Our findings can be summarized in a few points: *(i)* data-rich models and forecasts combination approaches are the best for predicting real series; *(ii)* ARMA(1,1) model predicts inflation change incredibly well and outperform data-rich models; *(iii)* the simple average of forecasts is the best approach to predict future SP500 returns; *(iv)* exchange rates can be predicted at short horizons mainly by univariate models but the random walk dominates at medium and long terms; *(v)* the optimal structure of forecasting equations changes much over time; and *(vi)* the dispersion of out-of-sample point forecasts is a good predictor of some macroeconomic and financial uncertainty measures as well as of the business cycle movements among real activity series.

JEL Classification: C55, C32, E17

Keywords: Forecasting, Factor Models, Data-rich environment, Model averaging.

*The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Department of Finance Canada. We thank Mehmet Caner, Todd Clark, Marco Del Negro, John Galbraith and Serena Ng for valuable comments.

[†]Department of Finance Canada, Ottawa, Email: maxime.leroux@canada.ca

[‡]Economix-CNRS, Université Paris Nanterre. Email: rachidi.kotchoni@u-paris10.fr

[§]Département des sciences économiques, Université du Québec à Montréal. 315, Ste-Catherine Est, Montréal, QC, H2X 3X2. Email: dstevanovic.econ@gmail.com
The author acknowledges financial support from the Fonds de recherche sur la société et la culture (Québec) and the Social Sciences and Humanities Research Council. Corresponding author.

1 Introduction

During recent decades, researchers and practitioners in economics and related fields have witnessed a fast improvement of data collection and storage capacity. Many economic and financial data sets have now reached astronomical sizes, both in terms of the dimension (number of variables) and the sample size (e.g. high-frequency data in finance). For instance, the Federal Reserve of St-Louis Economic Database (FRED) contains more than 390,000 macroeconomic and financial time series. As all these series will not be relevant for a given forecasting exercise, one will have to preselect relevant candidate predictors according to economic theories, the empirical literature and own heuristic arguments. In a data-rich environment, however, the econometrician can easily 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. Moreover, the numerical algorithms routinely used to train standard econometric models quickly break down as the number of variables increases. This is the well-known curse of dimensionality.

The new challenge faced by applied econometricians is to design computationally efficient methods capable of turning big datasets into concise information. Bayesian techniques developed in recent years to handle larger than usual VAR models can be viewed as an effort toward this objective. See (Banbura, Giannone & Reichlin 2010), (Koop 2013), (Carriero, Clark & Marcellino 2015) and (Giannone, Lenza & Primiceri 2015), among others. Another promising route to deal with big data is provided by factor models, where a potentially large number of variables is assumed to be driven by a smaller number of latent "*factors*". The latter class of models is largely advocated in the current paper, as done by precursors like (Stock & Watson 2002b) and (Ludvigson & Ng 2005), among others.

Since the seminal work of (Stock & Watson 2002), many forecasting methods utilizing factor-based models have been proposed to deal with data-rich environment. Given their growing popularity in the literature, there is a need for an extensive study that compares the performance of these methods. Few studies have done such a comparison exercise. See (Boivin & Ng 2005), (Kim & Swanson 2014), (Cheng & Hansen 2015), (Carrasco & Rossi 2016) and (Groen & Kapetanios 2016). This paper contributes to filling this gap by comparing the performance of five classes of models at forecasting four types of variables.

The first class of forecasting models considered consists of standard and univariate specifications, namely the Autoregressive Direct model (ARD), the Autoregressive Iterative model (ARI), the simplest Autoregressive Moving Average ARMA(1,1) model and the Autoregressive Distributed Lag (ADL) model. The second class of models consists of autoregressions that are augmented with exogenous factors: the Diffusion Indices (DI) of (Stock & Watson 2002b), the Targeted DI of (Bai & Ng 2008) and the DI with dynamic factors of (Forni, Hallin, Lippi

& Reichlin 2005). The third type of models assume that the factors are endogenous, meaning that the dynamics of the series being predicted obey the assumed factor structure. In the latter category, we have the Factor-Augmented VAR (FAVAR) of (Boivin & Ng 2005), the Factor-Augmented VARMA (FAVARMA) of (Dufour & Stevanovic 2013) and the Dynamic Factor Model (DFM) of (Forni et al. 2005). The fourth category consists of two recent and prominent data-rich models: the Three-pass Regression Filter (3PRF) of (Kelly & Pruitt 2015) and the Complete Subset Regression (CSR) of (Elliott, Gargano & Timmermann 2013). Finally, the fifth category consists of methods that average the previous forecasts. Here we consider the naive average of all forecasts (AVRG), the median of all forecasts (MED), the trimmed average of all forecasts (T-AVRG) and the inversely proportional average of all forecasts (IP-AVRG). The latter forecasting method is considered in (Stock & Watson 2004). For the sake of completeness, the simple random walk (RW) and the random walk with drift (RWD) are considered as well.

The data employed for this study come from an updated version of Stock and Watson's macroeconomic database that was used in (McCracken & Ng 2015). We compare the forecasting methods listed above on *(i)* real series (Industrial Production and Employment), *(ii)* nominal series (Consumer Price Index and Core CPI), *(iii)* financial series (SP500) and *(iv)* exchange rates (US-UK and US-Canada). These variables are selected for their popularity in the forecasting literature. The comparison approach is based on a pseudo out-of-sample scheme that uses the Mean Square Prediction Error (MSPE) and the Mean Absolute Prediction Error (MAPE) as metrics. More precisely, we compute the Relative MSPE and MAPE of all models with respect to the ARD model used as a benchmark. We also consider the sign prediction. For each series, horizon and out-of-sample period, the hyperparameters of our models (number of lags, number of factors, etc.) are calibrated using the Bayesian Information Criterion (BIC).

To the best of our knowledge, our paper is the first to put so many different forecasting models together and compare their performance on several types of data. Our first contribution is to compare these models in an extensive out-of-sample exercise and document their performance and robustness across the business cycles. Disentangling which type of models have significant forecasting power for real activity, prices, stock market and exchange rates is a valuable information for practitioners and policy makers. The second contribution of the current work is to provide a laboratory for future development of forecasting models. [1]

We find that data-rich models and forecast combination approaches perform well at predicting real series. The two dominating techniques to forecast Industrial Production growth and Employment growth are the IP-AVRG and the CSR. The worst performing here are the RW models, which suggests that real series are highly predictable. In addition, we identify the

[1]All the data used in the paper are public. Our Matlab codes will be made publicly available so that they can be updated to include any new forecasting model.

best models to be used during recession periods. In a real-life application, the most pessimistic forecast can be used as a worse case scenario that becomes more and more realistic on the eve of economic crises.

The ARMA(1,1) model emerges as the best to forecast inflation change. This result holds regardless of whether the CPI or the Core CPI is considered. Curiously enough, data-rich models do worse than the ARD benchmark at forecasting inflation growth. One possible explanation for this good performance of univariate models is that inflation change is largely exogenous with respect to the conditioning information set available to us. As a result, data-rich models are over-parameterized for this series and therefore have poor generalization performance.

The best forecasts of the SP500 returns at short horizons are obtained by averaging the other forecasts (AVRG and T-AVRG). At longer horizons, the CSR dominates. While the relative performance of the RW and RWD models with respect to the ARD benchmark improves much, these models underperform CSR, AVRG and T-AVRG at all horizons. This suggests that stock returns are predictable to some extent. Finally, we find that data-rich models are of no use when it comes to predicting exchange rates as they all underperform the ARD benchmark. At short horizon, univariate models deliver the best forecasts while at longer horizons, the RW models dominate. This suggests that of all series considered, exchange rates are the most difficult to predict. However, data-rich models improve the prediction of the direction of change.

Given the tremendous amount of information produced in this extensive out-of-sample horse race, we are able to study the stability of forecasting relationships as well as the out-of-sample forecasts dispersion. We find evidence of widespread structural changes in all dimensions of forecasting equations. However, they are not evenly distributed across the forecasted series and forecasting horizons. We find that the dispersion of out-of-sample point forecasts is significantly correlated with some macroeconomic and financial uncertainty measures used in the literature.

The remainder of the paper is organized as follows. Section 2 presents the standard time series models. Section 3 presents the data-rich environment and the corresponding models. Forecasts combinations are shown in Section 4. Section 5 presents the data, the design of the pseudo out-of-sample exercise and the results. Section 6 presents the stability of some forecasting models and the links between the forecasts dispersion and uncertainty. Section 7 concludes. Additional empirical results and simulation analysis are reported in a separate supplementary material.

2 Standard Forecasting Models

Let Y_t denote a macroeconomic or financial 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:

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

where $y_t \equiv \ln Y_t$ and $freq$ depends on the frequency of the data (400 if Y_t is quarterly, 1200 if Y_t is monthly, etc.).

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 & Watson 2002b) and (McCracken & Ng 2015). We shall therefore define $y_{t+h}^{(h)}$ as:

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

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

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

$$y_{t+h}^{(h)} = (freq/h) \sum_{k=1}^{h} y_{t+k} = (freq/h) \left[\ln(Y_{t+h}/Y_{t+h-1}) - \ln(Y_t/Y_{t-1})\right], \tag{3}$$

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

Indeed, $y_{t+h}^{(h)}$ is given by the same function of y_t everywhere while y_t is $\ln Y_t$ in (1), the first difference of $\ln Y_t$ in (2) and the second difference of $\ln Y_t$ in (3). In the remainder of the section, we describe the standard univariate and multivariate forecasting models advocated in the paper.

Autoregressive Direct (ARD) Our first univariate model is the so-called *autoregressive direct* (ARD) model, which is specified as:

$$y_{t+h}^{(h)} = \alpha^{(h)} + \sum_{l=1}^{L} \rho_l^{(h)} y_{t-l+1} + e_{t+h}, \quad t = 1, \ldots, T, \tag{4}$$

where $h \geq 1$ and $L \geq 1$. A direct prediction of y_{T+h}^{h} is deduced from the model above as follows:

$$\hat{y}_{T+h|T}^{h} = \hat{\alpha}^{(h)} + \sum_{l=1}^{L} \hat{\rho}_l^{(h)} y_{T-l+1},$$

where $\hat{\alpha}^{(h)}$ and $\hat{\rho}^{(h)}$ are OLS estimators of $\alpha^{(h)}$ and $\rho^{(h)}$. The optimal L will be selected using the Bayesian Information Criterion (BIC) for every out-of-sample (OOS) period. This makes the forecasting model more flexible by allowing the optimal L to vary over the OOS period.

Autoregressive Iterative (ARI) Our second univariate model is a standard AR(L) model specified as:

$$y_{t+1} = \alpha + \sum_{l=1}^{L} \rho_l y_{t+1-l} + e_{t+1}, \quad t = 1, \ldots, T. \tag{5}$$

where $L \geq 1$. This model is termed *autoregressive iterative* (ARI) because $\widehat{y}^h_{T+h|T}$ must be deduced from recursive calculations of $\widehat{y}_{T+1|T}$, $\widehat{y}_{T+2|T}$,...,$\widehat{y}_{T+h|T}$. We have:

$$\hat{y}_{T+k|T} = \hat{\alpha} + \sum_{l=1}^{L} \hat{\rho}_l \widehat{y}_{T+k-l|T}, \qquad k = 1, ..., h,$$

with the convention $\widehat{y}_{t|T} \equiv y_t$ for all $t \leq T$ and:

$$\widehat{y}^h_{T+h|T} = (freq/h) \sum_{k=1}^{h} \hat{y}_{T+k|T}. \tag{6}$$

Equation (6) will remain the appropriate prediction formula for all iterative models as long as the definition of y_t is adapted to whether $\ln Y_t$ is I(0), I(1) or I(2).

Here too, the optimal lag L will be selected using the Bayesian Information Criterion (BIC) for every out-of-sample period. If the true DGP of y_t is an AR(L), both the direct and iterative approaches should produce the same predictions for any horizon asymptotically as the sample size goes to infinity. However, none of the two specifications strictly dominates in finite samples. The iterative approach is found to be better when a true AR(L) process prevails for y_t while the direct approach is more robust to misspecification, see (Chevillon 2007). (Marcellino, Stock & Watson 2006) compare the forecasting performance of direct and iterative models for hundreds of time series. They conclude that the direct approach provides slightly better results but does not dominate uniformly across time and series.

ARMA(1,1) (Dufour & Stevanovic 2013) showed that ARMA models arise naturally as the marginal univariate representation of observables when they jointly follow a dynamic factor model. This suggests that the ARMA(1,1) is a natural benchmark against which to evaluate the performance of data-rich models.[2] The following representation is therefore considered and

[2]The ARMA(1,1) model has been used extensively in the empirical finance literature to forecast the realized volatility, but has been considered much less for the prediction of macroeconomic series.

estimated by maximum likelihood:

$$y_{t+1} = \alpha + \rho y_t + \theta e_t + e_{t+1}. \tag{7}$$

After estimation, the residuals $\hat{e}_T$ of the ARMA(1,1) model are generated in-sample using the recursion starting from the initial value $\widehat{e}_1 = 0$:

$$\widehat{e}_{t+1} = y_{t+1} - \widehat{\alpha} - \widehat{\rho} y_t - \widehat{\theta}\widehat{e}_t, \quad t = 1, ..., T.$$

The prediction of y_{T+h} for any horizon h is computed using the formula (6) along with the output of the following recursion:

$$\hat{y}_{T+k|T} = \hat{\alpha} + \hat{\rho}\hat{y}_{T+k-1|T} + \hat{\theta}\hat{e}_{T+k-1|T}, \quad k = 1, ..., h,$$

where $\hat{y}_{T|T} = y_T$, $\hat{e}_{T|T} = \hat{e}_T$ and $\hat{e}_{T+k|T} = 0$ for all $k = 1, ..., h$.

An alternative approach to forecast the ARMA(1,1) model is given by the finite order approximation of its AR(∞) representation. A truncation to L lags leads to the following forecasting formula:

$$\hat{y}_{T+k|T} = \frac{\hat{\alpha}}{1+\hat{\theta}} + \sum_{l=1}^{L} \left(\hat{\rho} + \hat{\theta}\right) \left(-\hat{\theta}\right)^{l-1} \hat{y}_{T+k-l|T}, \quad k = 1, ..., h,$$

with $\widehat{y}_{t|T} \equiv y_t$ for all $t \leq T$. This approach assumes that the ARMA(1,1) is invertible. The optimal truncation lag L could be chosen using the Bayesian Information Criterion (BIC).

Autoregressive Distributed Lag (ADL) A simple extension of the ARD model is obtained by adding exogenous predictors Z_t to its right-hand side. This leads to the so-called ADL model given by:

$$y_{t+h}^{(h)} = \alpha^{(h)} + \sum_{l=1}^{L} \rho_l^{(h)} y_{t-l+1} + \sum_{k=1}^{K} Z_{t-k+1}\beta_k^{(h)} + e_{t+h}, \tag{8}$$

where Z_t contains a small number of selected series. The precise content of Z_t is discussed in the empirical section.

3 Data-Rich Models

There is a growing literature on how to deal with a large number of predictors when forecasting macroeconomic time series. The factor-based approaches started with the diffusion indices model of (Stock & Watson 2002) and (Stock & Watson 2002b). Since then, several modifications and extensions of this model have been proposed. Among others, it has been combined with Lasso-type shrinkage in (Bai & Ng 2008), while (Elliott et al. 2013) proposed a forecast combination approach.

Let X_t be an N-dimensional stationary stochastic process. We consider a general DFM representation of X_t that will serve as a basis for subsequent analyses. Following the notation of (Dufour & Stevanovic 2013) and (Stock & Watson 2005), we assume that:

$$X_t = \lambda(L) f_t + u_t \,, \tag{9}$$

$$u_t = \delta(L) u_{t-1} + \nu_t \,, \tag{10}$$

$$f_t = \gamma(L) f_{t-1} + \theta(L)\eta_t \,, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{11}$$

where f_t is a $q \times 1$ vector of latent common factors, u_t is a $N \times 1$ vector of idiosyncratic components, ν_t is a $N \times 1$ vector of white noise that is uncorrelated with the $q \times 1$ vector of white noise η_t, $\lambda(L)$, $\delta(L)$, $\gamma(L)$ and $\theta(L)$ are matrices of lag polynomials.

We have:

$$\underset{(N\times q)}{\lambda(L)} = \sum_{k=0}^{p_\lambda - 1} \lambda_k L^k; \qquad \underset{(N\times N)}{\delta(L)} = \sum_{k=0}^{p_\delta - 1} \delta_k L^k,$$

$$\underset{(q\times q)}{\gamma(L)} = \sum_{k=0}^{p_\gamma - 1} \gamma_k L^k; \qquad \underset{q\times q}{\theta(L)} = I_q - \sum_{k=1}^{p_\theta} \theta_k L^{q_f}$$

with p_λ, p_δ, p_γ, $p_\theta \geq 1$ are the highest degrees of polynomials in each matrix. Indeed, the matrices of coefficients λ_k, δ_k, γ_k and θ_k are allowed to become sparse as k increases to the maximum degrees so that the orders of the polynomials in a given matrix may vary.

For instance, the i^{th} element of X_t is represented as:

$$X_{it} = \sum_{k=0}^{p_\lambda - 1} \lambda_{k,i} f_{t-k} + u_{i,t} \equiv \lambda^{(i)}(L) f_t + u_{it}, \tag{12}$$

$$u_{it} = \sum_{k=0}^{p_\delta - 1} \delta_{k,i} u_{i,t-1-k} + \nu_{it} \equiv \delta^{(i)}(L)\, u_{i,t-1} + \nu_{it}, \tag{13}$$

where $\lambda_{k,i}$ is the i^{th} row of λ_k, $\lambda^{(i)}(L) = \sum_{k=0}^{p_\lambda - 1} \lambda_{k,i} L^k$, $\delta_{k,i}$ is the i^{th} row of δ_k and $\delta^{(i)}(L) = \sum_{k=0}^{p_\delta - 1} \delta_{k,i} L^k$.

The exact DFM is obtained if the following assumption is satisfied:

$$E(u_{it} u_{js}) = 0 \,, \ \forall i, j, t, s, \quad i \neq j.$$

The approximate DFM is obtained by allowing for some limited cross-section correlations among the idiosyncratic components.[3] We assume the idiosyncratic errors ν_{it} are uncorrelated with the factors f_t at all leads and lags.

To obtain the static factor representation, we define $F_t = [f_t', f_{t-1}', \dots, f_{t-p_\lambda+1}']'$, a vector of size $K = qp_\lambda$ such that:

$$\begin{aligned} X_t &= \Lambda F_t + u_t, & (14)\\ u_t &= \delta(L)u_{t-1} + \nu_t, & (15)\\ F_t &= \Gamma F_{t-1} + \Theta(L)\eta_t, & (16) \end{aligned}$$

where

$$\underset{(N\times qp_\lambda)}{\Lambda} = \begin{bmatrix} \lambda_0 & \lambda_1 & \dots & \lambda_{p_\lambda-1} \end{bmatrix}$$

$$\underset{(qp_\lambda\times qp_\lambda)}{\Gamma(L)} = \begin{bmatrix} \gamma_0 & \gamma_1 & \dots & \gamma_{p_\gamma-1} \\ 0 & I & 0 & \dots \\ \vdots & \ddots & \ddots & \ddots \\ 0 & \dots & 0 & I \end{bmatrix}; \quad \underset{(qp_\lambda\times q)}{\Theta(L)} = \begin{bmatrix} \theta(L) \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Equations (14)-(16) define the FAVARMA model proposed in (Dufour & Stevanovic 2013). A simplified version of this model where $p_\lambda = 1$ (so that $K = q$ and $\Theta(L) = \theta(L)$) has been used in (Bedock & Stevanovic 2016) to estimate the effects of credit shocks. A similar model with $\theta(L) = I_q$ has been used to forecast time series in (Boivin & Ng 2005) and study the impact of monetary policy shocks in (Bernanke, Boivin & Eliasz 2005).

In practice, q and p_λ cannot be separately identified due to the latent nature of f. Therefore, we shall rewrite (16) in the static representation as a standard K-dimensional VARMA with no particular structure imposed on the matrices of coefficients. We have:

$$F_t = \Phi(L)F_{t-1} + \Theta(L)\eta_t, \tag{17}$$

where $\Phi(L) = \sum_{k=0}^{p_\phi-1} \phi_k L^k$ and $\Theta(L)$ is redefined as $\Theta(L) = \sum_{k=0}^{p_\theta-1} \theta_k L^k$. The optimal values of p_ϕ and p_θ can be selected by BIC.

[3]Intuitively, only a small number of largest eigenvalues of the covariance matrix of the common component, $\tilde{\lambda}_i(L)f_t$, may diverge when the number of series tends to infinity, while the remaining eigenvalues as well as the eigenvalues of the covariance matrix of specific components are bounded. See technical details in (Stock & Watson 2005) and (Bai & Ng 2008).

3.1 Factor-Augmented Regressions

The first category of forecasting models considered below are the factor-augmented regressions, where an autoregressive direct model is augmented with estimated static factors. In these models, there is no need to specify the dynamics of the factors as in (16) because static factors are extracted by principal component analysis. The second category of models are more directly related to the DFM model presented previously.

Diffusion Indices (ARDI) The first model is the (direct) autoregression augmented with diffusion indices from (Stock & Watson 2002b):

$$
\begin{aligned}
y_{t+h}^{(h)} &= \alpha^{(h)} + \sum_{l=1}^{p_y^h} \rho_l^{(h)} y_{t-l+1} + \sum_{l=1}^{p_f^h} F_{t-l+1} \beta_l^{(h)} + e_{t+h}, \quad t = 1, \ldots, T && (18) \\
X_t &= \Lambda F_t + u_t && (19)
\end{aligned}
$$

where F_t are $K^{(h)}$ *consecutive static* factors and the superscript h stands for the value of K when forecasting h periods ahead. The optimal values of p_y^h, p_f^h and $K^{(h)}$ are simultaneously selected by BIC. The h-step ahead forecast is obtained as:

$$
\hat{y}_{T+h|T}^{h} = \hat{\alpha}^{(h)} + \sum_{l=1}^{p_y^h} \widehat{\rho}_l^{(h)} y_{T-l+1} + \sum_{l=1}^{p_f^h} F_{T-l+1} \widehat{\beta}_l^{(h)}.
$$

The feasible ARDI model is obtained after estimating F_t as the first $K^{(h)}$ principal components of X_t. See (Stock & Watson 2002) for technical details on the estimation of F_t as well as their asymptotic properties.

Below, we consider two variations of the ARDI model. In the first version, we select only a subset of $K^{(h)}$ factors to be included in (18) while in the second the F_t are obtained as dynamic principal components.

Variation I: ARDI-tstat The importance of the factors as predictors of $y_{t+h}^{(h)}$ may be independent of their importance as principal components. Indeed, the ordering of the factors in F_t is related to their capacity to explain the (co-)variations in X_t. The selection of factors into the ARDI model automatically includes the first $K^{(h)}$ principal components. A natural variation of this approach is to select only those that have significant coefficients in the regression (18). This leads to forecast $y_{t+h}^{(h)}$ as:

$$\hat{y}^h_{T+h|T} = \hat{\alpha}^{(h)} + \sum_{l=1}^{p_y^h} \widehat{\rho}_l^{(h)} y_{T-l+1} + \sum_{i \in K^*} \hat{F}_{i,T} \beta_i^{(h)} \quad (20)$$
$$K^* = \{i \in 1, \ldots, K \mid t_i > t_c\}.$$

where $K^* \in K$ refers to elements of F_t corresponding to coefficients β_i^h having their t-stat larger (in absolute terms) than the critical value t_c (here we omit the superscript h for simplicity). Another difference with respect to the ARDI model is that the optimal number of factors changes over time.

Variation II: ARDI-DU The second variation of the ARDI model is taken from (Boivin & Ng 2005). The model is the same as the ARDI except that F_t is estimated by one-sided generalized principal components as in (Forni et al. 2005). Hence, the working hypothesis behind the dimensionality reduction is the DFM equation (9).

Targeted Diffusion Indices (ARDIT) Another critique of the ARDI model is that not necessarily all series in X_t are equally important to predict $y^{(h)}_{t+h}$. The ARDIT model of (Bai & Ng 2008) takes this aspect into account. Instead of shrinking the factors space as in ARDI-tstat variation, the idea is first to pre-select a subset X_t^* of the series in X_t that are relevant for forecasting $y^{(h)}_{t+h}$ and next predict the factors using this subset. (Bai & Ng 2008) propose two ways to construct the subset X_t^*:

- Hard threshold (OLS): **ARDIT-hard**

$$y^{(h)}_{t+h} = \alpha^{(h)} + \sum_{j=0}^{3} \rho_j^{(h)} y_{t-j} + \beta_i^{(h)} X_{i,t} + \epsilon_t \quad (21)$$
$$X_t^* = \{X_i \in X_t \mid t_{Xi} > t_c\} \quad (22)$$

- Soft threshold (LASSO): **ARDIT-soft**

$$\hat{\beta}^{lasso} = arg\,min_\beta \left[RSS + \lambda \sum_{i=1}^{N} |\beta_i| \right] \quad (23)$$
$$X_t^* = \{X_i \in X_t \mid \beta_i^{lasso} \neq 0\} \quad (24)$$

In the hard threshold case, a univariate regression (21) is performed for each predictor X_{it} at the time. The subset X_t^* is then obtained by gathering those series whose coefficients $\beta_i^{(h)}$ have their t-stat larger than the critical value t_c. We follow (Bai & Ng 2008) and consider 3 lags of y_t in (21), and set t_c to 1.28 and 1.65. The second approach uses the LASSO technique to select X_t^* by regressing y_{t+h}^h on all elements of X_t and using LASSO penalty to discard uninformative predictors.[4]

3.2 Factor-Structure-Based Models

The second category of forecasting models relies directly on the factor structure when predicting the series of interest. The working hypothesis will be the DFM (9)-(11) or its static form (SFM) (14)- (16) with some variations. Another important difference is that the series of interest, y_t, is now included in the informational set X_t.

Factor-Augmented VAR (FAVAR) Suppose that X_t obeys the SFM representation (14)-(16) with $\Theta(L) = \theta(L) = I$. We have:

$$
\begin{aligned}
X_t &= \Lambda F_t + u_t && (25)\\
u_t &= \delta(L)u_{t-1} + v_t && (26)\\
F_t &= \Phi F_{t-1} + \eta_t. && (27)
\end{aligned}
$$

This model implicitly assumes that $p_\lambda = 1$ so that $K = q$ and F_t reduces to a first order VAR. The optimal order of the polynomial $\delta(L)$ is selected with BIC while the optimal number of static factors is chosen by (Bai & Ng 2002) IC_{p2} criterion. After estimation, one forecasts the factors using (27) upon assuming stationarity. The idiosyncratic component is predicted using (26) and then $\widehat{F}_t$ and $\widehat{u}_t$ are combined into (25) to obtain a prediction of X_t. (Boivin & Ng 2005) compare the direct and iterative approaches:

- Iterative

$$
\begin{aligned}
\hat{F}_{T+h|T} &= \widehat{\Phi}\hat{F}_{T+h-1|T}\\
\hat{u}_{T+h|T} &= \hat{\delta}(L)\hat{u}_{T+h-1|T}\\
\hat{X}_{T+h|T} &= \hat{\Lambda}\hat{F}_{T+h|T} + \hat{u}_{T+h|T}
\end{aligned}
$$

- Direct

$$
\begin{aligned}
\hat{F}^{(h)}_{T+h|T} &= \widehat{\Phi}^{(h)}\hat{F}_T\\
\hat{u}^{(h)}_{T+h|T} &= \hat{\delta}(L)^{(h)}\hat{u}_T\\
\hat{X}^{(h)}_{T+h|T} &= \hat{\Lambda}\hat{F}^{(h)}_{T+h|T} + \hat{u}^{(h)}_{T+h|T}
\end{aligned}
$$

[4] As in (Bai & Ng 2008) we target 30 series. It is possible to optimally select the number of retained series, but the procedure is very long and (Bai & Ng 2008) did not find significant improvements.

The forecast of interest, $\hat{y}^{(h)}_{T+h|T}$, is then extracted from $\hat{X}_{T+h|T}$ or $\hat{X}^{(h)}_{T+h|T}$. The accuracy of the predictions depends on the validity of the restrictions imposed by the factor model. As ARDI type models are simple predictive regressions, they are likely to be more robust to misspecification than the factor model.

Factor-Augmented VARMA (FAVARMA) (Dufour & Stevanovic 2013) show that the dynamics of the factors should be modeled as a VARMA and suggest the class of Factor-Augmented VARMA models represented in (14)-(16). Since the VARMA representation is not estimable in general, they suggest four identified forms of Equation (16): Final AR (FAR), Final MA (FMA), Diagonal AR (DAR) and Diagonal MA (DMA). Only the iterative version is considered:

$$\begin{aligned}\hat{F}_{T+h|T} &= \widehat{\Phi}\hat{F}_{T+h-1|T} + \sum_{k=1}^{p_\theta}\widehat{\theta}_k\hat{\eta}_{T+h-k|T} \\ \hat{u}_{T+h|T} &= \hat{\delta}(L)\hat{u}_{T+h-1|T} \\ \hat{X}_{T+h|T} &= \hat{\Lambda}\hat{F}_{T+h|T} + \hat{u}_{T+h|T}\end{aligned}$$

with $\hat{\eta}_{T+h-k|T} = 0$ if $h-k>0$. The forecast $\hat{y}^{h}_{T+h|T}$ is extracted from $\hat{X}_{T+h|T}$.

DFM Contrary to the FAVAR(MA) approach, (Forni et al. 2005) propose to use a nonparametric estimate of the common component to forecast the series of interest.[5] The forecasting formula for the idiosyncratic component remains the same. The forecast of X_t is constructed as follows:

$$\begin{aligned}\hat{u}_{T+h|T} &= \hat{\delta}(L)\hat{u}_{T+h-1|T} \\ \hat{X}_{T+h|T} &= \widehat{\lambda}(L)\widehat{f}_{T+h|T} + \hat{u}_{T+h|T}\end{aligned}$$

and $\hat{y}^{(h)}_{T+h|T}$ is extracted from $\hat{X}_{T+h|T}$. The number of underlying dynamic factors f_t is selected by (Hallin & Liska 2007)'s test. The advantage of the current approach over the FAVAR(MA) clearly lies in the nonparametric treatment of the common component, which might be more robust to misspecifications. However, the nonparametric method may struggle in finite samples.

[5] See (Boivin & Ng 2005) for discussion about this forecasting models. It is the 'DN' specification in their paper.

3.3 Other data-rich methods

We now present two recent methodologies that have been shown to compare favorably to ARDI models.

Three-Pass Regression Filter (3PRF) (Kelly & Pruitt 2015) propose another approach to construct predicting factors from a large data set. The factors approximation is in the spirit of the Famma-MacBeath two-step procedure:

1. Time series regression of X_{it} on Z_t for $i = 1, \ldots, N$

$$X_{i,t} = \phi_{0,i} + Z_t'\phi_i + \varepsilon_{i,t}$$

2. Cross-section regression of X_{it} on $\hat{\phi}_i$ for $t = 1, \ldots, T$

$$X_{i,t} = \varsigma_{0,t} + \hat{\phi}_i' f_t + \epsilon_{i,t}$$

3. Time series regression of $y_{t+h}^{(h)}$ on $\hat{f}_t$

$$y_{t+h}^{(h)} = \beta_0 + \beta \hat{f}_t' + \eta_{t+h}$$

4. Prediction

$$\hat{y}_{T+h|T}^{(h)} = \hat{\beta}_0 + \hat{\beta}\hat{f}_T$$

We follow (Kelly & Pruitt 2015) and use 4 lags of y_t as proxies for Z_t. They also suggest an information criterion to optimally select the proxy variables.

Complete Subset Regression (CSR) (Elliott et al. 2013) do not use directly the factor structure of the data. They generate a large number of forecasts of $y_{T+h|T}^{(h)}$ using several subsets of the predictors in X_t. The final forecast is then obtained as the average of the individual forecasts:

$$\hat{y}_{T+h|T,m}^{(h)} = \hat{c} + \hat{\rho} y_t + \hat{\beta} X_{t,m} \tag{28}$$

$$\hat{y}_{T+h|T}^{(h)} = \frac{\sum_{m=1}^{M} \hat{y}_{T+h|T,m}}{M} \tag{29}$$

where $X_{t,m}$ contains L series for each model $m = 1, \ldots, M$.[6] Note that this method can be computationally demanding when the number of predictors in X_t is large.

[6]In (Elliott et al. 2013) L is set to 1, 10 and 20. M is the maximum number of models that is set to 20,000 when needed.

4 Forecasts Combinations

Instead of looking at individual forecasts, one can also aggregate them into a single prediction.

Equal-Weighted Forecast (AVRG) The simplest, but often very robust, method is to set equal weights on each individual forecast, $w_{it} = \frac{1}{M}$, i.e. take a simple average over all forecasts:

$$y_{t+h|t}^{(h,ew)} = \frac{1}{M} \sum_{i=1}^{M} y_{t+h|t}^{(h,i)}$$

Trimmed Average (T-AVRG) Another approach consists of removing the most extreme forecasts. First, order the M forecasts from the lowest to the highest value $\left(y_{t+h|t}^{(h,1)} \leq y_{t+h|t}^{(h,2)} \ldots \leq y_{t+h|t}^{(h,M)} \right)$. Then trim a proportion λ of forecasts from both sides:

$$y_{t+h|t}^{(h,trim)} = \frac{1}{M(1-2\lambda)} \sum_{i=\lceil \lambda M \rceil}^{\lfloor (1-\lambda)M \rfloor} y_{t+h|t}^{(h,i)}$$

where $\lceil \lambda M \rceil$ is the integer immediately larger than λM and $\lfloor (1-\lambda)M \rfloor$ is the integer immediately smaller than $(1-\lambda)M$.

Inversely Proportional Average (IP-AVRG) A more flexible solution is to produce weights that depend inversely on the historical performance of individual forecasts as in (Diebold & Pauly 1987). Here, we follow (Stock & Watson 2004) and define the discounted weight on the i^{th} forecast as follows

$$w_{it} = \frac{m_{it}^{-1}}{\sum_{j=1}^{M} m_{jt}^{-1}},$$

where m_{it} is the discounted MSPE for the forecast i:

$$m_{it} = \sum_{s=T_0}^{t-h} \rho^{t-h-s} (y_{s+h} - y_{s+h|s}^{(h,i)})^2,$$

and ρ is a discount factor. In our applications, we consider $\rho = 1$ and $\rho = 0.95$.

Median Finally, instead of averaging forecasts one can use the median, another measure of central location, that is less subject to extreme values than the mean:

$$y_{t+h|t}^{(h,median)} = \text{median}(y_{t+h|t}^{(h,i)})_{i=1}^{M}$$

The median further avoids the dilemma regarding which proportion of forecasts to trim.

5 Empirical Performance of the Forecasting Models

We use historical data to evaluate and compare the performance of all the forecasting models described previously.[7] The data employed consists of an updated version of Stock and Watson macroeconomic panel available at Federal Reserve of St-Louis's web site (FRED). It contains 134 monthly macroeconomic and financial indicators observed from 1960M01 to 2014M12. Details on the construction of the series can be found in (McCracken & Ng 2015).

The empirical exercise is easier when the data set is balanced. In practice, there is usually a trade-off between the relevance and the availability (and frequency) of a time series. Not all series are available from the starting date 1960M01 in the (McCracken & Ng 2015) database, but this can be accommodated when a rolling window is used. Indeed, a series that is not available at the starting date will eventually appear in the informational set as the window moves forward.[8]

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 & Ng 2015) and (Stock & Watson 2002b) and assume that price indexes are all $I(2)$ while interest and unemployment rates are $I(1)$.[9]

5.1 Pseudo-Out-of-Sample Setup

The pseudo-out-of-sample period is 1970M01 - 2014M12. The forecasting horizons considered are 1, 2, 3, 4, 6, 8 and 12 months. There are 540 evaluation periods for each horizon. All models are estimated recursively on rolling windows. For each model, the optimal hyperparameters

[7] In principle, a real-time forecasting exercise could be preferable but not all the data are yet available in real-time vintages. Hence, we choose to evaluate the models with the most recent releases and not consider their performance in the presence of revisions.

[8] However, this is a problem when conducting a structural FAVAR analysis as in (Bernanke et al. 2005). Another source of unbalanced panels is mixing frequencies. (Stock & Watson 2002b) construct a monthly data set using monthly and quarterly series. They transform the quarterly series into monthly indicators using an expectation-maximization (EM) procedure that also works to fill the holes of unobserved monthly data points. This EM technique has also been used in (Boivin, Giannoni & Stevanović 2013) when estimating the effects of credit shocks.

[9] (Bernanke et al. 2005) keep inflation, interest and unemployment rates in levels in X_t. Choosing (SW) or (BBE) transformations has important effects on correlation patterns in X_t. Under (BBE), the group of interest rates is highly correlated as well as the inflation and unemployment rates. Hence, the principal components will tend to exploit these clusters such that the initial factors will be related to those groups of series. As pointed out by (Boivin & Ng 2006), the presence of these clusters may alter the estimation of *common* factors. Under (SW), these correlation clusters are less important. Recently, procedures have been proposed to deal directly with the unit root instead of differentiating the data, see (Banerjee, Marcellino & Masten 2014) and (Barigozzi, Lippi & Luciani 2016).

(number of factors, number of lags, etc.) are selected specifically for each evaluation period and forecasting horizon. The size of the rolling window is $120 - h$ months, where h is the forecasting horizon. The following metrics are used to compare the models: *(i)* mean squared predictive error (MSPE); *(ii)* mean absolute predictive error (MAPE); *(iii)* pseudo-R^2 from (Galbraith 2003); *(iv)* (Pesaran & Timmermann 1992) sign test. MSPE and MAPE will often be reported relative to ARD (autoregressive direct) used as a benchmark, along with the (Diebold & Mariano 1995) (DM) test.

We consider the prediction of four types of macroeconomic and financial variables: real activity, prices, stock market returns and exchange rates. As in (McCracken & Ng 2015), the real activity measures to be forecasted are Industrial Production (IP), Employment (EMP), while the Consumer Price Index (CPI) and the Core Consumer Price Index (CPICORE) represent the nominal sector. As usual, the stock market is represented by the SP500. Finally, we consider two US bilateral exchange rates: the UK (EXUSUK) and Canada (EXUSCA). The real activity series (IP and EMP) as well as the SP500 and the exchange rates are treated as I(1). The CPI and CPICORE are assumed to be I(2), as in (Stock & Watson 2002b) and (McCracken & Ng 2015).

5.2 Empirical Results

This subsection presents the main results separately for each series. There are in total 19 individual forecasts and five forecast combinations. The supplementary material contains additional results.

5.2.1 Real Activity

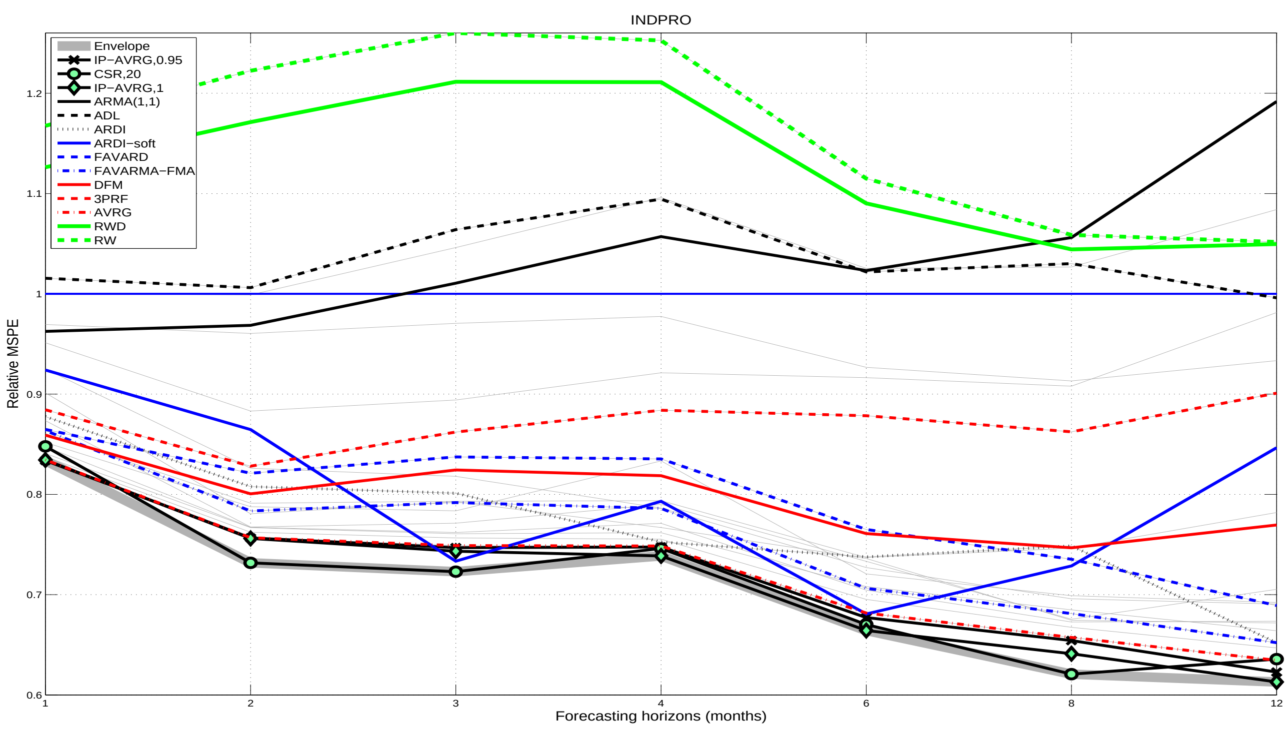
Industrial Production We first examine the performance of the various models for the industrial production. Figure (1) plots the relative MSPE (RMSPE) with respect to ARD for all models and all horizons. Table (1) reports the RMSPEs with the results of the standard Diebold-Mariano (DM) test at 1%, 5% and 10% significance levels.

First, we note that all data-rich models outperform the standard univariate models. Depending on the forecasting horizon, the relative efficiency gains approach 40%. The two dominating techniques to forecast the industrial production are the IP-AVRG of all forecasts and the CSR with 20 effective predictors (CSR-20). This clearly suggests that model averaging is the best approach (in the MSPE sense) to predict the annualized growth rate of industrial production. As shown by Figure (1), most data-rich models are close to the envelope. The 3PRF and the FAVARMA-FAR models are the worst performing data rich models.

Table (2) reports the minimum RMSPE and relative MAPE (RMAPE) values for the Full Out-of-sample period (1970-2014), and for NBER recession and non-recession periods. Again, model averaging dominates under the RMAPE criterion for the full sample. However, during NBER recessions the best models are FAVAR(MA) and ARDI-soft.[10] Note that the RMSPE and RMAPE are systematically smaller during recessions than during non-recession periods. This indicates that the relative efficiency gain of the data-rich models improves during recessions.

Finally, Figure (3) plots the realized industrial production growth, the forecast of the best RMSPE model as well as the distribution of all forecasts 1-3 periods ahead during the full out-of-sample period. The trajectories of forecasts delivered by the IP-AVRG and CSR track the actual industrial production quite well although they are less volatile. The most pessimistic forecasts appear to be the best predictor of the industrial production during downturns (see e.g. the 2007-09 recession). In a real-life application, the most pessimistic forecast can be used as a worse case scenario that becomes more realistic on the eve of economic crises.

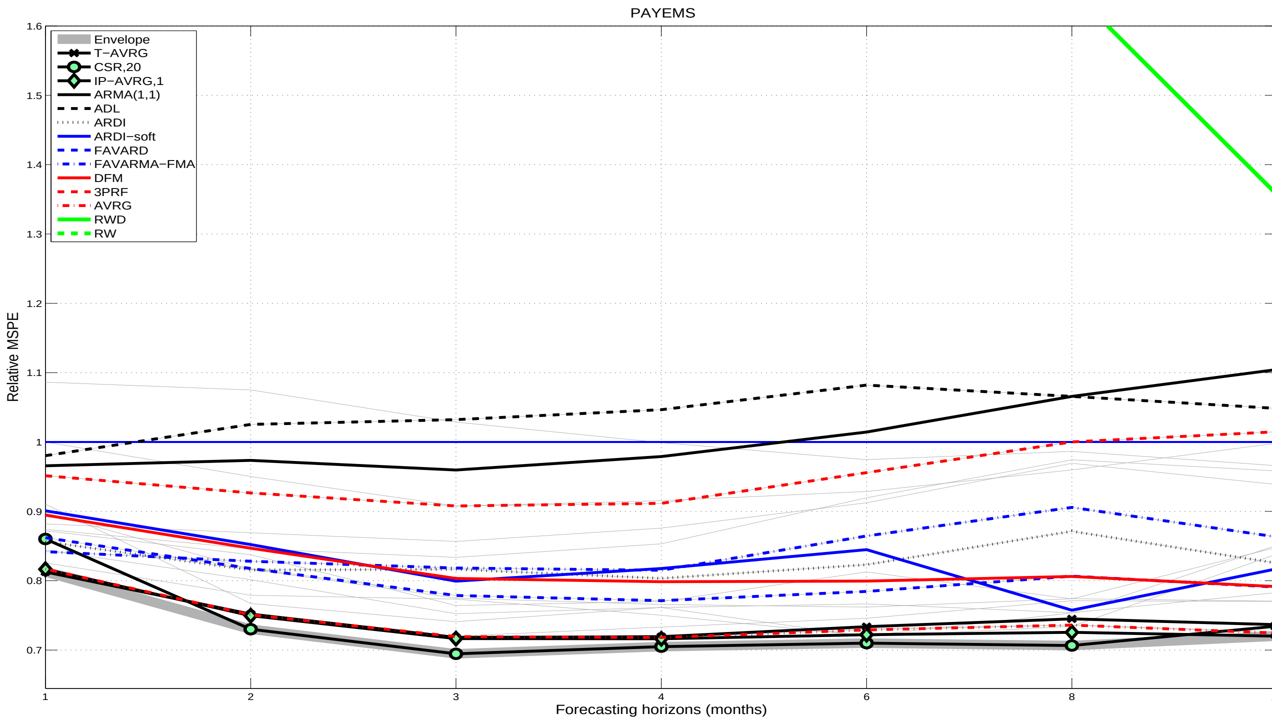
Figure 1: Relative MSPE: Industrial Production



Employment We now examine the empirical results for the employment growth. Figure (2) shows the RMSPE while Table (3) shows the results of the DM test for all models and horizons. Here too, it appears that forecast combination is the best technique to forecast employment growth. The two dominating methods are the CSR-20 and the IP-AVRG-1. Factor-based models (DFM, FAVARD and FAVARMA-FMA) continue to dominate univariate models as

[10] That is, if the horizon of interest $(t+h)$ belongs to recession episodes.

Figure 2: Relative MSPE: Employment



The figure shows the MSPE of all models relative to ARD. When the value is below the blue line the corresponding model produces smaller MSPE than ARD. The thick gray line shows the inferior envelope, i.e. the lowest RMSPE for each horizon.

well. The ARMA(1,1) dominates the ARD model at between 1 and 4 horizons but not at longer horizons.

Table (4) shows the best performing models in terms of RMSPE and RMAPE under recession and non-recession episodes. The results suggest that employment growth is better forecasted by the ARDI-tstat and ARDI-hard models during recessions. This finding is robust to the metrics used to evaluate the forecasts. Forecast combination approaches dominate during expansions.

Figure (4) shows the trajectories of forecasts of the best performing models as well as the actual series and the distribution of all forecasts for horizons 1 to 3. We see that the best performing models track the actual data quite well, although they are slightly optimistic during recessions and pessimistic during expansions. This is not surprising since forecasts have to be smoother than the actual data.

Table 1: Relative MSPE for INDPRO

	h=1	h=2	h=3	h=4	h=6	h=8	h=12
Standard Time Series Models							
ARI	1.000	0.999	1.046*	1.096**	1.025	1.027	1.084
ARMA(1,1)	0.963	0.969	1.011	1.057	1.023	1.056	1.192*
ADL	1.016	1.006	1.064**	1.094**	1.022	1.030	0.996
Factor-Augmented Regressions							
ARDI	0.878**	0.808***	0.801***	0.753**	0.738**	0.749**	0.652***
ARDI-soft	0.924	0.865*	0.733***	0.793**	0.681**	0.729**	0.847
ARDI-hard,1.28	0.901*	0.781***	0.793***	0.767**	0.736**	0.674***	0.672***
ARDI-hard,1.65	0.874**	0.768***	0.771***	0.790**	0.733**	0.676***	0.705**
ARDI-tstat,1.96	0.925	0.826**	0.818**	0.786**	0.727***	0.696**	0.691***
ARDI-DU	0.852***	0.788***	0.784***	0.833**	0.721**	0.699**	0.692**
Factor-Structure-Based Models							
FAVARI	0.861**	0.791***	0.793***	0.794***	0.737***	0.746***	0.782***
FAVARD	0.865**	0.821**	0.837**	0.835*	0.765**	0.735**	0.689**
FAVARMA-FMA	0.864**	0.783***	0.792***	0.786***	0.706***	0.681***	0.652***
FAVARMA-FAR	0.951	0.883*	0.894	0.921	0.916	0.908	0.981
DFM	0.859***	0.801***	0.824***	0.819***	0.761***	0.747***	0.769***
Other Data-Rich Models							
3PRF	0.884**	0.828***	0.862**	0.884**	0.878**	0.862***	0.901*
CSR,1	0.969*	0.960	0.971	0.977	0.927**	0.913***	0.933***
CSR,10	0.848***	0.767***	0.762***	0.771***	0.705***	0.673***	0.674***
CSR,20	0.848***	**0.732*** **	**0.723*** **	0.746***	0.670***	**0.621*** **	0.636***
Forecasts Combinations							
AVRG	0.834***	0.757***	0.749***	0.749***	0.682***	0.657***	0.635***
Median	0.841***	0.767***	0.761***	0.762***	0.708***	0.685***	0.664***
T-AVRG	0.838***	0.762***	0.757***	0.755***	0.695***	0.667***	0.647***
IP-AVRG,1	0.834***	0.756***	0.743***	**0.739*** **	**0.664*** **	0.641***	**0.613*** **
IP-AVRG,0.95	**0.834*** **	0.757***	0.747***	0.747***	0.677***	0.654***	0.623***

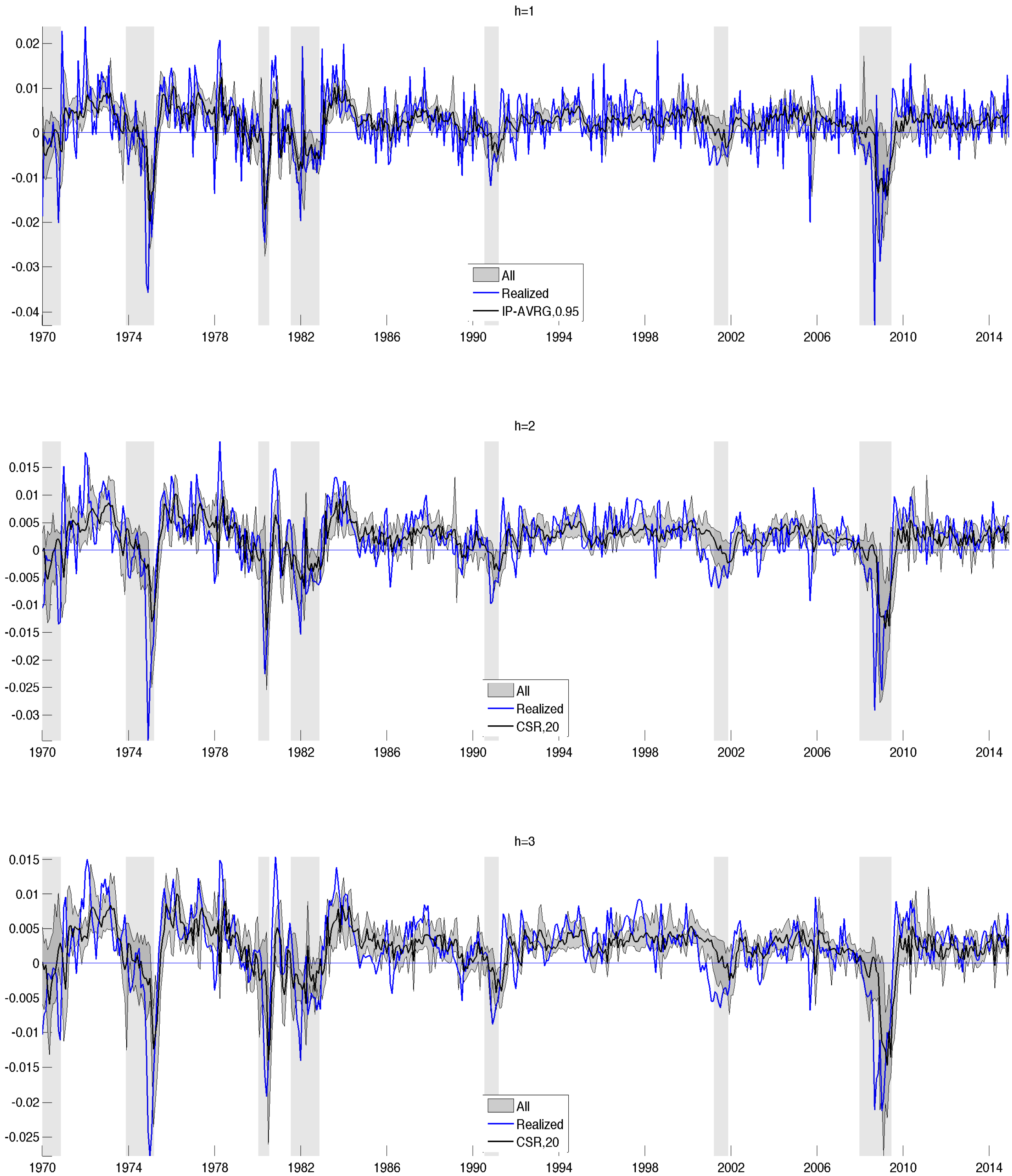
*Note: Minimum values are in bold, while ***, **, * stand for 1%, 5% and 10% significance of Diebold-Mariano test.*

Table 2: Relative MSPE and MAPE across business cycles for INDPRO

	Full Out-of-Sample		NBER Recessions		NBER Non-Recessions	
	RMSPE	RMAPE	RMSPE	RMAPE	RMSPE	RMAPE
h=1	IP-AVRG,0.95	CSR,20	FAVARMA-FMA	FAVARI	CSR,10	CSR,10
	0.834***	0.913***	0.640***	0.795***	0.901***	0.942***
h=2	CSR,20	CSR,20	FAVARMA-FMA	FAVARI	CSR,10	CSR,10
	0.732***	0.870***	0.607***	0.733***	0.786***	0.906***
h=3	CSR,20	CSR,20	ARDI-soft	ARDI-soft	CSR,10	CSR,10
	0.723***	0.882***	0.547***	0.707***	0.767***	0.903***
h=4	IP-AVRG,1	CSR,10	ARDI-soft	ARDI-soft	CSR,10	CSR,10
	0.739***	0.879***	0.581***	0.729***	0.772***	0.901***
h=6	IP-AVRG,1	CSR,20	FAVARD	FAVARMA-FMA	CSR,10	CSR,10
	0.664***	0.835***	0.484***	0.648***	0.722***	0.882***
h=8	CSR,20	CSR,20	FAVARMA-FAR	FAVARMA-FMA	CSR,20	CSR,20
	0.621***	0.798***	0.533***	0.667***	0.635***	0.830***
h=12	IP-AVRG,1	IP-AVRG,1	FAVARMA-FMA	FAVARMA-FMA	IP-AVRG,1	IP-AVRG,1
	0.613***	0.770***	0.526***	0.682***	0.585***	0.775***

*Note: This table shows the minimum relative (to ARD model) MSPE and MAPE values as well as the corresponding models for the full out-of-sample and during recessions and non-recession periods during recessions as declared by the NBER. ***, **, * stand for 1%, 5% and 10% significance of Diebold-Mariano test.*

Figure 3: Out-of-sample forecasts: Industrial Production



The figure shows the pseudo-out-of-sample forecasts of the Industrial Production annualized monthly growth rate for horizons 1, 2, and 3 months. The bleu line presents the historical data and the black line the forecast of the best MSPE model. The gray area around these lines presents the forecasts of all models considered in this exercise.

Table 3: Relative MSPE for Employment

	h=1	h=2	h=3	h=4	h=6	h=8	h=12
Standard Time Series Models							
ARI	1.000	0.950**	0.908**	0.915**	0.929**	0.960*	0.998
ARMA(1,1)	0.966	0.974	0.960	0.979	1.014	1.066	1.105
ADL	0.980	1.025	1.032*	1.047*	1.082*	1.066	1.048
Factor-Augmented Regressions							
ARDI	0.856*	0.815**	0.816**	0.803**	0.823**	0.871*	0.825*
ARDI-soft	0.901	0.852	0.799*	0.818*	0.845	0.758**	0.817*
ARDI-hard,1.28	0.826**	0.779**	0.773**	0.750**	0.720***	0.753**	0.851
ARDI-hard,1.65	0.844*	0.801**	0.752**	0.761**	0.720***	0.734***	0.839*
ARDI-tstat,1.96	0.911	0.767**	0.741**	0.761**	0.767**	0.754***	0.783**
ARDI-DU	0.872*	0.838*	0.764**	0.770**	0.813**	0.774**	0.848*
Factor-Structure-Based Models							
FAVARI	0.882	0.869	0.857	0.876	0.912	0.969	0.939
FAVARD	0.862*	0.818*	0.779**	0.771**	0.785**	0.807**	0.791***
FAVARMA-FMA	0.842*	0.828*	0.818*	0.815*	0.865	0.906	0.863**
FAVARMA-FAR	0.874	0.846*	0.834*	0.853*	0.920	0.974	0.958
DFM	0.895	0.847*	0.803**	0.799**	0.799**	0.806***	0.791***
Other Data-Rich Models							
3PRF	0.951	0.927	0.908	0.912	0.956	1.000	1.015
CSR,1	1.086***	1.075*	1.029	0.999	0.975	0.987	0.966**
CSR,10	0.898**	0.817***	0.775***	0.766***	0.762***	0.774***	0.771***
CSR,20	0.860*	**0.730*** **	**0.695*** **	**0.705*** **	**0.710*** **	**0.707*** **	0.735***
Forecasts Combinations							
AVRG	0.817**	0.751***	0.719***	0.718***	0.729***	0.736***	0.724***
Median	0.820**	0.748***	0.720***	0.733***	0.746***	0.771***	0.770***
T-AVRG	**0.813** **	0.752***	0.718***	0.719***	0.734***	0.745***	0.737***
IP-AVRG,1	0.817**	0.750***	0.716***	0.716***	0.722***	0.726***	**0.720*** **
IP-AVRG,0.95	0.817**	0.750***	0.718***	0.721***	0.731***	0.736***	0.727***

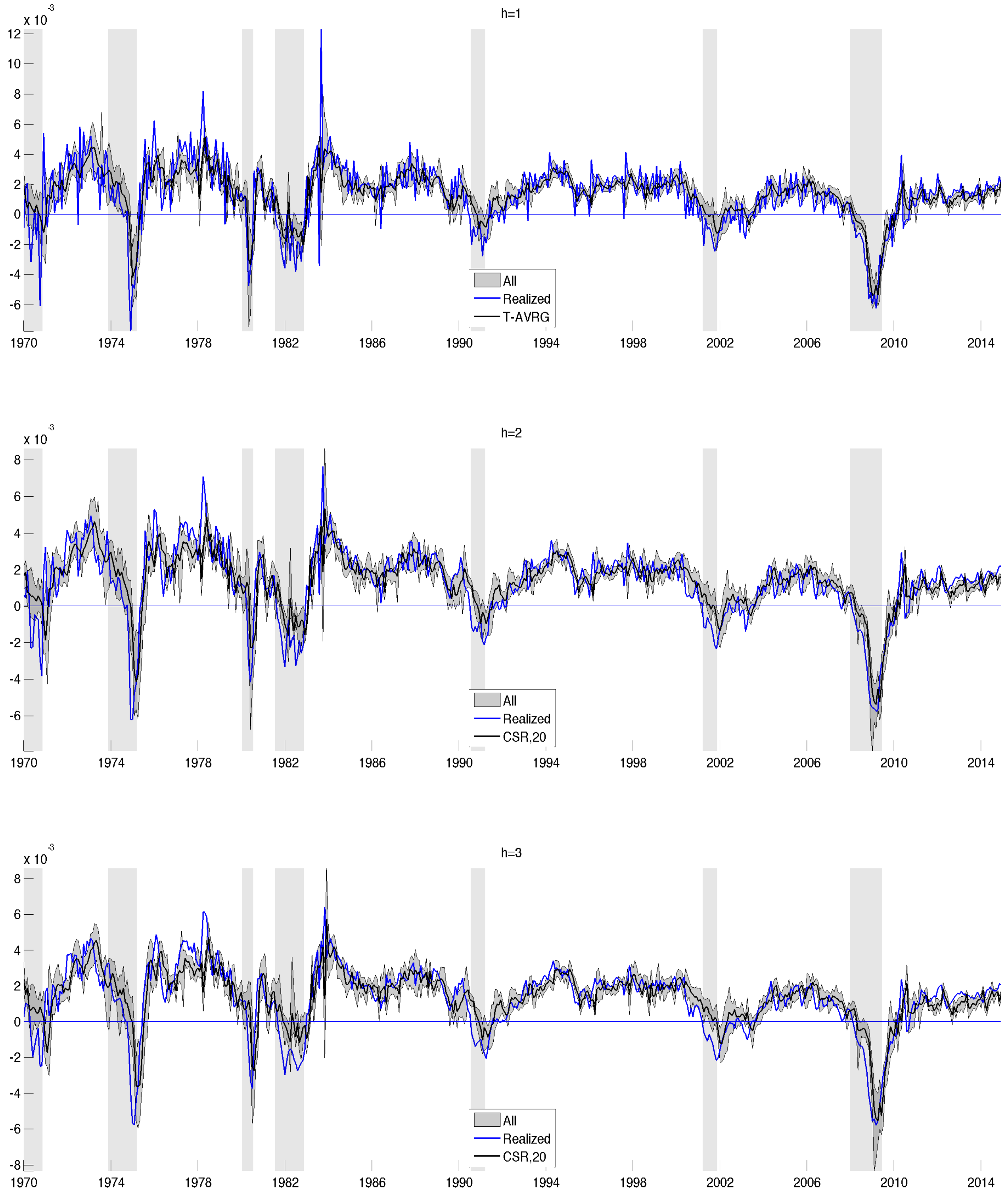
*Note: Minimum values are in bold, while ***, **, * stand for 1%, 5% and 10% significance of Diebold-Mariano test.*

Table 4: Relative MSPE and MAPE across business cycles for Employment

	Full Out-of-Sample		NBER Recessions		NBER Non-Recessions	
	RMSPE	RMAPE	RMSPE	RMAPE	RMSPE	RMAPE
h=1	T-AVRG	IP-AVRG,0.95	ARDI-soft	ARDI	T-AVRG	T-AVRG
	0.813**	0.934***	0.659**	0.838***	0.835*	0.944**
h=2	CSR,20	CSR,20	ARDI-tstat,1.96	ARDI-tstat,1.96	Median	CSR,20
	0.730***	0.884***	0.636***	0.817***	0.724**	0.885***
h=3	CSR,20	CSR,20	ARDI-tstat,1.96	ARDI-tstat,1.96	Median	AVRG
	0.695***	0.873***	0.637***	0.796***	0.670**	0.872***
h=4	CSR,20	AVRG	ARDI-tstat,1.96	ARDI-tstat,1.96	T-AVRG	AVRG
	0.705***	0.864***	0.681**	0.822***	0.639***	0.842***
h=6	CSR,20	CSR,20	ARDI-hard,1.28	ARDI-hard,1.28	AVRG	AVRG
	0.710***	0.856***	0.653***	0.791***	0.665***	0.845***
h=8	CSR,20	CSR,20	ARDI-hard,1.65	ARDI-hard,1.65	IP-AVRG,1	IP-AVRG,1
	0.707***	0.845***	0.688**	0.773***	0.666***	0.846***
h=12	IP-AVRG,1	IP-AVRG,1	CSR,20	CSR,20	AVRG	AVRG
	0.720***	0.834***	0.702***	0.771***	0.700***	0.835***

*Note: This table shows the minimum relative (to ARD model) MSPE and MAPE values as well as the corresponding models for the full out-of-sample and during recessions and non-recession periods during recessions as declared by the NBER. ***, **, * stand for 1%, 5% and 10% significance of Diebold-Mariano test.*

Figure 4: Out-of-sample forecasts: Employment



The figure shows the pseudo-out-of-sample forecasts of the Employment annualized monthly growth rate for horizons 1, 2, and 3 months. The bleu line presents the historical data and the black line the forecast of the best MSPE model. The gray area around these lines presents the forecasts of all models considered in this exercise.

5.2.2 Prices

CPI We now examine the performance of the various models at forecasting the inflation change deduced from the consumer price index (CPI). Indeed, the series of interest here is the second difference of the CPI, which basically is the CPI acceleration. Figure (5) shows the RMSPE while Table (5) displays the significance levels for the DM tests. Surprisingly, the ARMA(1,1) dominates all data-rich models at almost all forecasting horizons (from $h = 1$ to 8). The performance of data-rich models first deteriorate as h increases and then start improving at long horizons ($h > 6$), but not enough to be attractive in front of the ARD and the ARMA(1,1). The efficiency gains of the best performing models over the ARD benchmark are lower than what we obtained for the industrial production growth and employment growth. This may be suggesting that as a nominal variable, inflation change is harder to predict than real variables. The best performing data-rich model is the IP-AVRG-1, which slightly dominates the ARD. Factor based models are not recommended for predicting inflation growth. In particular, FAVARMA should be avoided.

Table (6) shows the best performing models for each forecast horizon during expansions and recessions. The results confirm the superiority of the ARMA(1,1) model. The ARDI-hard appears as the winning model at long horizons ($h \geq 8$) during recessions. During expansions, however, the IP-AVRG-0.95 does better than the ARMA(1,1) at horizons $h \geq 8$.

Figure (7) shows the trajectories of the best performing models in terms of RMSPE and RMAPE for horizon $h = 1$ to $h = 3$. Clearly, the inflation growth is much harder to track than the two previous real activity variables. Indeed, the trajectories of forecasts delivered by the best performing models are much smoother than the realized data, especially during non-recession episodes.

Core CPI We now present the empirical results for the inflation growth deduced from the Core consumer price index (CPI). Figure (6) shows the RMSPE of each model at all horizons while Table (7) shows the results of formal DM tests. The results confirm the domination of the ARMA(1,1) model at predicting inflation growth. Indeed, data-rich models perform worse than previously. Taking the simplicity and ease of implementation into account, we must recognize that the ARD model dominates all data-rich models as well. One possible explanation for this good performance of univariate models is that inflation growth is largely exogenous with respect to the conditioning information set available to us. This explanation implies that all data-rich models are over-parameterized for this series and are therefore doomed to perform poorly out of samples.

Table (8) shows that the ARDI and ARDI-hard models slightly dominate at the horizon $h = 12$, but the previous Figure and Table shows that the improvement over the benchmark is tiny. Figure (8) confirms the difficulty of the best forecasting models at replicating the volatility of the inflation growth.

Figure 5: Relative MSPE: CPI

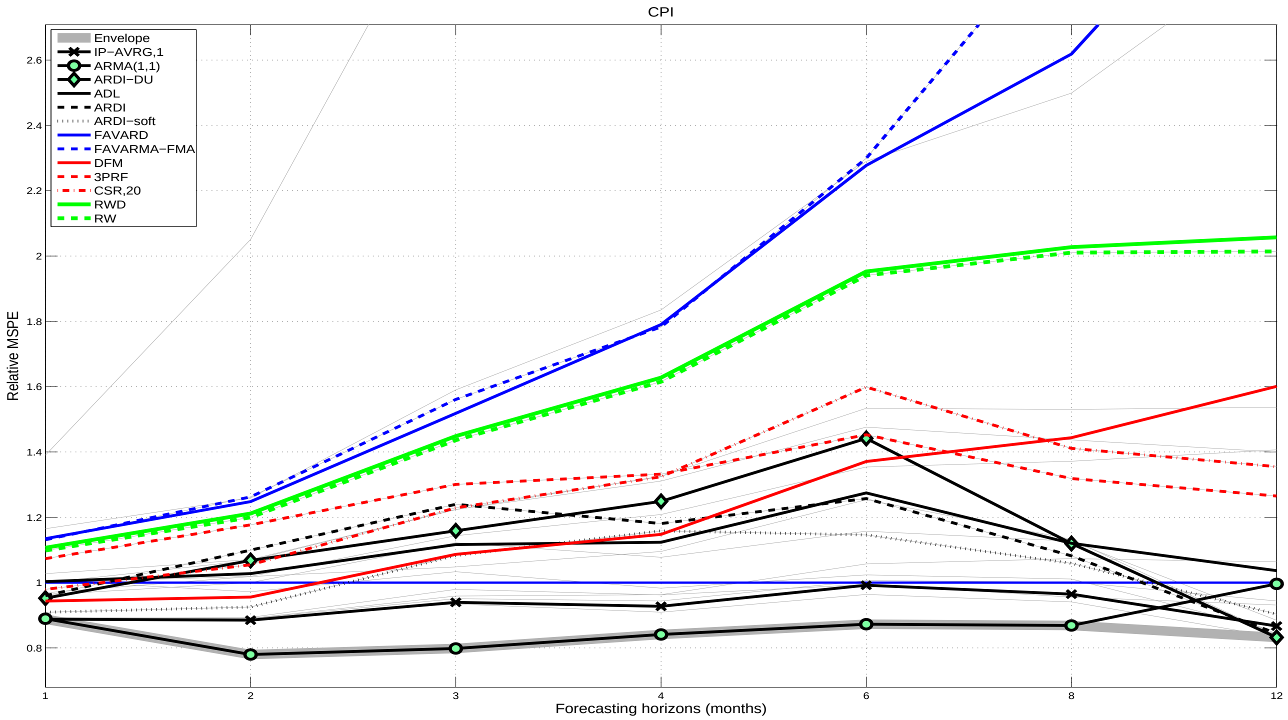
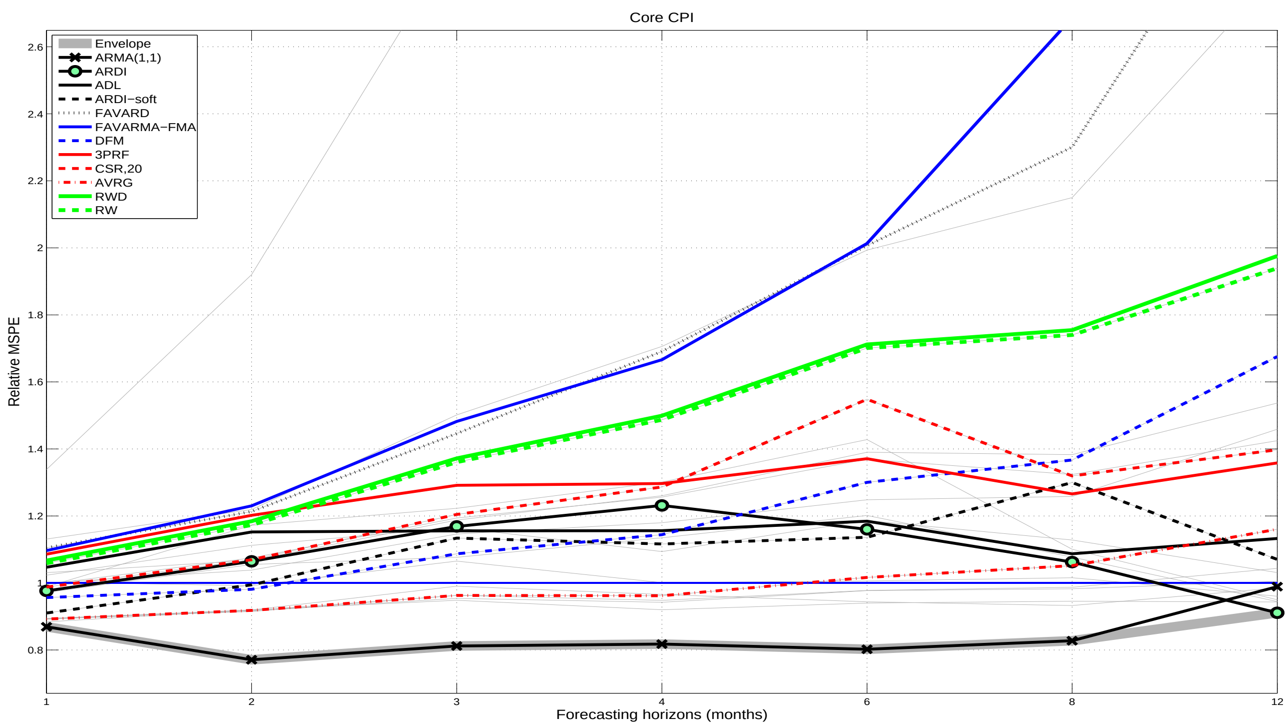


Figure 6: Relative MSPE: Core CPI



The figure shows the MSPE of all models relative to ARD. When the value is below the blue line the corresponding model produces smaller MSPE than ARD. The thick gray line shows the inferior envelope, i.e. the lowest RMSPE for each horizon.

Table 5: Relative MSPE for CPI

	h=1	h=2	h=3	h=4	h=6	h=8	h=12
Standard Time Series Models							
ARI	1.000	1.019	1.143**	1.209**	1.354***	1.372***	1.407**
ARMA(1,1)	0.889**	**0.780****	**0.798****	**0.841****	**0.873**	**0.869**	0.996
ADL	1.003	1.027	1.116	1.124	1.275	1.120	1.037
Factor-Augmented Regressions							
ARDI	0.960	1.099	1.240	1.181	1.257	1.082	0.844**
ARDI-soft	0.909	0.925	1.083	1.158	1.146	1.059	0.903
ARDI-hard,1.28	0.969	1.018	1.048	1.095	1.254	1.133	0.838*
ARDI-hard,1.65	0.963	1.000	1.122	1.077	1.158	1.127	0.884
ARDI-tstat,1.96	1.007	0.972	1.033	0.983	1.024	1.011	0.852**
ARDI-DU	0.952*	1.068	1.159	1.249	1.441	1.120	**0.832****
Factor-Structure-Based Models							
FAVARI	1.165***	1.259***	1.590***	1.835***	2.290***	2.499***	2.953**
FAVARD	1.134**	1.248**	1.518***	1.790***	2.277***	2.618***	3.308***
FAVARMA-FMA	1.131**	1.262**	1.561***	1.783***	2.299***	3.045**	2.921**
FAVARMA-FAR	1.390***	2.051***	3.195***	4.430***	7.364***	10.502***	15.809***
DFM	0.944*	0.956	1.086	1.148	1.371**	1.444**	1.601*
Other Data-Rich Models							
3PRF	1.073*	1.177**	1.301***	1.332***	1.453***	1.319***	1.265***
CSR,1	1.027	1.071	1.224**	1.326***	1.534***	1.530***	1.537***
CSR,10	0.999	1.068	1.224**	1.311**	1.476***	1.437***	1.400**
CSR,20	0.979	1.055	1.229**	1.324**	1.599***	1.411***	1.355**
Forecasts Combinations							
AVRG	0.889***	0.891**	0.959	0.963	1.057	1.073	1.064
Median	0.889**	0.894**	0.979	0.962	0.994	0.956	0.931
T-AVRG	0.894**	0.890**	0.951	0.944	1.006	1.002	0.944
IP-AVRG,1	**0.888*****	0.885**	0.940	0.927*	0.992	0.965	0.867*
IP-AVRG,0.95	0.891***	0.891**	0.935	0.909**	0.964	0.941	0.833**

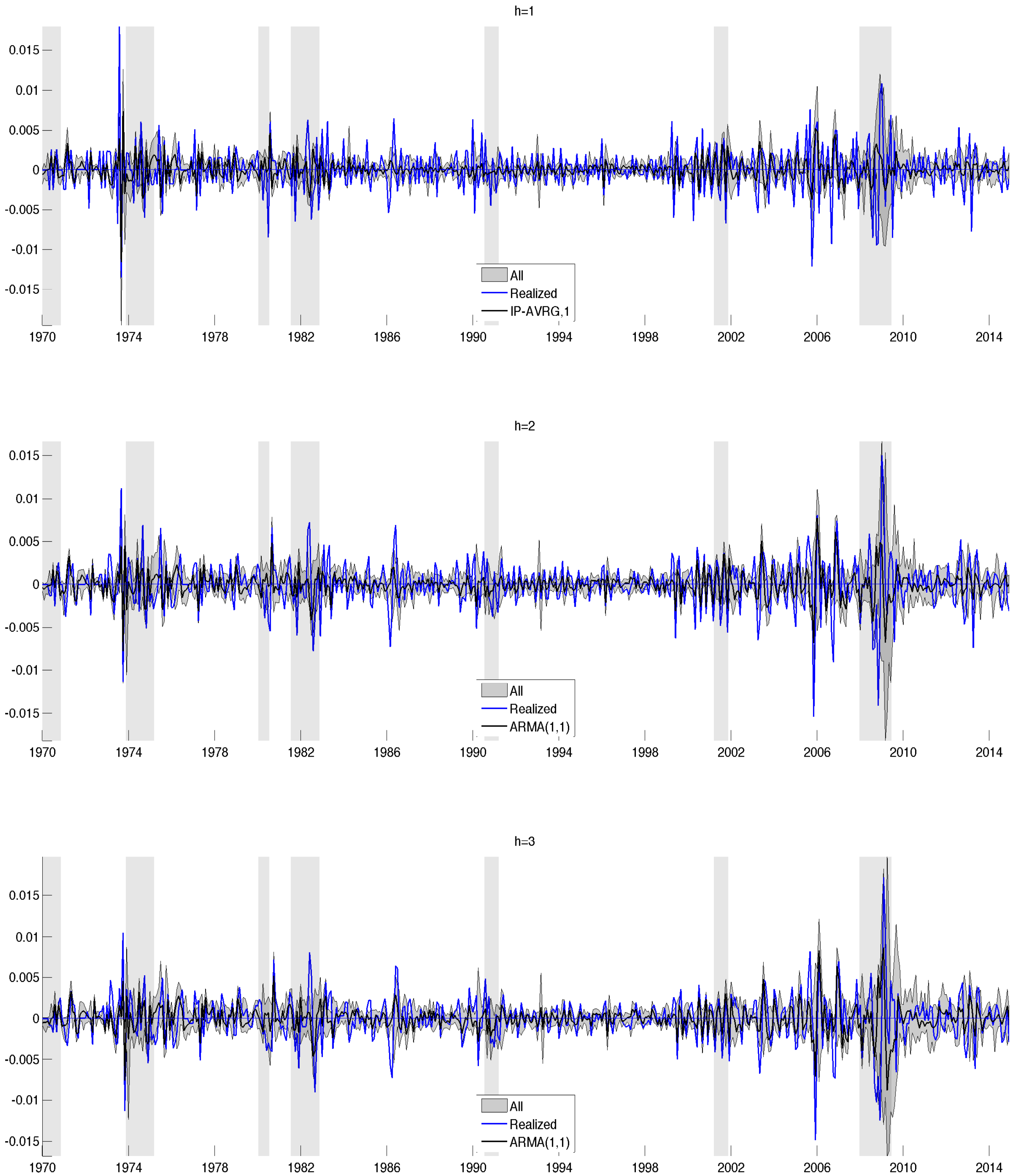
*Note: Minimum values are in bold, while ***, **, * stand for 1%, 5% and 10% significance of Diebold-Mariano test.*

Table 6: Relative MSPE and MAPE across business cycles for CPI

	Full Out-of-Sample		NBER Recessions		NBER Non-Recessions	
	RMSPE	RMAPE	RMSPE	RMAPE	RMSPE	RMAPE
h=1	IP-AVRG,1	ARMA(1,1)	Median	ARDI-soft	ARMA(1,1)	ARMA(1,1)
	0.888***	0.940**	0.772**	0.879**	0.867***	0.927***
h=2	ARMA(1,1)	ARMA(1,1)	ARMA(1,1)	ARDI-soft	ARMA(1,1)	ARMA(1,1)
	0.780**	0.911***	0.725**	0.895**	0.823***	0.915***
h=3	ARMA(1,1)	ARMA(1,1)	ARMA(1,1)	ARMA(1,1)	ARMA(1,1)	ARMA(1,1)
	0.798**	0.936**	0.725**	0.894**	0.864*	0.954
h=4	ARMA(1,1)	ARMA(1,1)	ARMA(1,1)	ARDI	ARMA(1,1)	ARMA(1,1)
	0.841**	0.940**	0.946	0.979	0.777**	0.922**
h=6	ARMA(1,1)	IP-AVRG,0.95	ADL	ADL	ARMA(1,1)	IP-AVRG,0.95
	0.873	0.992	0.765*	0.921*	0.799	0.966
h=8	ARMA(1,1)	IP-AVRG,0.95	ARDI-hard,1.65	ARDI-hard,1.65	ARMA(1,1)	IP-AVRG,0.95
	0.869	0.963*	0.689*	0.908	0.833	0.968
h=12	ARDI-DU	IP-AVRG,0.95	ARDI-hard,1.65	ARDI-hard,1.65	IP-AVRG,0.95	IP-AVRG,0.95
	0.832**	0.934**	0.640**	0.811**	0.846**	0.945*

*Note: This table shows the minimum relative (to ARD model) MSPE and MAPE values as well as the corresponding models for the full out-of-sample and during recessions and non-recession periods during recessions as declared by the NBER. ***, **, * stand for 1%, 5% and 10% significance of Diebold-Mariano test.*

Figure 7: Out-of-sample forecasts: CPI



The figure shows the pseudo-out-of-sample forecasts of the CPI annualized monthly growth rate for horizons 1, 2, and 3 months. The bleu line presents the historical data and the black line the forecast of the best MSPE model. The gray area around these lines presents the forecasts of all models considered in this exercise.

Table 7: Relative MSPE for Core CPI

	h=1	h=2	h=3	h=4	h=6	h=8	h=12
Standard Time Series Models							
ARI	1.000	1.038	1.146*	1.180**	1.248**	1.258**	1.459**
ARMA(1,1)	**0.869****	**0.771****	**0.812****	**0.818****	**0.802***	**0.827***	0.989
ADL	1.047	1.152	1.156	1.156	1.184	1.087	1.133
Factor-Augmented Regressions							
ARDI	0.976	1.064	1.168	1.231	1.160	1.062	**0.911***
ARDI-soft	0.910	0.994	1.134*	1.116	1.136	1.300	1.070
ARDI-hard,1.28	0.972	1.055	1.077	1.135	1.201	1.077	0.944
ARDI-hard,1.65	1.023	1.112***	1.165	1.094	1.186	1.128	1.031
ARDI-tstat,1.96	1.003	0.992	1.065	1.001	1.006	1.016	0.960
ARDI-DU	0.988	1.172	1.223	1.301	1.428	1.100	0.950
Factor-Structure-Based Models							
FAVARI	1.131**	1.219**	1.501**	1.706**	1.993**	2.150**	2.821**
FAVARD	1.105**	1.213**	1.446**	1.691**	2.007**	2.301**	3.249**
FAVARMA-FMA	1.097*	1.230**	1.482**	1.666**	2.013**	2.687*	2.817**
FAVARMA-FAR	1.338***	1.920***	2.921***	3.926***	6.112***	8.631***	14.611***
DFM	0.956	0.981	1.087	1.144	1.300*	1.367*	1.676*
Other Data-Rich Models							
3PRF	1.086**	1.202***	1.291**	1.297***	1.371***	1.266***	1.358***
CSR,1	1.031	1.067	1.187**	1.260**	1.390**	1.383**	1.537**
CSR,10	0.996	1.075	1.194*	1.256**	1.369**	1.324**	1.424**
CSR,20	0.988	1.069	1.204*	1.286**	1.548**	1.319**	1.397**
Forecasts Combinations							
AVRG	0.892**	0.918*	0.963	0.962	1.016	1.052	1.160
Median	0.884**	0.919*	0.990	0.966	0.944	0.933	0.985
T-AVRG	0.894**	0.915*	0.963	0.948	0.978	0.987	1.044
IP-AVRG,1	0.892**	0.916*	0.954	0.942	0.977	0.982	1.005
IP-AVRG,0.95	0.894**	0.920*	0.948	0.920**	0.940	0.945	0.943

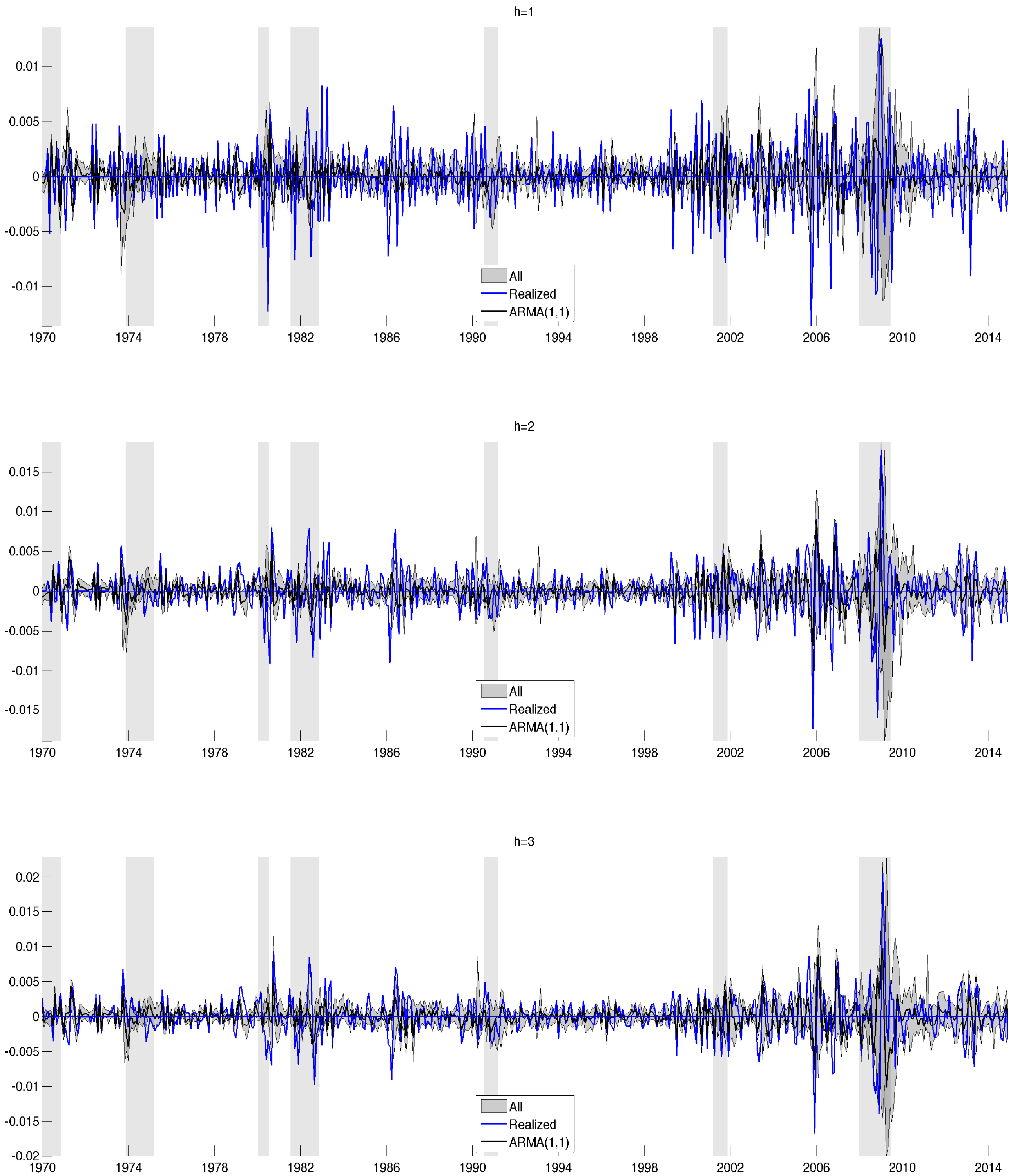
*Note: Minimum values are in bold, while ***, **, * stand for 1%, 5% and 10% significance of Diebold-Mariano test.*

Table 8: Relative MSPE and MAPE across business cycles for Core CPI

	Full Out-of-Sample		NBER Recessions		NBER Non-Recessions	
	RMSPE	RMAPE	RMSPE	RMAPE	RMSPE	RMAPE
h=1	ARMA(1,1)	ARMA(1,1)	ARDI-soft	ARDI-soft	ARMA(1,1)	ARMA(1,1)
	0.869**	0.936**	0.738*	0.864**	0.835***	0.918***
h=2	ARMA(1,1)	ARMA(1,1)	ARMA(1,1)	ARMA(1,1)	ARMA(1,1)	ARMA(1,1)
	0.771**	0.914**	0.705*	0.886	0.835***	0.930**
h=3	ARMA(1,1)	ARMA(1,1)	ARMA(1,1)	ARMA(1,1)	ARMA(1,1)	ARMA(1,1)
	0.812**	0.929**	0.840*	0.957	0.796**	0.923**
h=4	ARMA(1,1)	ARMA(1,1)	ARMA(1,1)	ARDI-tstat,1.96	ARMA(1,1)	ARMA(1,1)
	0.818**	0.934**	0.919	0.921**	0.744*	0.920**
h=6	ARMA(1,1)	ARMA(1,1)	ADL	ADL	ARMA(1,1)	ARMA(1,1)
	0.802*	0.966	0.753*	0.867**	0.701*	0.950
h=8	ARMA(1,1)	ARMA(1,1)	ARDI-hard,1.65	ADL	ARMA(1,1)	ARMA(1,1)
	0.827*	0.965	0.736*	0.887**	0.767*	0.962
h=12	ARDI	ARDI	ARDI-hard,1.65	ARDI-hard,1.65	ARDI	ARDI
	0.911*	0.983	0.783*	0.907	0.956	1.005

*Note: This table shows the minimum relative (to ARD model) MSPE and MAPE values as well as the corresponding models for the full out-of-sample and during recessions and non-recession periods during recessions as declared by the NBER. ***, **, * stand for 1%, 5% and 10% significance of Diebold-Mariano test.*

Figure 8: Out-of-sample forecasts: Core CPI

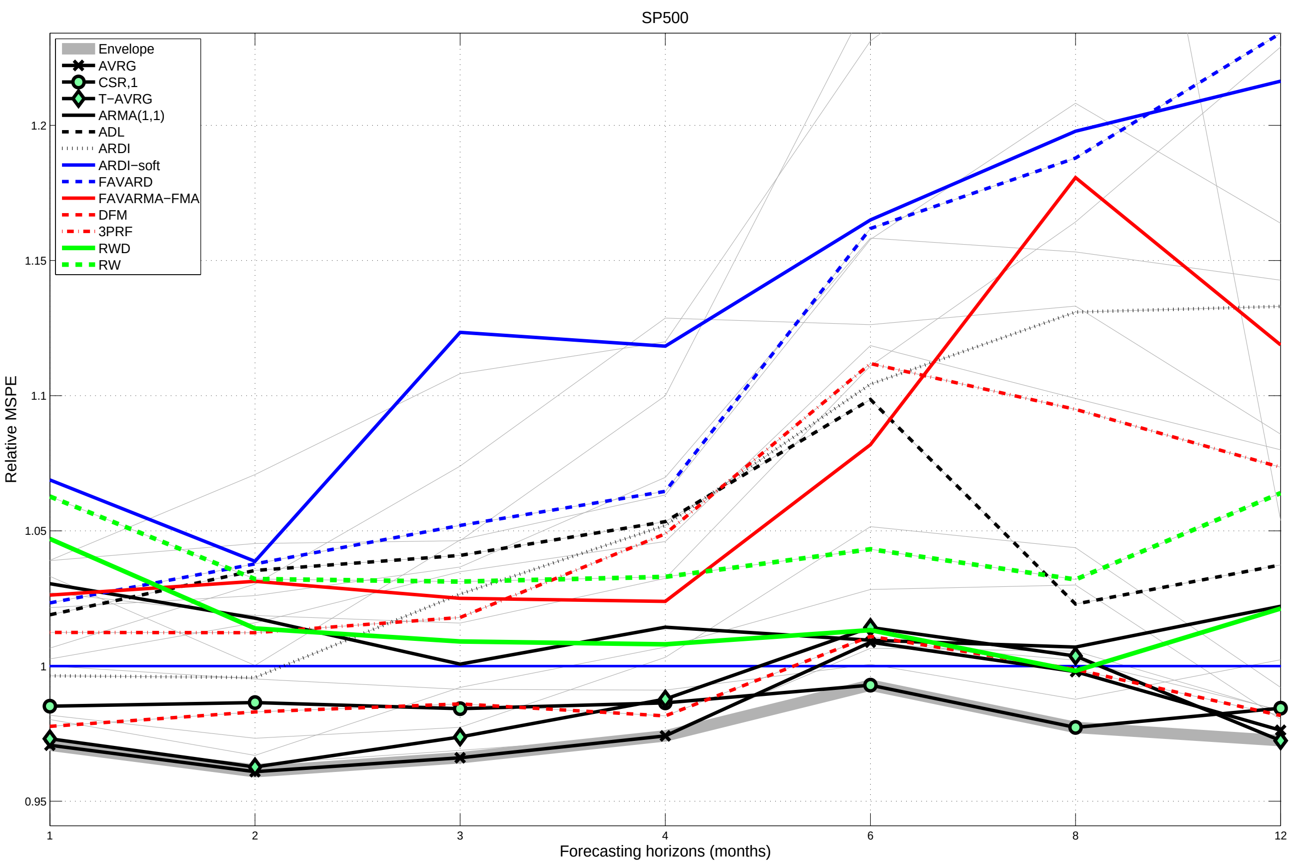


The figure shows the pseudo-out-of-sample forecasts of the Core CPI annualized monthly growth rate for horizons 1, 2, and 3 months. The bleu line presents the historical data and the black line the forecast of the best MSPE model. The gray area around these lines presents the forecasts of all models considered in this exercise.

5.2.3 Stock Market

We now examine the empirical results for the SP500 returns. As usual, Figure (9) shows the RMSPE and Table (9) displays the significance levels of the DM test statistics. Given the perpetual debate on the efficiency of stock markets, we have highlighted the performance of the RW and RWD models. According to the results, these two models are dominated by the ARD benchmark. This immediately suggests that stock returns are predictable to some extent.

Figure 9: Relative MSPE: SP500



The figure shows the MSPE of all models relative to ARD. When the value is below the blue line the corresponding model produces smaller MSPE than ARD. The thick gray line shows the inferior envelope, i.e. the lowest RMSPE for each forecasting horizon.

At horizons between $h = 1$ and $h = 4$, the best performing models are the AVRG and T-AVRG. The CSR-1 dominates at $h = 5$ and beyond. The latter forecasting model is the only one that consistently dominates the benchmark at all horizons. Except the DFM whose performance is comparable to that of the CSR-1 at horizons between $h = 1$ and $h = 4$, all other individual data-rich models are dominated by the ARD. It should be noted that the efficiency gain of the best performing models over the standard RW approaches 5% at all horizons. At the same time, the efficiency gain of the best performing models over the ARD benchmark lies below 4% at all horizons. This shows that stock market returns are much harder to forecast than any of the series considered previously.

Table 9: Relative MSPE for SP500

	h=1	h=2	h=3	h=4	h=6	h=8	h=12
Standard Time Series Models							
ARI	1.000	0.995	0.991*	0.991	1.001	0.988	1.002
ARMA(1,1)	1.030**	1.018*	1.001	1.014	1.010	1.007	1.022
ADL	1.019	1.035	1.041	1.053*	1.099*	1.023	1.037
Factor-Augmented Regressions							
ARDI	0.996	0.996	1.027	1.052	1.104	1.131	1.133
ARDI-soft	1.069*	1.039	1.123*	1.118	1.165	1.198*	1.216*
ARDI-hard,1.28	1.022	1.026	1.037	1.070	1.158	1.153	1.143*
ARDI-hard,1.65	1.033	1.000	1.047	1.063	1.158	1.208	1.164*
ARDI-tstat,1.96	1.007	1.030	1.074	1.129*	1.126	1.133	1.086
ARDI-DU	1.003	1.016	1.035	1.046	1.119	1.099	1.080
Factor-Structure-Based Models							
FAVARI	1.027	1.019	1.016	1.032	1.111	1.164	1.229
FAVARD	1.023	1.038	1.052	1.065	1.162	1.188	1.234
FAVARMA-FMA	1.026	1.031	1.025	1.024	1.082	1.181	1.119
FAVARMA-FAR	1.039	1.071	1.108*	1.120*	1.231*	1.287*	1.353*
DFM	0.978	0.983	0.986	0.982	1.011	0.999	0.982
Other Data-Rich Models							
3PRF	1.012	1.012	1.018	1.049	1.112	1.095	1.074
CSR,1	0.985***	0.986**	0.984**	0.986	**0.993**	**0.977**	0.984
CSR,10	0.982	0.973	0.977	1.003	1.052	1.044	0.992
CSR,20	1.039	1.045	1.046	1.100	1.248*	1.450*	1.054
Forecasts Combinations							
AVRG	**0.971**	**0.961**	**0.966**	**0.974**	1.009	0.998	0.976
Median	0.980	0.967	0.992	1.007	1.028	1.030	0.980
T-AVRG	0.973	0.963	0.974	0.988	1.014	1.004	**0.972**
IP-AVRG,1	0.971	0.962	0.967	0.975	1.009	1.002	0.984
IP-AVRG,0.95	0.972	0.964	0.969	0.974	1.007	1.005	0.983

*Note: Minimum values are in bold, while ***, **, * stand for 1%, 5% and 10% significance of Diebold-Mariano test.*

Table (10) shows that the CSR approach dominates during non-recession periods while other data-rich models (ARDI-hard, 3PFR, FAVARMA-FMA) do quite well during recession episodes. This suggests that data-rich models (other than the CSR) capture bearish signals better than the benchmark model. Figure (10) explains why the efficiency gain of our best models over the RW is small. In fact, our best forecasts are much less volatile than the SP500 returns in general in addition to being too optimistic during recession periods. Given our findings at Table (10), we can infer that the most pessimistic forecasts during recessions are delivered by factor-based models and these forecasts are closer to the realized data than the average forecast.

It may be of interest to compare the forecasting methods in terms of their ability to correctly match the signs of the target series. Indeed, a forecasting model that is outperformed by the RW according to the MSPE or MAPE can still have significant predictive power for the sign of the target, see (Satchell & Timmermann 1995). This possibility can be assessed by means

Table 10: Relative MSPE and MAPE across business cycles for SP500

	Full Out-of-Sample		NBER Recessions		NBER Non-Recessions	
	RMSPE	RMAPE	RMSPE	RMAPE	RMSPE	RMAPE
h=1	AVRG	CSR,10	ARDI-tstat,1.96	ARDI-tstat,1.96	CSR,10	CSR,10
	0.971	0.974**	0.940	0.952*	0.978	0.969***
h=2	AVRG	T-AVRG	ARDI-hard,1.65	ARDI-tstat,1.96	CSR,1	T-AVRG
	0.961	0.967**	0.892	0.910*	0.991*	0.975*
h=3	AVRG	AVRG	IP-AVRG,1	ARDI-hard,1.65	CSR,1	AVRG
	0.966	0.980	0.916	0.936	0.988**	0.993
h=4	AVRG	AVRG	DFM	DFM	CSR,1	CSR,1
	0.974	0.981	0.903	0.936	0.990	0.990**
h=6	CSR,1	ARI	3PRF	3PRF	CSR,1	ARI
	0.993	0.999	0.877*	0.918*	0.995	0.999
h=8	CSR,1	AVRG	3PRF	FAVARMA-FMA	CSR,1	CSR,1
	0.977	0.986	0.807**	0.882*	0.974	0.988
h=12	T-AVRG	DFM	FAVARMA-FMA	FAVARMA-FMA	CSR,1	CSR,1
	0.972	0.976	0.736**	0.823**	0.987	0.990

*Note: This table shows the minimum relative (to ARD model) MSPE and MAPE values as well as the corresponding models for the full out-of-sample and during recessions and non-recession periods during recessions as declared by the NBER. ***, **, * stand for 1%, 5% and 10% significance of Diebold-Mariano test.*

of the (Pesaran & Timmermann 1992) sign forecast test. The test statistic is given by:

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

where $\hat{p}$ is the sample proportion of correctly signed forecasts (or the success ratio) 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 or MAPE. Under the null hypothesis that the signs of the forecasts are independent of the signs of the target, we have $S_n \longrightarrow \mathrm{N}(0,1)$. [11]

Table (11) reports the success ratios with the standard significance levels. At short horizons, most models have significant predictive power for the sign of the SP500 return, with a maximum of 64% at $h = 1$ for the CSR and forecast combinations, and 66% at $h = 2$ for the ARDIT. At longer horizons, the factor-structure-based models correctly predict up to 72% of the signs of the SP500 return. Interestingly, the performance of the models seems to improve with the forecast horizon. The null hypothesis of independence between the forecast and realized values is rejected most of the time for ADL, data-rich models and forecast combinations, but not for univariate autoregressive models. Therefore, adding external information improves the prediction of the sign of the SP500 returns.

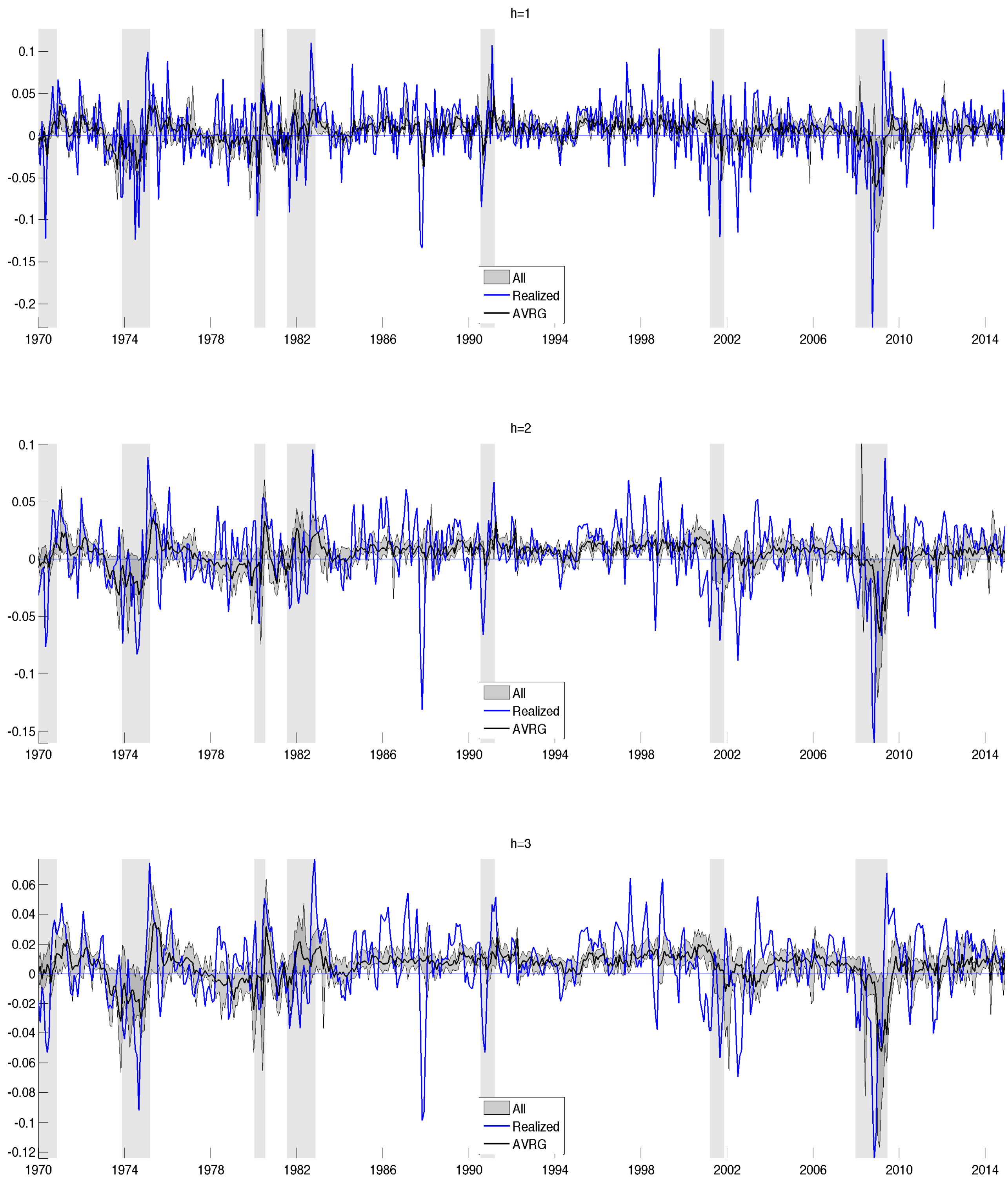
[11] Let q denote the proportion of positive realizations in the actual data and $\widehat{q}$ the proportion of positive forecasts. Under H_0, the estimated theoretical number of correctly signed forecast is $\hat{p}^* = q\widehat{q} + (1-q)(1-\widehat{q})$.

Table 11: Success ratio for SP500

	h=1	h=2	h=3	h=4	h=6	h=8	h=12
Standard Time Series Models							
ARD	0.593	0.619	0.630	0.635*	0.661*	0.643	0.680
ARI	0.593*	0.619*	0.626	0.633*	0.646	0.646	0.663
ARMA(1,1)	0.593**	0.583	0.639**	0.617	0.644	0.637	0.654**
ADL	0.606***	0.598**	0.611**	0.637***	0.663***	0.656**	0.661**
Factor-Augmented Regressions							
ARDI	0.624***	0.657***	0.648***	0.665***	0.678***	0.681***	0.689***
ARDI-soft	0.630***	0.652***	0.654***	0.669***	0.667***	0.676***	0.661***
ARDI-hard,1.28	0.630***	**0.663*** **	0.652***	0.657***	0.659***	0.669***	0.674***
ARDI-hard,1.65	0.613***	0.652***	0.650***	0.661***	0.667***	0.654***	0.657*
ARDI-tstat,1.96	0.604***	0.652***	0.631***	0.626***	0.661***	0.656***	0.670***
ARDI-DU	0.613***	0.654***	0.637***	0.654***	0.674***	0.672***	0.698***
Factor-Structure-Based Models							
FAVARI	0.630***	0.654***	**0.669*** **	**0.687*** **	0.707***	0.698***	0.700***
FAVARD	0.631***	0.637***	0.652***	0.676***	0.685***	0.698***	0.709***
FAVARMA-FMA	0.630***	0.648***	**0.669*** **	0.678***	**0.711*** **	**0.711*** **	0.706***
FAVARMA-FAR	0.600***	0.628***	0.635***	0.659***	0.674***	0.681***	0.672***
DFM	0.624***	0.644***	0.643***	0.670***	0.700***	0.696***	**0.720*** **
Other Data-Rich Models							
3PRF	0.620***	0.641***	0.661***	0.672***	0.700***	0.687***	0.680***
CSR,1	0.613***	0.628**	0.639**	0.657***	0.669**	0.667	0.693
CSR,10	**0.639*** **	0.650***	0.665***	0.674***	0.681***	0.674***	0.681***
CSR,20	0.620***	0.656***	0.650***	0.663***	0.670***	0.674***	0.650*
Forecasts Combinations							
AVRG	**0.639*** **	0.656***	0.661***	0.670***	0.693***	0.685***	0.689**
Median	0.622***	0.661***	0.659***	0.676***	0.683***	0.678***	0.696***
T-AVRG	0.635***	0.661***	0.661***	0.670***	0.693***	0.683***	0.694***
IP-AVRG,1	**0.639*** **	0.657***	0.663***	0.674***	0.694***	0.685***	0.691**
IP-AVRG,0.95	**0.639*** **	0.657***	0.663***	0.676***	0.696***	0.680***	0.693***

*Note: This table shows the success ratio with the (Pesaran & Timmermann 1992) sign forecast test significance where ***, **, * stand for 1%, 5% and 10% levels. Bold characters represent the maximum success ratio.*

Figure 10: Out-of-sample forecasts: SP500



The figure shows the pseudo-out-of-sample forecasts of the SP500 annualized monthly returns for horizons 1, 2, and 3 months. The bleu line presents the historical data and the black line the forecast of the best MSPE model. The gray area around these lines presents the forecasts of all models considered in this exercise.

5.2.4 Exchange Rates

(Rossi 2013) has reviewed the recent literature on exchange rates forecasting and concluded that some exchange rates may be predictable at short horizons. This conclusion is reached by comparing some models to the driftless RW using the MSPE criterion. This subsection provides further evidence on the predictability of exchange rates by evaluating the performance of a large number of recent data-rich models in an extensive out-of-sample experiment.

EXUSUK We first examine the results of the EXUSUK annualized growth. Figure (12) shows the RMSPE of all models while Table (12) shows the formal DM tests. First of all, we note that the exchange rate growth is difficult to predict pointwise. The efficiency gains of the best performing models over the standard RW are maximized at around 12% at $h = 1$ and decreases fast with h to become negative beyond $h = 4$. Indeed, the ARMA(1,1) is the best model at horizons between $h = 1$ and $h = 4$ and while the standard RW is the best at horizons $h = 5$ and beyond. The CSR-1 does as well as the ARD benchmark at all horizons. Clearly, data-rich methods are of no help when it comes to predicting the US-UK exchange rate growth.

Table (13) shows that the superiority of the ARMA(1,1) comes from its good performance during non-recession periods. The best data-rich models are just as good as the ARD model during recession periods. Figure (13) shows the same pattern as in the case of the SP500 returns, namely that the forecast is too smooth in general and more optimistic than the actual series during recessions.

Although most of the models struggle with the point prediction of exchange rate growth, Table (14) shows that they have good predictive power for the sign of the target. Indeed, some models delivered up to 63% of correctly signed forecast for the US-UK exchange rate growth at the horizon $h = 1$ and up to 58% at the horizon $h = 2$. DFM, CSR, ADL and ARMA exhibit all achieve approximately 55% of correctly signed forecasts for exchange rate growth at horizons $h = 3$, 4 and 6 months. The proportion of correctly signs forecasts drops below 50% at longer horizons. Our results are in line with those of (Satchell & Timmermann 1995), who found that nonlinear models produce a larger proportion of correctly signed exchange rates forecasts than a RW model despite the fact that their MSPE is higher.

EXUSCA Figure (12) shows the RMSPEs for the Canada-US exchange rate and Table (15) shows the DM tests at different significant levels. The main difference with the US-UK exchange rate is that the standard RW dominates all models as soon as the forecast horizon exceeds $h = 1$. The efficiency gain of the RW over the benchmark increases with the forecast horizon and approaches 11% for $h = 12$. ARI and ARMA(1,1) slightly dominate the ARD for horizons between $h = 2$ and $h = 8$.

Table (16) shows that they are fewer models that significantly outperform the benchmark during recession periods than during no recession periods. Indeed, the only model that beats the ARD during the recession is the ARMA(1,1) model and this occurs at horizons $h = 2$ and $h = 3$. Figure (14) confirms the well-known stylized fact that the volatility of exchange rates is hard to replicate.

While the RW model dominates in predicting the point values of US-CA exchange rate growth, Table (17) shows that several data-rich models and forecast combinations have predictive power for the sign of the EXUSCA growth at all horizons. The ARMA(1,1) model delivers 57% of correctly signed predictions at horizon $h = 1$. The ARDI-tstat is the only model with significant predictive power at the horizon $h = 2$. The AVRG is correctly signed 55% (resp. 56%) at horizon $h = 3$ (resp. $h = 8$). The ARDIT model with hard threshold predicts correctly the sign of EXUSCA returns 58% of time at horizons 6 and 12 months ahead. Interestingly, most data-rich models and all forecasts combinations have significant predictive power at 1-year horizon despite the fact that they are all outperformed by the RW in terms of the MSPE and MAPE. Overall, the aggregation of forecasts is a robust strategy as it permits to better forecast the sign of the target than individual models at most horizons.

Figure 11: Relative MSPE: EXUSUK

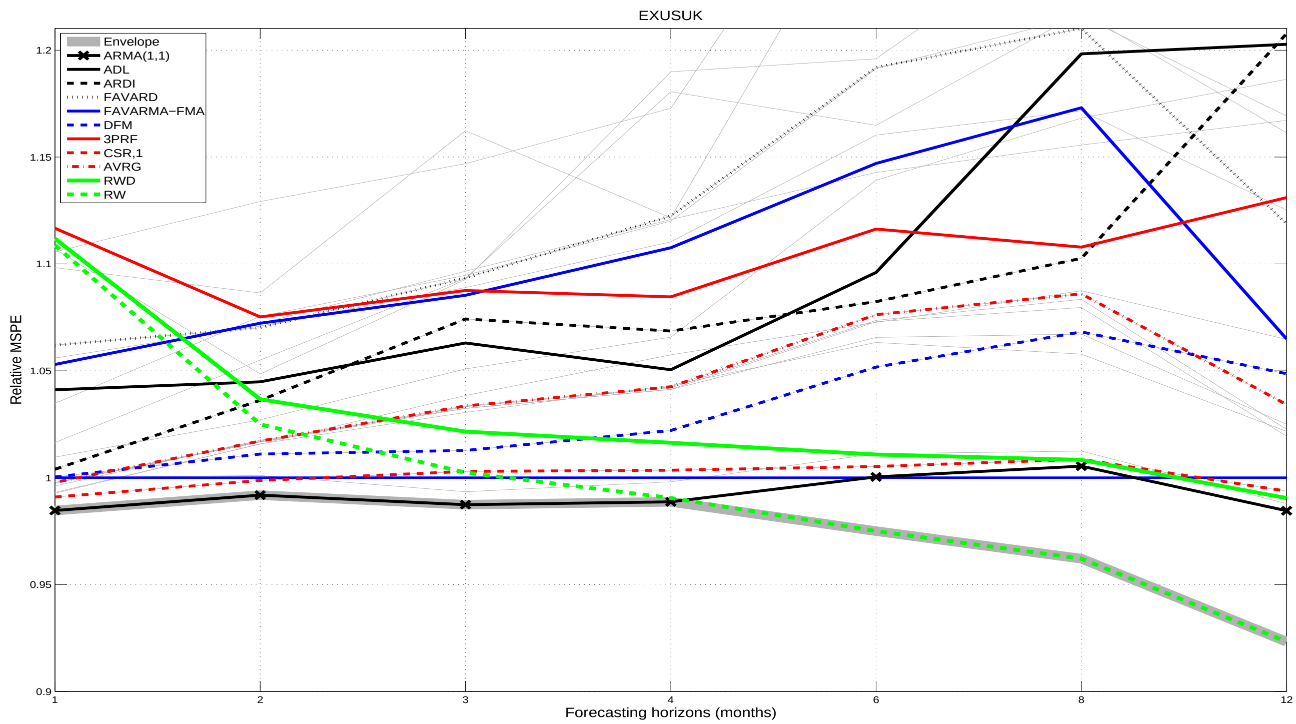
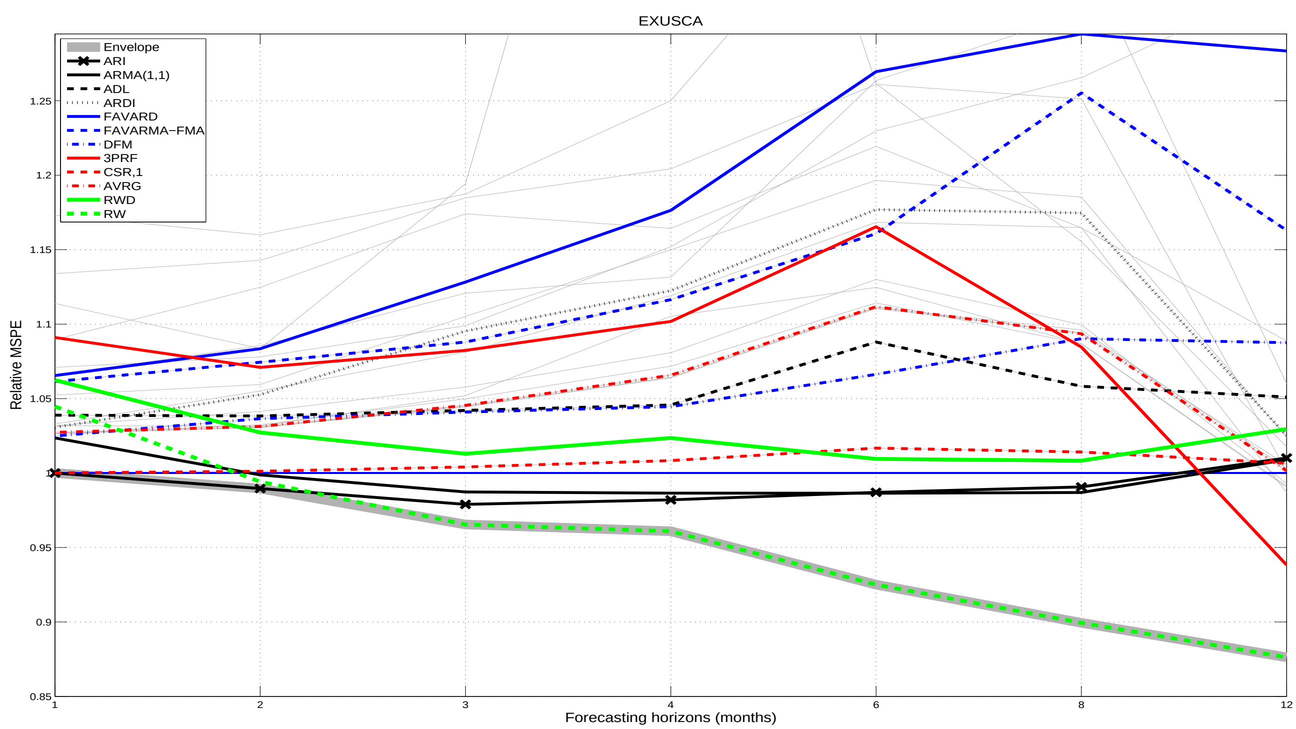


Figure 12: Relative MSPE: EXUSCA



The figure shows the MSPE of all models relative to ARD. When the value is below the blue line the corresponding model produces smaller MSPE than ARD. The thick gray line shows the inferior envelope, i.e. the lowest RMSPE for each horizon.

Table 12: Relative MSPE for EXUSUK

	h=1	h=2	h=3	h=4	h=6	h=8	h=12
Standard Time Series Models							
ARI	1.000	1.001	0.993	0.998	1.011*	1.012	0.988
ARMA(1,1)	**0.985**	**0.992**	**0.987**	**0.989***	**1.000**	**1.005**	**0.985**
ADL	1.041*	1.045*	1.063*	1.051	1.096**	1.198**	1.203***
Factor-Augmented Regressions							
ARDI	1.004	1.036	1.074*	1.069*	1.082**	1.103***	1.208***
ARDI-soft	1.107**	1.049*	1.093**	1.190**	1.196***	1.264***	1.328***
ARDI-hard,1.28	1.098***	1.086***	1.162***	1.122***	1.286***	1.365***	1.275***
ARDI-hard,1.65	1.106***	1.129***	1.147***	1.173***	1.315***	1.363***	1.324***
ARDI-tstat,1.96	1.053**	1.072***	1.097***	1.121***	1.143***	1.156***	1.167***
ARDI-DU	1.010	1.027	1.051**	1.066**	1.139***	1.168***	1.186***
Factor-Structure-Based Models							
FAVARI	1.056**	1.071***	1.089**	1.110**	1.160**	1.172**	1.125
FAVARD	1.062**	1.070**	1.093**	1.122**	1.192**	1.210**	1.119**
FAVARMA-FMA	1.053**	1.072***	1.085**	1.108**	1.147**	1.173**	1.065
FAVARMA-FAR	1.035*	1.075**	1.095**	1.120***	1.191***	1.216**	1.169**
DFM	1.000	1.011	1.013	1.022	1.052**	1.068**	1.049*
Other Data-Rich Models							
3PRF	1.117***	1.075**	1.088**	1.085**	1.116**	1.108**	1.131***
CSR,1	0.991*	0.999	1.003	1.003	1.005	1.009**	0.994
CSR,10	0.993	1.016	1.038**	1.057**	1.073**	1.088***	1.065*
CSR,20	1.016	1.055**	1.093***	1.181***	1.165***	1.218***	1.161***
Forecasts Combinations							
AVRG	0.998	1.017	1.034*	1.043*	1.076**	1.086**	1.034
Median	0.993	1.016	1.031	1.043*	1.063**	1.058**	1.022
T-AVRG	0.996	1.018	1.032	1.041*	1.066**	1.067**	1.025
IP-AVRG,1	0.997	1.017	1.032	1.041*	1.073**	1.080**	1.019
IP-AVRG,0.95	0.998	1.017	1.033*	1.042*	1.074**	1.083**	1.023

*Note: Minimum values are in bold, while ***, **, * stand for 1%, 5% and 10% significance of Diebold-Mariano test.*

Table 13: Relative MSPE and MAPE across business cycles for EXUSUK

	Full Out-of-Sample		NBER Recessions		NBER Non-Recessions	
	RMSPE	RMAPE	RMSPE	RMAPE	RMSPE	RMAPE
h=1	ARMA(1,1)	ARMA(1,1)	CSR,10	Median	ARMA(1,1)	ARMA(1,1)
	0.985	0.990*	0.972	0.991	0.985	0.990*
h=2	ARMA(1,1)	CSR,1	ARI	ARI	ARMA(1,1)	CSR,1
	0.992	0.998	0.950	0.984	0.999	0.999
h=3	ARMA(1,1)	ARMA(1,1)	ARI	DFM	ARMA(1,1)	ARMA(1,1)
	0.987	0.998	0.953	0.978	0.996	1.000
h=4	ARMA(1,1)	ARMA(1,1)	3PRF	Median	ARMA(1,1)	ARMA(1,1)
	0.989*	0.994*	0.982	0.982	0.988*	0.994*
h=6	ARMA(1,1)	CSR,1	ADL	3PRF	ARMA(1,1)	ARMA(1,1)
	1.000	1.002	0.994	0.959	0.998	1.001
h=8	ARMA(1,1)	CSR,1	CSR,1	Median	ARMA(1,1)	ARMA(1,1)
	1.005	1.005	1.006	0.988	1.004	1.004
h=12	ARMA(1,1)	ARMA(1,1)	CSR,20	CSR,20	ARMA(1,1)	ARMA(1,1)
	0.985	0.992	0.922	0.979	0.967	0.987

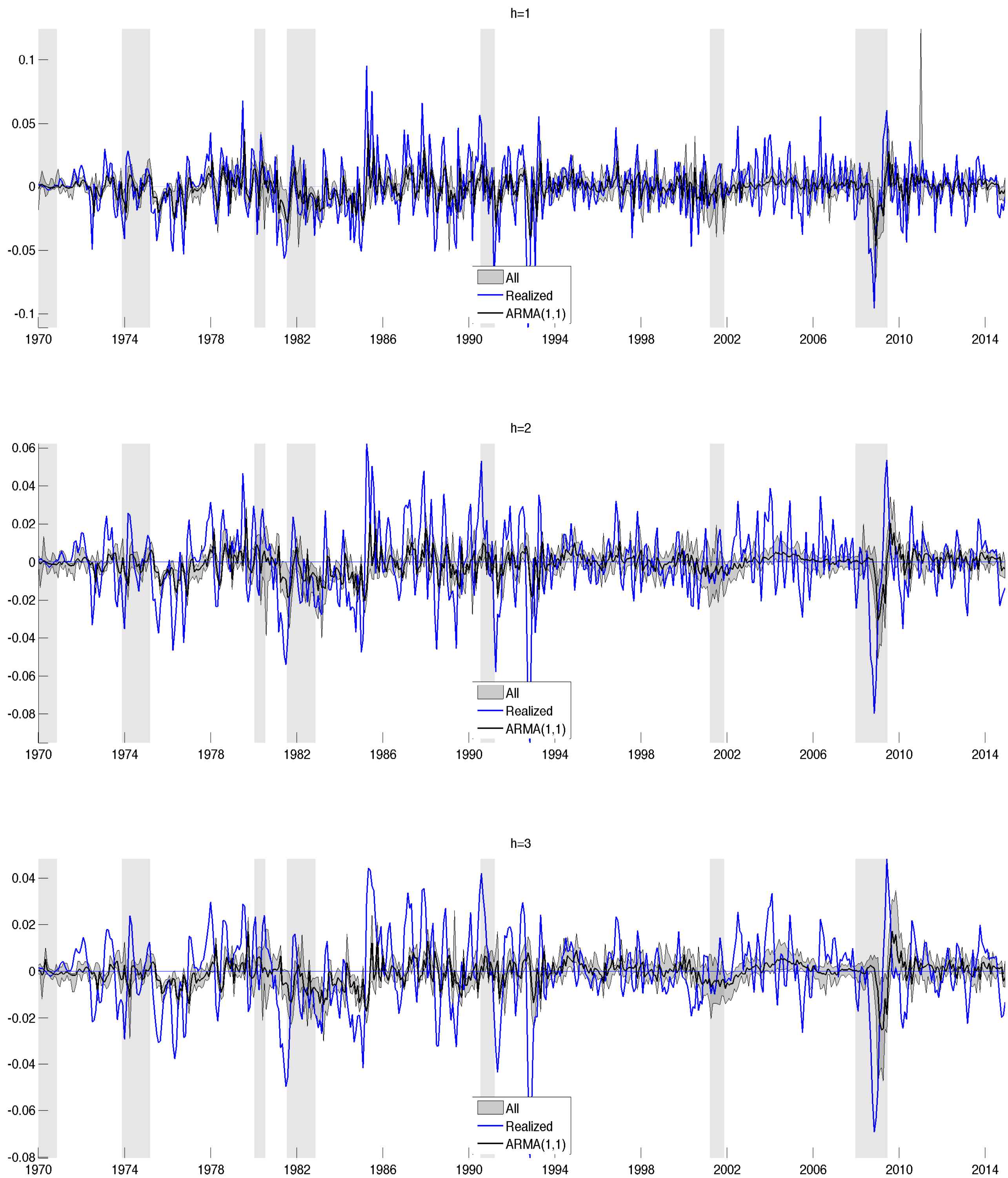
*Note: This table shows the minimum relative (to ARD model) MSPE and MAPE values as well as the corresponding models for the full out-of-sample and during recessions and non-recession periods during recessions as declared by the NBER. ***, **, * stand for 1%, 5% and 10% significance of Diebold-Mariano test.*

Table 14: Success ratio for EXUSUK

	h=1	h=2	h=3	h=4	h=6	h=8	h=12
Standard Time Series Models							
ARD	0.607	0.543**	0.535*	0.535*	0.491	0.472	0.480
ARI	0.607***	0.546**	0.552**	0.522	0.507	0.491	0.476
ARMA(1,1)	0.619***	0.561***	0.550**	0.533*	**0.530***	**0.504**	0.480
ADL	0.622***	0.559***	0.528	**0.554****	0.507	0.476	0.493
Factor-Augmented Regressions							
ARDI	0.628***	**0.581*****	0.537*	0.520	0.485	0.494	0.439***
ARDI-soft	0.615***	0.552**	0.531	0.493	0.470	0.489	0.483
ARDI-hard,1.28	0.596***	0.559***	0.548**	0.524	0.483	0.474	0.487
ARDI-hard,1.65	0.617***	0.541*	0.500	0.502	0.506	0.476	0.478
ARDI-tstat,1.96	0.593***	0.515	0.517	0.522	0.502	0.491	0.491
ARDI-DU	0.628***	0.578***	0.554**	0.504	0.483	0.483	0.461*
Factor-Structure-Based Models							
FAVARI	0.606***	0.559***	0.522	0.509	0.472	0.480	0.456**
FAVARD	0.598***	0.550**	0.517	0.493	0.467	0.456**	0.443***
FAVARMA-FMA	0.602***	0.543**	0.509	0.506	0.459*	0.474	0.459*
FAVARMA-FAR	0.585***	0.519	0.502	0.509	0.469	0.465	0.463*
DFM	0.619***	0.569***	**0.556*****	0.528	0.478	0.481	0.452**
Other Data-Rich Models							
3PRF	0.563***	0.519	0.509	0.507	0.470	0.480	0.448**
CSR,1	0.619***	0.550**	0.539*	0.546**	0.500	0.480	**0.494**
CSR,10	0.620***	0.544**	**0.556*****	0.531*	0.494	0.476	0.454**
CSR,20	0.591***	0.561***	0.554**	0.537*	0.487	0.483	0.454**
Forecasts Combinations							
AVRG	0.620***	0.557***	0.533	0.520	0.476	0.463	0.446**
Median	**0.633*****	0.552**	0.539*	0.519	0.478	0.469	0.441***
T-AVRG	0.630***	0.559***	0.541*	0.517	0.472	0.452**	0.446**
IP-AVRG,1	0.619***	0.554**	0.531	0.517	0.478	0.465	0.443***
IP-AVRG,0.95	0.619***	0.557***	0.528	0.519	0.476	0.459*	0.446**

*Note: This table shows the success ratio with the (Pesaran & Timmermann 1992) sign forecast test significance where ***, **, * stand for 1%, 5% and 10% levels.*

Figure 13: Out-of-sample forecasts: EXUSUK



The figure shows the pseudo-out-of-sample forecasts of the EXUSUK annualized monthly returns for horizons 1, 2, and 3 months. The bleu line presents the historical data and the black line the forecast of the best MSPE model. The gray area around these lines presents the forecasts of all models considered in this exercise.

Table 15: Relative MSPE for EXUSCA

	h=1	h=2	h=3	h=4	h=6	h=8	h=12
Standard Time Series Models							
ARI	**1.000**	**0.990**	**0.979****	**0.982****	0.987	0.991	1.010
ARMA(1,1)	1.024	0.999	0.987**	0.986*	**0.986**	**0.987**	1.009
ADL	1.039	1.038	1.042	1.046	1.088*	1.058*	1.051*
Factor-Augmented Regressions							
ARDI	1.031*	1.053	1.095*	1.123	1.177*	1.175*	1.025
ARDI-soft	1.090***	1.125**	1.174**	1.164*	1.219*	1.165**	1.089
ARDI-hard,1.28	1.134***	1.143***	1.185**	1.204*	1.261**	1.251**	1.001
ARDI-hard,1.65	1.173***	1.160***	1.187**	1.250**	1.409**	1.352**	1.060
ARDI-tstat,1.96	1.052**	1.059*	1.105**	1.150*	1.197*	1.185*	1.024
ARDI-DU	1.033*	1.055*	1.081	1.119*	1.168*	1.165*	0.987
Factor-Structure-Based Models							
FAVARI	1.071**	1.078*	1.099	1.152*	1.230*	1.265	1.330
FAVARD	1.065*	1.083*	1.128*	1.176*	1.270*	1.295*	1.283*
FAVARMA-FMA	1.061*	1.074*	1.088*	1.116*	1.161*	1.255	1.163
FAVARMA-FAR	1.066**	1.086**	1.121**	1.132**	1.264*	1.309*	1.370*
DFM	1.025	1.037	1.041	1.045	1.066	1.090	1.088
Other Data-Rich Models							
3PRF	1.091**	1.071*	1.082	1.102	1.165	1.085	**0.938****
CSR,1	1.000	1.001	1.004	1.008	1.017	1.014	1.007
CSR,10	1.032**	1.031	1.052*	1.105*	1.125*	1.087	0.990
CSR,20	1.114**	1.084***	1.194**	1.673*	1.262**	1.155**	1.018
Forecasts Combinations							
AVRG	1.027*	1.031	1.045	1.066	1.112	1.093	1.002
Median	1.032*	1.041	1.058	1.081	1.130	1.100	0.993
T-AVRG	1.026*	1.032	1.050	1.072	1.114	1.087	0.991
IP-AVRG,1	1.027*	1.031	1.044	1.064	1.110	1.094	1.005
IP-AVRG,0.95	1.026*	1.031	1.045	1.064	1.111	1.096	1.003

*Note: Minimum values are in bold, while ***, **, * stand for 1%, 5% and 10% significance of Diebold-Mariano test.*

Table 16: Relative MSPE and MAPE across business cycles for EXUSCA

	Full Out-of-Sample		NBER Recessions		NBER Non-Recessions	
	RMSPE	RMAPE	RMSPE	RMAPE	RMSPE	RMAPE
h=1	ARI	CSR,1	CSR,1	CSR,1	ARI	CSR,1
	1.000	0.998*	1.000	0.995	1.000	0.998
h=2	ARI	ARMA(1,1)	ARI	ARMA(1,1)	ARI	ARDI-tstat,1.96
	0.990	0.996	0.976	0.984***	0.995	0.996
h=3	ARI	ARI	ARI	ARMA(1,1)	ARI	ARI
	0.979**	0.990**	0.981	0.990*	0.978**	0.990**
h=4	ARI	ARI	ADL	ARI	ARI	ARI
	0.982**	0.992*	0.955	0.993	0.981*	0.991*
h=6	ARMA(1,1)	ARMA(1,1)	ADL	CSR,1	ARMA(1,1)	ARMA(1,1)
	0.986	0.988*	0.970	1.002	0.980	0.985*
h=8	ARMA(1,1)	ARMA(1,1)	FAVARMA-FMA	CSR,10	ARMA(1,1)	ARMA(1,1)
	0.987	0.987	0.915	0.979	0.983	0.981*
h=12	3PRF	3PRF	ARDI-soft	T-AVRG	3PRF	3PRF
	0.938**	0.940***	0.806	0.965	0.929**	0.928***

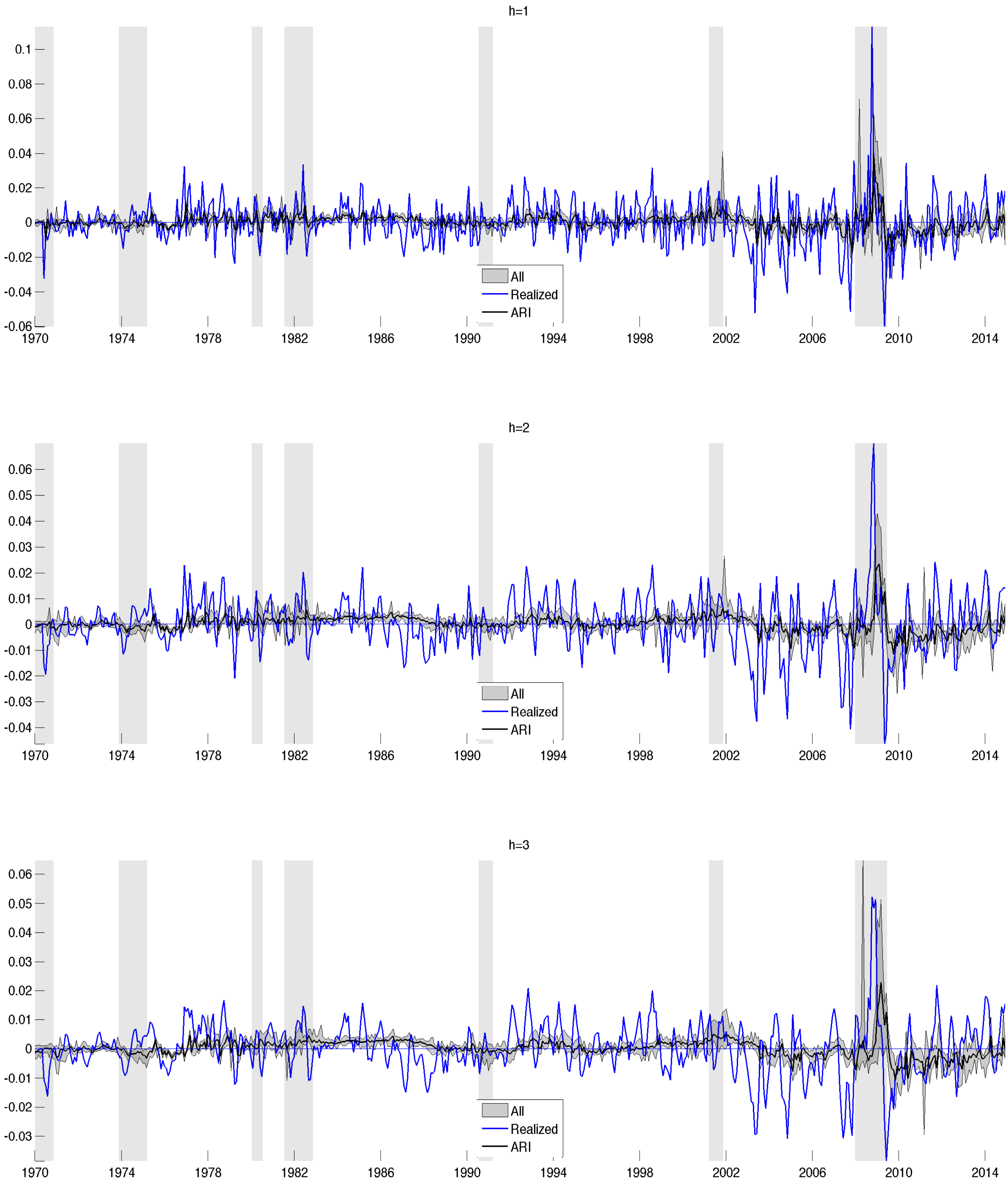
*Note: This table shows the minimum relative (to ARD model) MSPE and MAPE values as well as the corresponding models for the full out-of-sample and during recessions and non-recession periods during recessions as declared by the NBER. ***, **, * stand for 1%, 5% and 10% significance of Diebold-Mariano test.*

Table 17: Success ratio for EXUSCA

	h=1	h=2	h=3	h=4	h=6	h=8	h=12
Standard Time Series Models							
ARD	0.554	0.530	0.515	0.489	0.507	0.526	0.530
ARI	0.554**	0.530	0.526	0.517	0.515	0.522	0.531
ARMA(1,1)	**0.572*** **	0.535	0.533	0.511	0.517	0.526	0.541*
ADL	0.567***	0.526	0.528	0.509	0.487	0.496	0.507
Factor-Augmented Regressions							
ARDI	0.546**	0.528	0.537	0.524	0.491	0.496	0.533
ARDI-soft	0.533	0.524	0.519	0.506	0.519	0.519	0.559***
ARDI-hard,1.28	0.543**	0.522	0.544**	**0.531**	0.548**	0.511	0.557***
ARDI-hard,1.65	0.548**	0.519	0.530	0.530	**0.583*** **	0.517	**0.583*** **
ARDI-tstat,1.96	0.537*	**0.546** **	0.496	0.511	0.528	0.528	0.544**
ARDI-DU	0.541*	0.519	0.535	0.502	0.504	0.487	0.548**
Factor-Structure-Based Models							
FAVARI	0.535	0.531	0.524	0.504	0.511	0.515	0.557***
FAVARD	0.535*	0.526	0.515	0.489	0.513	0.528	0.570***
FAVARMA-FMA	0.552**	0.531	0.530	0.506	0.519	0.507	0.561***
FAVARMA-FAR	0.552**	0.511	0.535	0.504	0.507	0.513	0.563***
DFM	0.543**	0.522	0.513	0.498	0.506	0.504	0.552**
Other Data-Rich Models							
3PRF	0.543**	0.530	0.513	0.513	0.520	0.530	0.570***
CSR,1	0.556***	0.519	0.507	0.485	0.504	0.511	0.522
CSR,10	0.561***	0.519	0.519	0.515	0.539*	0.519	0.548**
CSR,20	0.550**	0.517	0.507	0.528	0.531	0.511	0.531
Forecasts Combinations							
AVRG	0.539*	0.511	**0.546** **	0.524	0.556**	**0.561*** **	0.570***
Median	0.537*	0.520	0.524	0.507	0.543*	0.541*	0.572***
T-AVRG	0.528	0.522	0.537*	0.524	0.552**	0.548**	0.561***
IP-AVRG,1	0.539*	0.507	0.543*	0.522	0.552**	0.552**	0.569***
IP-AVRG,0.95	0.537*	0.509	0.544**	0.524	0.550**	0.552**	0.563***

*Note: This table shows the success ratio with the (Pesaran & Timmermann 1992) sign forecast test significance where ***, **, * stand for 1%, 5% and 10% levels.*

Figure 14: Out-of-sample forecasts: EXUSCA



The figure shows the pseudo-out-of-sample forecasts of the EXUSCA annualized monthly returns for horizons 1, 2, and 3 months. The bleu line presents the historical data and the black line the forecast of the best MSPE model. The gray area around these lines presents the forecasts of all models considered in this exercise.

5.2.5 Summary

The Table (18) summarizes the main findings from this out-of-sample exercise and the implications for the prediction of US economic activity. We group the horizons into short-term (1-3 months), mid-term (4 and 6 months), and long-term (8 and 12 months). In addition, we report the recommended model during NBER recession periods and from the end of the Great Recession onward (since 2009M06).

Industrial Production According to MSPE, the industrial production growth rate is best predicted by the forecasts combinations: CSR-20 of (Elliott et al. 2013) and the IP-AVRG. However, the results are sensitive to the business cycle: the best forecasting models during recessions are the FAVARMA of (Dufour & Stevanovic 2013) and the ARDIT of (Bai & Ng 2008). The FAVARMA model is also the best since the end of the Great Recession. When it comes to the prediction of the sign of the industrial production change, the factor-structure-based models dominate (except at short horizons with CSR-20, and during recessions with ARDIT).

Employment growth The conclusions drawn previously are valid for the second real activity series as well. Overall, the CSR-20 and IP-AVRG are the models that minimize MSPE for all horizons. During recessions, ARDI and its variations perform best while the recommendation is not clear for the period following the Great Recession where CSR, DFM and ARDI emerge as winners. In the case of the sign forecast, IP-AVRG and autoregressive iterative model are the best at short horizon while ARDIT and ARDI-DU are winners at mid and long-term (and also during recessions). Since 2009M06, the models ARI, CSR and ARDI-DU share the throne.

Total and Core CPI The ARMA(1,1) model emerges as the best in terms of the MSPE criterion when it comes to predicting the CPI inflation growth. IP-AVRG is the second-best approach to predict inflation growth pointwise. When only the sign of the target matter, other models like ARDI-DU, ARDIT, ADL and forecast combinations are also useful as well, especially at longer horizons and during expansions. The results for the Core CPI inflation growth are very similar.

Stock market Few studies have found evidence of predictability of stock market returns, even in a data-rich context, see (Ludvigson & Ng 2005). Our results show that several data-rich models and forecast combinations are able to achieve lower MSPE and produce a higher proportion of correctly signed predictions for the SP500 returns than the RW. AVRG and CSR are the best models across all horizons according to distance-based criteria while factor-structure-based models emerge as better candidates in terms of correctly signed forecast. Since 2009M06 the winning model is the FAVARMA, followed by the ARMA(1,1).

Exchange rates As suggested by the literature, the RW is a very tough benchmark to beat when it comes to point-forecasting the exchange rate growth. However, the ARMA(1,1) model outperforms the RW at the short horizon while CSR, ARI and ARD have quite good performance during recessions and also since 2009M06. Data-rich models and the ARMA(1,1) become more relevant here when the signs of the predicted values are compared to the sign of actual exchange rates growth. The predictability is very low except for the short horizon and during recessions.

Table 18: Summary of forecasting performance

	Industrial Production			Employment		
	Distance-based criterion	Predictability	Direction-based criterion	Distance-based criterion	Predictability	Direction-based criterion
	MSPE	Pseudo-R^2	Sign forecast	MSPE	Pseudo-R^2	Sign forecast
Short term	CSR,20 / IP-AVRG	0.26-0.40	CSR,20 / FAVAR	CSR,20 / IP-AVRG	0.56-0.68	IP-AVRG / ARI
Mid term	CSR,20 / IP-AVRG	0.39	FAVAR(MA)	CSR,20 / IP-AVRG	0.65-0.60	ARDIT
Long term	CSR,20 / IP-AVRG	0.41	FAVAR	CSR,20 / IP-AVRG	0.57-0.47	ARDIT / ARDI-DU
Recessions	FAVARMA / ARDI-soft	0.5-0.63	FAVAR(MA) / ARDIT	ARDI*	0.40-0.82	ARDIT / ARDI-DU
Since Great Recession	FAVARMA	0.18-0.47	FAVARMA	CSR / DFM / ARDI	0.60-0.80	ARI / CSR / ARDI-DU
	CPI			Core CPI		
	Distance-based criterion	Predictability	Direction-based criterion	Distance-based criterion	Predictability	Direction-based criterion
	MSPE	Pseudo-R^2	Sign forecast	MSPE	Pseudo-R^2	Sign forecast
Short term	ARMA / IP-AVRG	0.20-0.45	ARMA / ARDI-DU	ARMA	0.19-0.41	AVRGs / ARDI-DU
Mid term	ARMA	0.48-0.55	ARDI-DU / ADL	ARMA	0.46-0.53	ARDI-DU / ADL
Long term	ARMA / IP-AVRG	0.57-0.60	AVRG / Median	ARMA/ARDI	0.53-0.54	ADL / ARDI-DU
Recessions	ARMA / ARDIT / ADL	0.07-0.66	ARDIT / ADL / AVRGs	ARMA / ARDIT / ADL	0.11-0.60	ARDIT / CSR / ADL
Since Great Recession	ARMA / ARDI	0.25-0.57	ARDI-DU / ARMA	ARMA / ARDI	0.22-0.61	ARMA / ARDI-DU
	SP500					
	Distance-based criterion	Predictability	Direction-based criterion			
	MSPE	Pseudo-R^2	Sign forecast			
Short term	AVRG / CSR	0.07-0.04	AVRGs / ARDIT			
Mid term	AVRG / CSR	0.03-0.02	FAVAR(MA)			
Long term	AVRG / CSR	0.04-0.05	FAVARMA / DFM			
Recessions	ARDI* / 3PRF / FAVARMA	0.08-0.31	CSR / ARDIT / FAVAR(MA)			
Since Great Recession	FAVARMA / ARMA	0.02-0.15	FAVAR(MA)			
	Exchange rate: US-UK			Exchange rate: US-CA		
	Distance-based criterion	Predictability	Direction-based criterion	Distance-based criterion	Predictability	Direction-based criterion
	MSPE	Pseudo-R^2	Sign forecast	MSPE	Pseudo-R^2	Sign forecast
Short term	ARMA	0.11-0.02	ARDI / Median	ARD / RW	0.06-0.05	ARMA / ARDI-tstat
Mid term	RW	0.03-0.04	ARMA / ADL	RW	0.06-0.08	ARDIT
Long term	RW	0.05-0.07		RW	0.11-0.15	AVRG / ARDIT
Recessions	CSR,10 / ARI	0.05-0.17	AVRG / ARDI	ARD / RW	0.08-0.18	ADL / ARDIT / AVRG*
Since Great Recession	ARI / RW	0.01-0.08	ARDIT / ARDI-DU	ARI / RW	0.05-0.17	CSR / ARDI*

Note: This table resumes the forecasting performance comparison. Under the Distance-based criterion the best models are chosen according to MSPE. The Direction-based criterion is the sign forecast test of (Pesaran & Timmermann 1992) where the best models have the highest significant success ratio. Predictability is measured by the pseudo-R^2 as in (Galbraith 2003). We report maximum values per horizon. Short-term consists of 1-3 months ahead; mid-term 4 and 6; long-term consists of 8 and 12 months horizons. Recessions stand for all NBER recession periods. Since 2009M06 *corresponds to the period since the end of the Great Recession. We report the best models for each category and across different horizons: 1-3 for the short term; 4 and 6 for the mid term; 8 and 12 for the long term; and all horizons combined for Recessions and since 2009M06.*

6 Miscellaneous

This horse race involving a large number of forecasting models in a very long out-of-sample period of 44 years produces a tremendous amount of secondary information such as the selection of models hyperparameters (lag polynomial orders, number of static and dynamic factors, selected predictors, etc.) as well as the distribution of forecasts for all series and horizons of interest. In this section we examine the stability of the forecasting equations over time and relate the out-of-sample forecast dispersion to macroeconomic uncertainty.

6.1 Stability of Forecasting Relationships

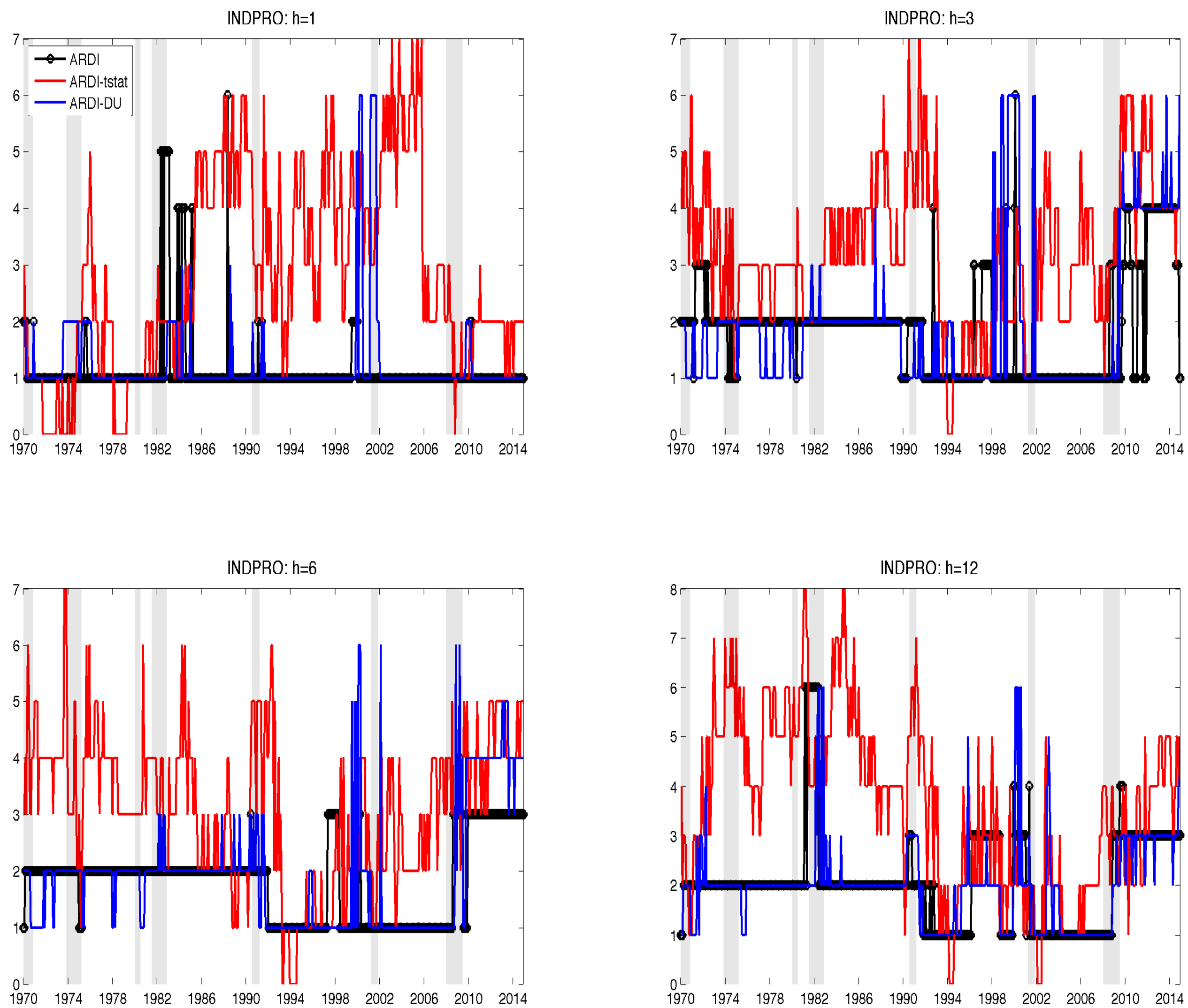
Recently, several studies have suggested that factor loadings and the number of factors are likely to change over time.[12]. Figure (15) shows the number of factors retained in ARDI, ARDI-tstat and ARDI-DU models when predicting the industrial production growth rate 1, 3, 6 and 12 months ahead. Recall that in the case of ARDI (ARDI-DU) models, the BIC selects the number of consecutive (generalized) principal components to be used , while the t-test selects a subset of principal components. Figure (16) plots the selected hyperparameters of the ARDI model (18) for all seven series of interest with 1-month horizon. The bottom right panel shows the estimated number of factors in FAVAR(MA) and DFM models using (Bai & Ng 2002) and (Hallin & Liska 2007) respectively.

The number of PCs used in ARDI-tstat and ARDI-DU models vary considerably across the out-of-sample period as well as for different forecasting horizons. The number of factors in ARDI model is more stable, but other hyperparameters of that model are quite unstable. From figure (16) we see that the instability is not the same for all series. Forecasting real activity measures require more factors (and their lags) than when predicting inflation and exchange rate growths. In the case of SP500 we note that more lags of F_t are systematically used since 1997. Finally, the number of static factors used in FAVAR(MA) models is much larger than the number of dynamic factors in the common component of DFM.

We now turn to the ARDIT model of (Bai & Ng 2008) to see whether the choice of pre-selected predictors and their nature are stable over time. Figure (17) plots the number of series selected by soft and hard thresholds when forecasting industrial production growth 1, 3, 6 and 12 months ahead. We note a lot of instability in the hard threshold model, with the number of selected series ranging from a minimum of 20 when predicting industrial production in early 70s to almost 80 series at the end of the sample. In addition, there are large swings between the 80s and the Great Recession. Figure (18) shows the type of series selected by ARDIT-hard

[12] See, among others, (Breitung & Eickmeier 2011), (D'Agostino, Gambetti & Giannone 2013), (Eickmeier, Lemke & Marcellino 2015), (Cheng, Liao & Schorfheide 2016), (Mao Takongmo & Stevanovic 2015), (Stevanovic 2016) and (Guerin, Leiva-Leon & Marcellino 2016).

Figure 15: Number of factors selected by BIC: Industrial Production

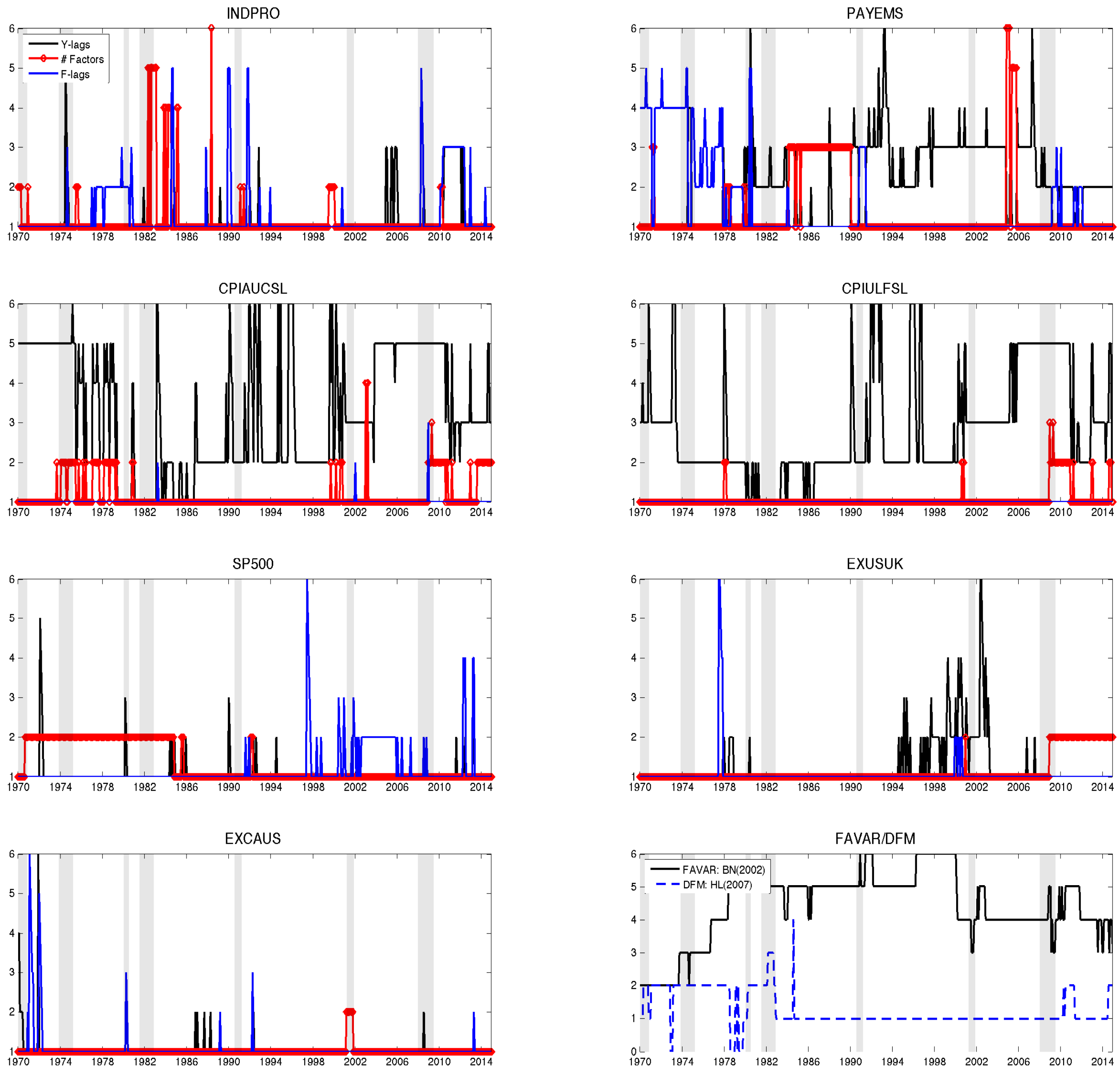


The figure shows the number of factors selected by the Bayesian Information Criterion (BIC) when predicting Industrial Production annualized monthly growth rate at horizons 1, 3, 6 and 12 months.

model with $t_c = 1.65$. We group the data as in (McCracken & Ng 2015) and calculate the proportion of selected series from each of those groups. This proportion is relatively stable in the case of industrial production, except that the stock prices start being important from late 90s for 1 and 2 months ahead and dominate the price series for longer horizons.

Figures (19) and (20) shows the same quantities but for all series and for 1-month horizon. The patterns are similar to real activity and price series, except that the number of candidate predictors is generally lower in the case of CPI and Core CPI inflation growths. In the case of stock prices the number of selected series is declining until the Great Recession while it remains relatively stable and small in the case of exchange rates. The proportions of variables per group

Figure 16: ARDI specification and # factors in FAVARs: All series



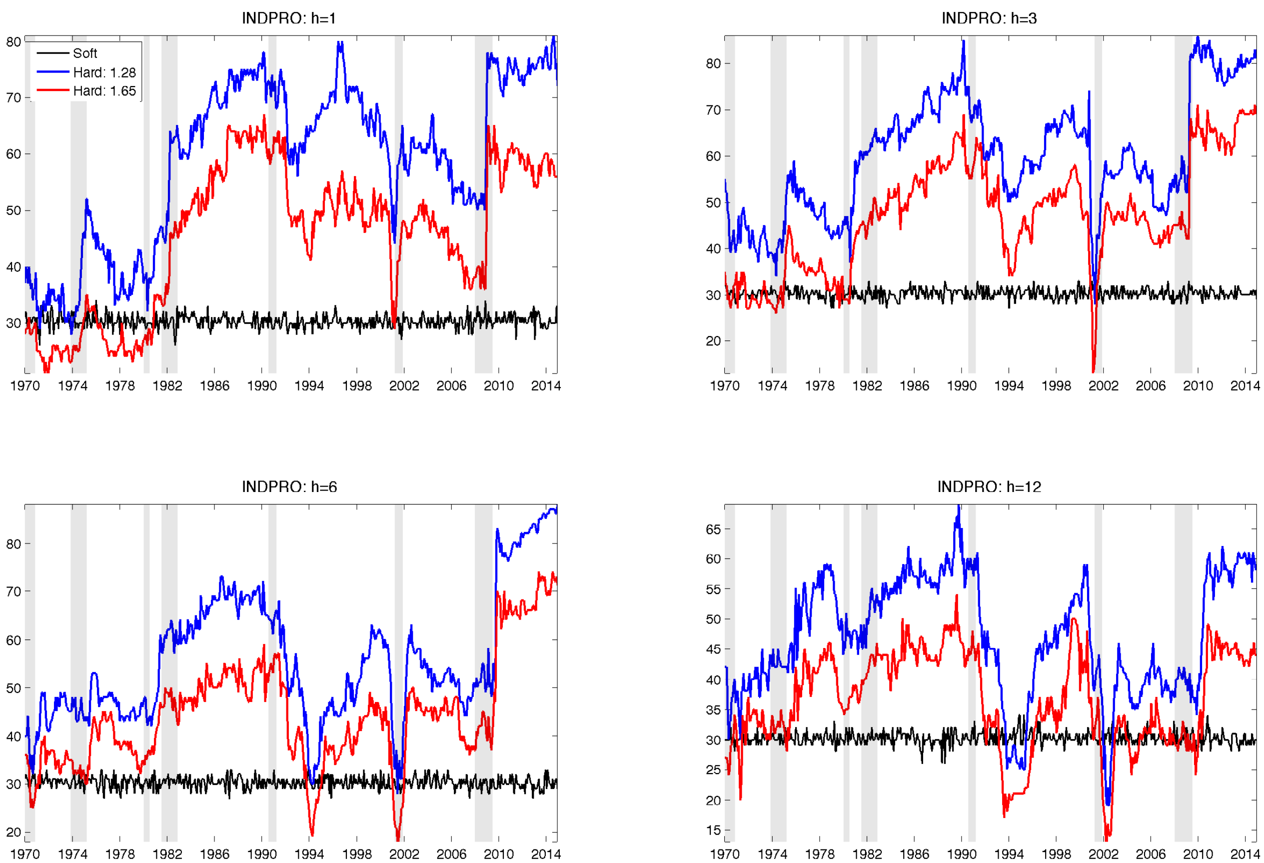
The figure shows the specification of ARDI model selected by BIC for all series at 1-month horizons. $Y - lags$ *stands for the estimated autoregressive lag order* p_y *in (18), while* $\#Factors$ *is the number of consecutive PCs kept in* F_t *and* $F - lags$ *is the lag order* p_f. *The bottom right panel plots the estimated number of static and dynamic factors by (Bai & Ng 2002) and (Hallin & Liska 2007) respectively.*

are quite different across the predicted series. They are very similar and stable in the case of industrial production and employment, but vary a lot when predicting the inflation growths.

Interestingly, there have been periods in the second half of 70s when almost no price series have been selected to forecast the inflation growth. This can be explained by the aggressive monetary policy during the Volcker period which shows up in the importance of interest rates group. The interest rates were important when predicting the stock returns in the first half of the out-of-sample period, but they almost disappeared since 2004, replaced by money and credit aggregates. Since the Great Recession, prices and real output and income series are clearly the most important.

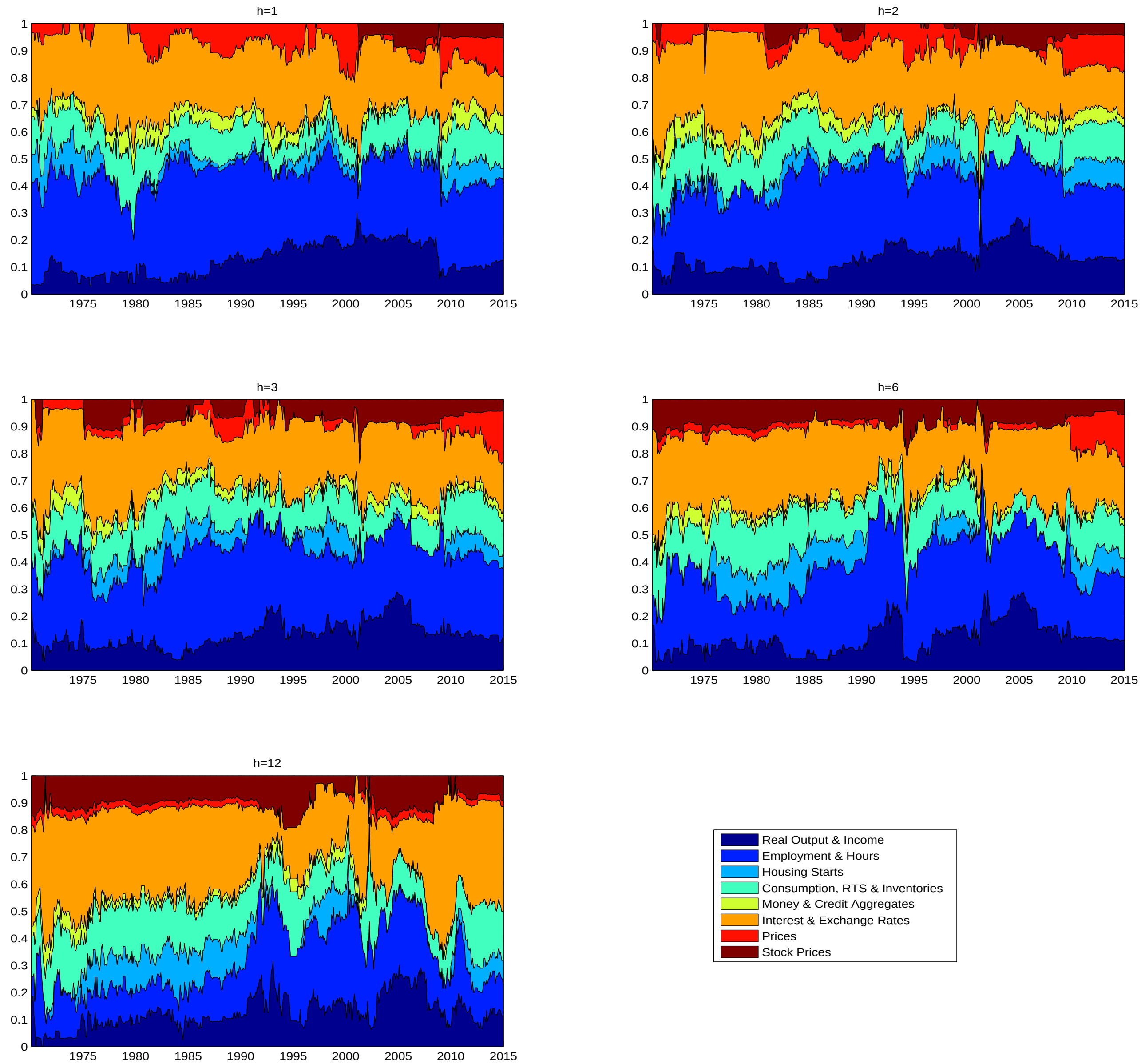
Overall, our very long out-of-sample period and the variety of forecasting models may serve as a good laboratory to study the stability of factor structures and the forecasting relationships. The results presented in this section document the prevalence of structural changes in all dimensions. However, the occurrence of these changes are not evenly distributed across the forecasted series and forecasting horizons.

Figure 17: Number of series selected by ARDIT-hard: Industrial Production



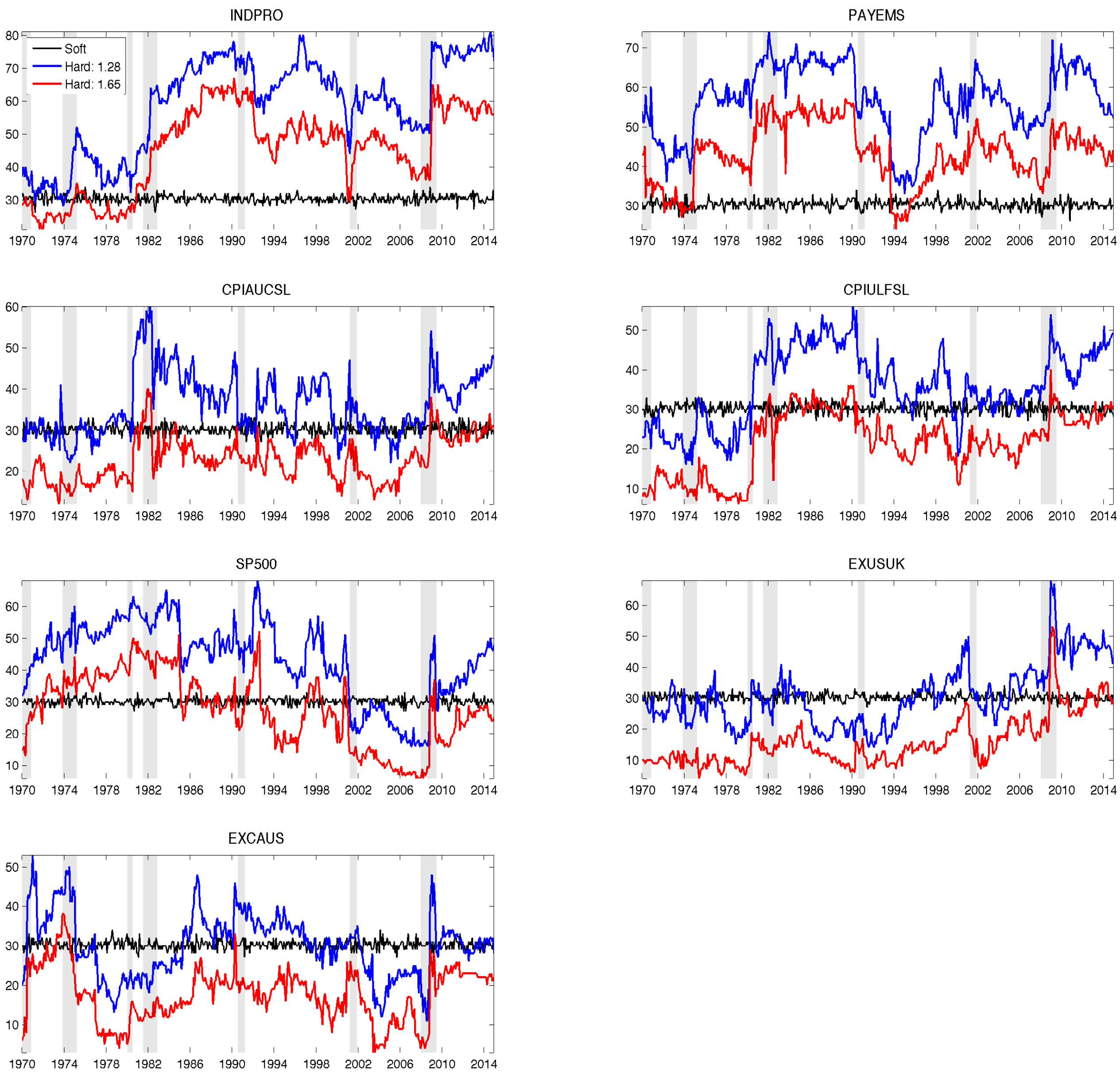
The figure shows the number of series selected by the hard and soft threshold ARDIT model when predicting Industrial Production annualized monthly growth rate at horizons 1, 2, 3, 6 and 12 months.

Figure 18: Type of series selected by ARDIT-hard: Industrial Production



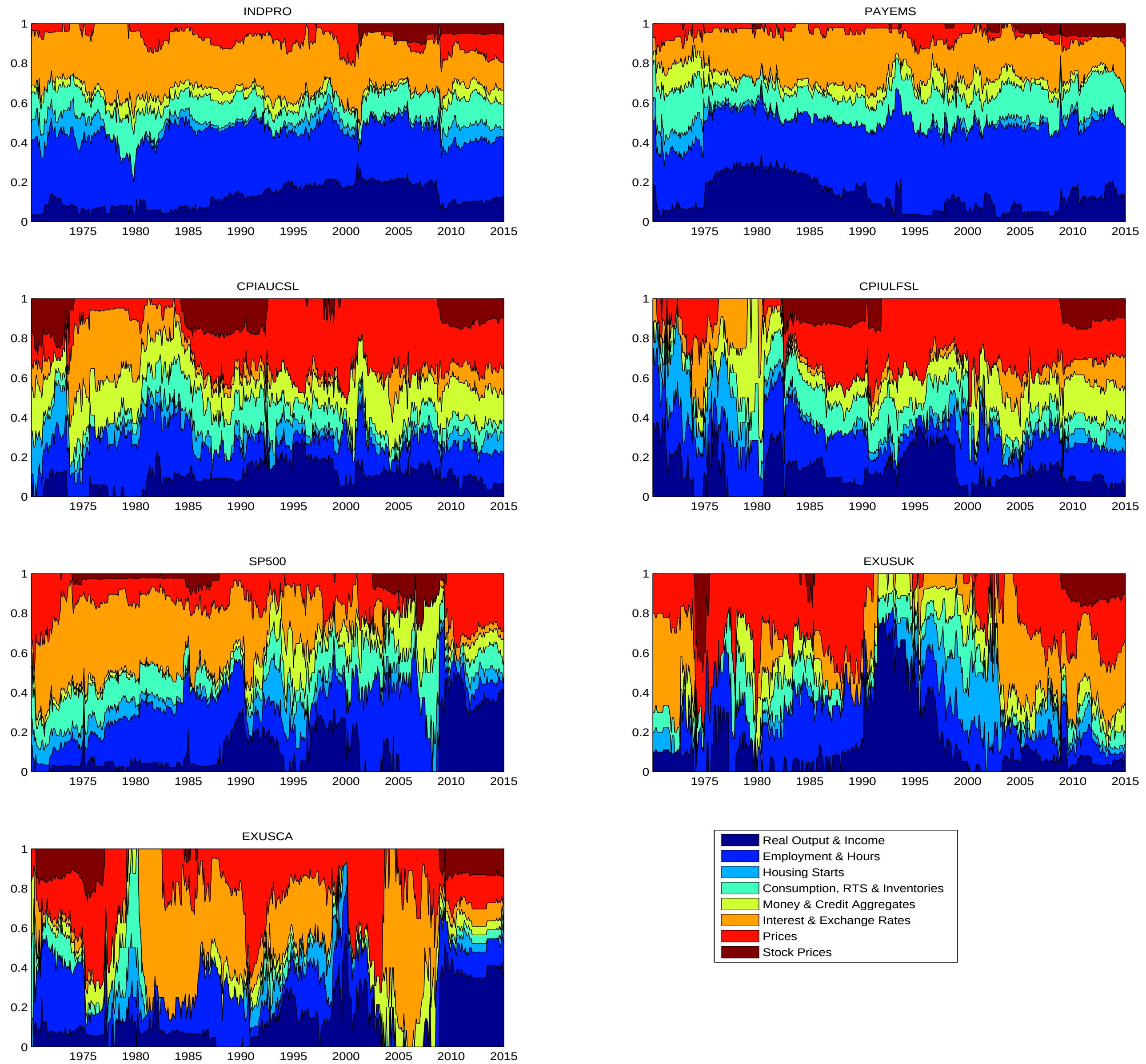
The figure shows the proportion of series selected by the hard threshold ARDIT model with $t_c = 1.65$ *when predicting Industrial Production annualized monthly growth rate at horizons 1, 2, 3, 6 and 12 months. The content of each group is described in (McCracken & Ng 2015).*

Figure 19: Number of series selected by ARDIT-hard: All series



The figure shows the number of series selected by the hard and soft threshold ARDIT model when predicting all series at 1-month horizon.

Figure 20: Type of series selected by ARDIT-hard: All series



The figure shows the proportion of series selected by the hard threshold ARDIT model with $t_c = 1.65$ *when predicting all series at horizon 1 month. The content of each group is described in (McCracken & Ng 2015).*

6.2 Forecasts Dispersion and Uncertainty

Since the seminal work by (Bloom 2009) there is a growing literature on the measurement of the macroeconomic uncertainty and its relationship with economic activity. For instance, (Bloom 2009) used the realized volatility of SP500 (and VIX) as a proxy for macroeconomic uncertainty while (Jurado, Ludvigson & Ng 2015) (JLN) measure it as the common stochastic volatility factor of forecasting errors estimated for more than a hundred of series. Another way of measuring the forecasting error volatility is to take the dispersion of individual forecasts for every out-of-sample period and every series for a particular forecasting horizon, see (Rossi, Sekhposyan & Soupre 2016).

Here we consider two measures of dispersion: the standard deviation (STD) and the interquartile range (IQR). Figure (21) plots the average (across seven series) of STD and IQR against the JLN macroeconomic uncertainty measure and SP500 realized volatility for horizons 1, 3 and 12 months ahead. We see that the out-of-sample forecast dispersion co-moves with the macroeconomic uncertainty during the business cycles irrespective of the forecasting horizon. It increases during NBER recessions, except for the 1991 recession, and the peak dispersion is observed in the middle of the 2007-09 recession. Compared to JLN our measures present higher peaks during recessions but the two are fairly correlated.

Figure (22) shows disaggregated forecasts dispersion for 1-month horizon as well as the first principal component of the STD and IQR measures of seven series. For instance, we see on the (1,1) Figure that the average dispersion (across series) and principal component of the dispersions look very similar. The (1,2) Figure shows the STD and IQR for industrial production against the JLN macroeconomic uncertainty measure and the SP500. λ^{STD} and λ^{IQR} represent the principal component loadings of the industrial production's dispersion. We see that the aggregate dispersion loads more on the dispersions of the CPI, SP500 and EXUSCA than on the dispersions of real activity variables. On the one hand, the forecast dispersion associated with employment growth is much higher until 1984 but stays low for the rest of the sample, even during the Great Recession. On the other hand, the dispersion in both CPI forecasts are quite larger since 2000.

Table (19) reports the proportion of variance of several uncertainty measures explained by our two aggregate dispersion measures and the time series-specific forecast dispersions. For instance, the aggregate STD dispersion explains 58% of variation in JLN macro uncertainty at 1-month horizon while SP500 realized volatility explains 22%. We also consider the VIX as well as the economic policy uncertainty (Policy) of (Baker, Bloom & Davis 2015). VIX is very correlated with SP500 realized volatility hence the R^2s are similar. In the case of Policy, the highest R^2 is 0.13.

Finally, we consider verifying whether the uncertainty measured by our out-of-sample forecast dispersion has a significant impact on the business cycle. We consider the 8 variables VAR from (Bloom 2009) and (Jurado et al. 2015) with the same recursive ordering but replacing their series of uncertainty by our aggregate STD dispersion. Figure (23) plots the impulse responses to the 100 basis points shock on forecast dispersion equation. This increase in the forecast dispersion generates a significant and persistent fall in employment and industrial production as well as in consumer prices. The federal funds rate decreases that can be interpreted as the systematic response of the central bank. Worked hours decline in the short term. These results are in line with the findings of (Bloom 2009) and (Jurado et al. 2015).

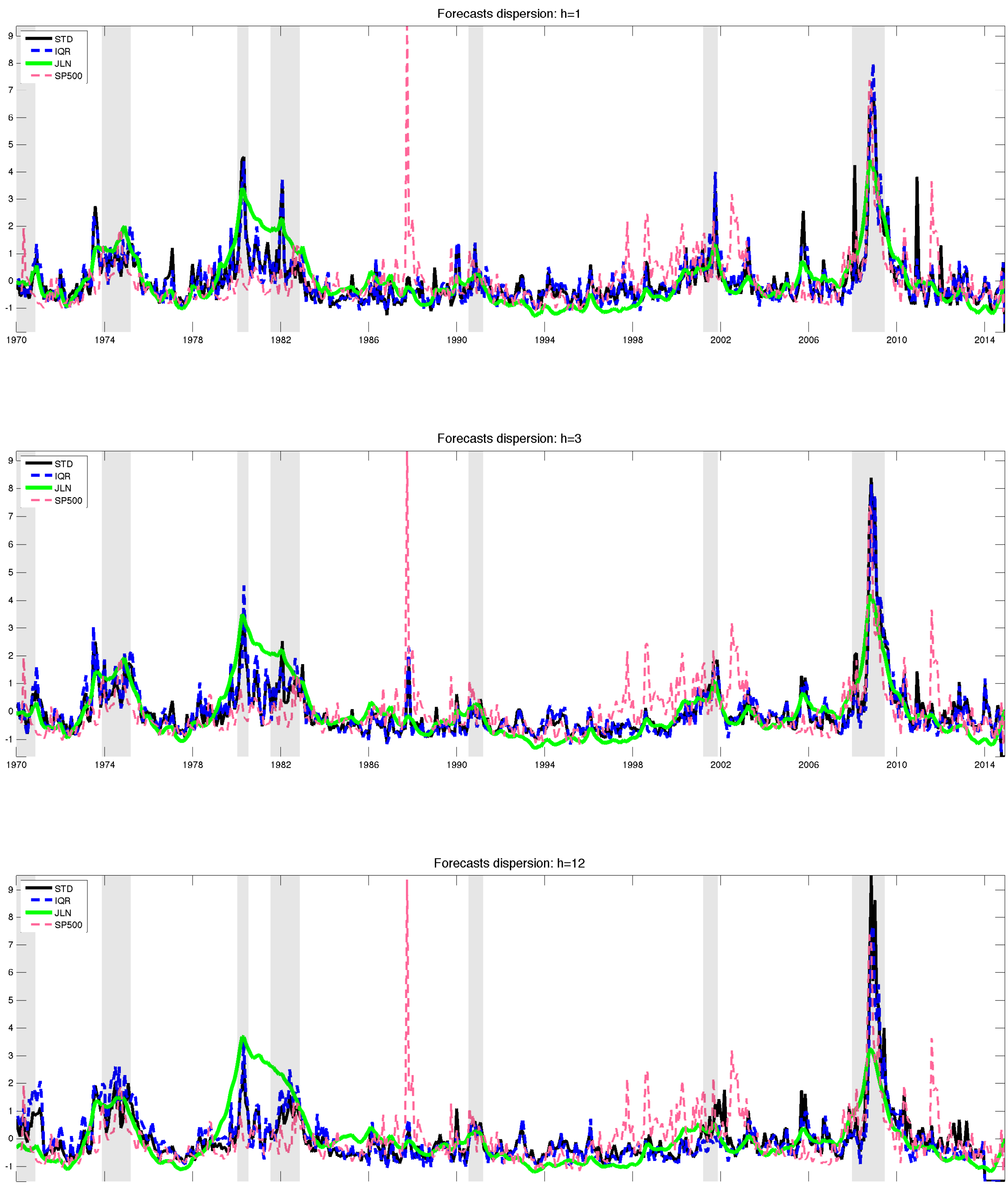
Overall, our out-of-sample forecast dispersion measures are good predictors of the macroeconomic and financial uncertainty measures used in the literature. Our results suggest that an unanticipated shock to forecast dispersion can generate business cycle movements among several real activity variables.

Table 19: Forecasts dispersions and measures of uncertainty

		$h=1$				$h=3$				$h=12$			
		Macro	SP500	VIX	Policy	Macro	SP500	VIX	Policy	Macro	SP500	VIX	Policy
All series	STD	0,58	0,22	0,29	0,09	0,60	0,28	0,28	0,13	0,54	0,27	0,27	0,13
	IQR	0,58	0,22	0,29	0,09	0,59	0,23	0,23	0,10	0,55	0,20	0,24	0,09
INDPRO	STD	0,41	0,10	0,21	0,06	0,47	0,11	0,27	0,12	0,52	0,09	0,25	0,11
	IQR	0,29	0,06	0,14	0,04	0,38	0,07	0,25	0,11	0,39	0,05	0,21	0,13
PAYEMS	STD	0,17	0,00	0,10	0,01	0,31	0,04	0,25	0,05	0,37	0,07	0,24	0,06
	IQR	0,16	0,00	0,08	0,01	0,26	0,02	0,18	0,05	0,30	0,04	0,24	0,06
CPI	STD	0,34	0,14	0,09	0,03	0,36	0,17	0,12	0,07	0,34	0,14	0,12	0,06
	IQR	0,21	0,10	0,04	0,00	0,30	0,15	0,08	0,04	0,35	0,16	0,11	0,05
CoreCPI	STD	0,28	0,15	0,10	0,04	0,29	0,18	0,11	0,08	0,25	0,16	0,12	0,07
	IQR	0,19	0,11	0,05	0,01	0,26	0,16	0,07	0,04	0,27	0,17	0,11	0,06
SP500	STD	0,50	0,15	0,30	0,10	0,51	0,21	0,30	0,13	0,43	0,22	0,29	0,15
	IQR	0,44	0,11	0,30	0,10	0,42	0,14	0,25	0,11	0,36	0,13	0,25	0,10
EXUSUK	STD	0,20	0,09	0,14	0,03	0,31	0,17	0,22	0,06	0,27	0,14	0,17	0,06
	IQR	0,20	0,14	0,18	0,04	0,26	0,11	0,15	0,02	0,22	0,08	0,12	0,02
EXUSCA	STD	0,25	0,20	0,19	0,08	0,19	0,19	0,17	0,12	0,23	0,29	0,25	0,12
	IQR	0,24	0,22	0,20	0,09	0,17	0,16	0,11	0,07	0,13	0,14	0,10	0,04

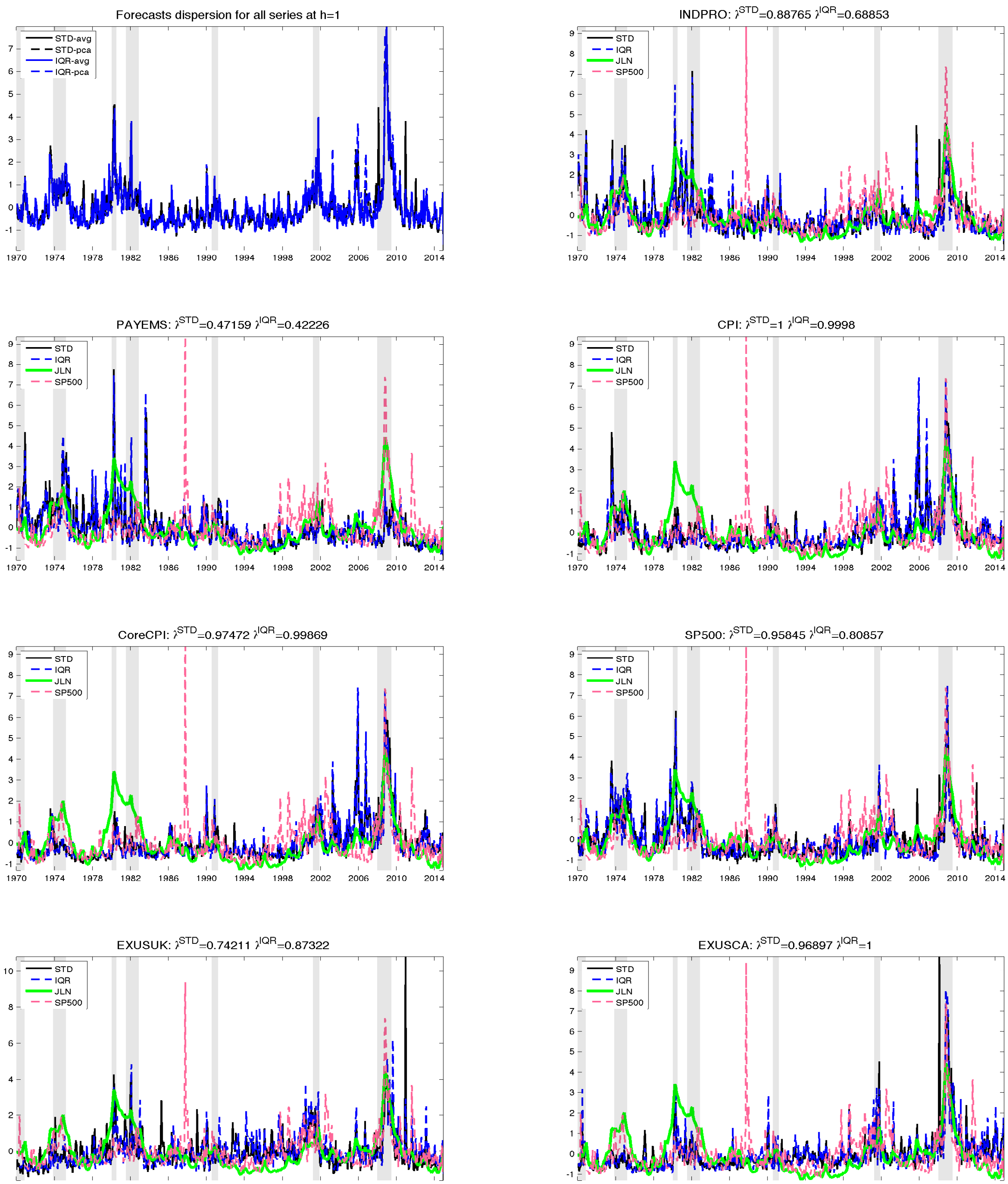
Note: This table shows the proportion of the variance (R^2) of uncertainty measures (columns) explained by the forecasts dispersion average measures STD and IQR for horizons 1, 3 and 12 months ahead. The uncertainty measures are: Macro *uncertainty from (Jurado et al. 2015), implied volatility of SP500 index options* VIX, SP500 *realized volatility (measured as a standard deviation of daily returns for each month) and economic* Policy *uncertainty from (Baker et al. 2015).*

Figure 21: Average forecasts dispersion and macroeconomic uncertainty



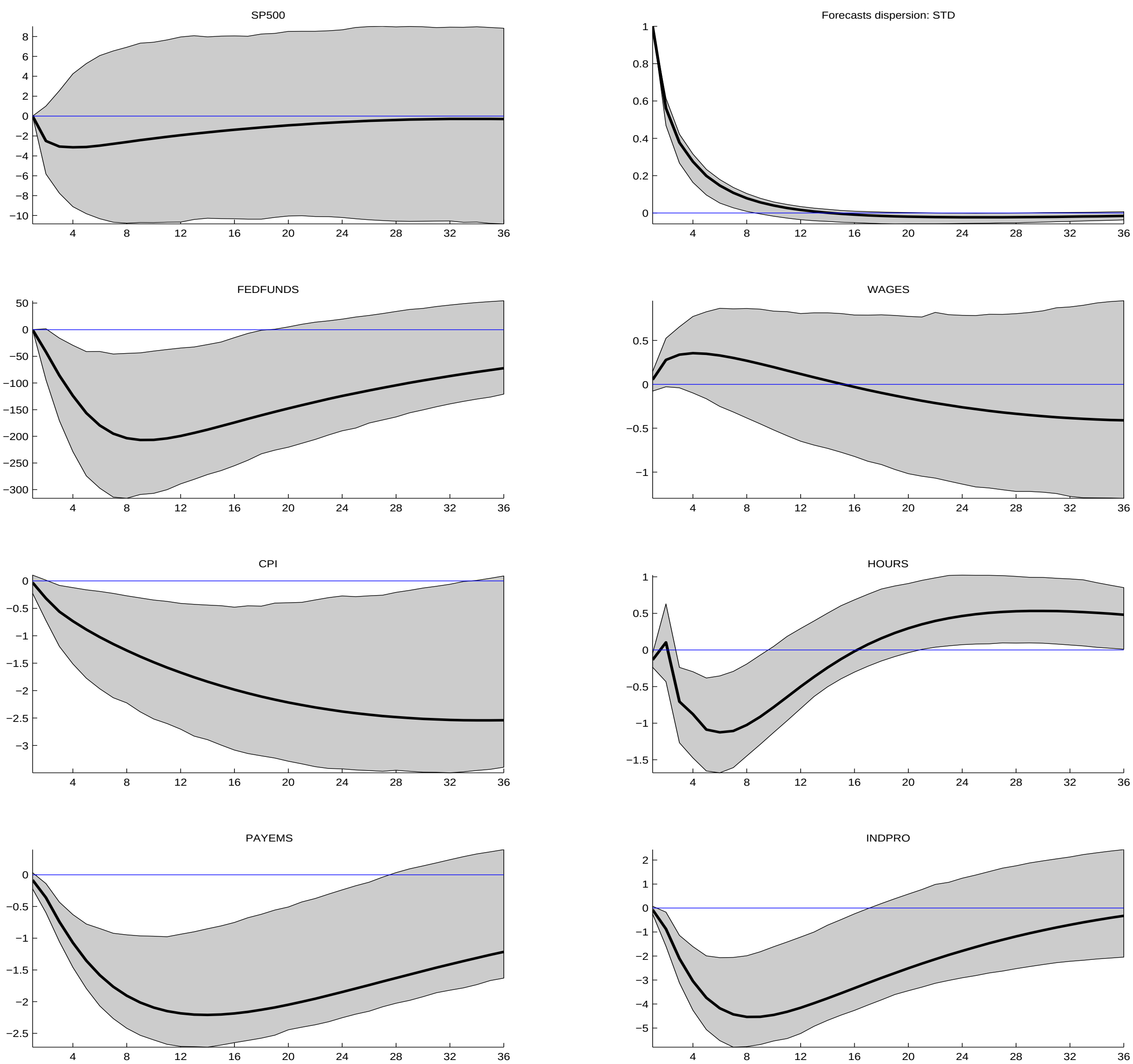
The figure shows the forecasts dispersion averaged across all 7 series for forecasting horizons of 1, 3 and 12 months. Two dispersion measures are taken: standard error (STD) and interquartile range (IQR). JLN represent the macro uncertainty from (Jurado et al. 2015) and SP500 represent the realized volatility of SP500. All series are standardized.

Figure 22: 1-month ahead forecasts dispersion across series



The figure shows the forecasts dispersion for each series and for 1-month forecasting horizon. The (1,1) panel compares the average dispersion measures against the first principal component from standardized individual dispersion series. The rest of the panels plot the two dispersion measures for each series as well as the macro uncertainty from (Jurado et al. 2015). All series are standardized.

Figure 23: Impulse responses to the shock on forecasts dispersion



This figure plots the impulse responses to the orthogonalized shock on forecast dispersion equation in the VAR-8 *model as in (Jurado et al. 2015). The lag order is set to 2 according to BIC. The gray represent 90% bootstrap confidence bands.*

7 Conclusion

This paper compares the performance of five classes of forecasting models on four types of time series in an extensive out-of-sample exercise. The classes of models considered are standard univariate models (Autoregressive Direct, Autoregressive Iterative, ADL and ARMA(1,1)), factor-augmented regressions (e.g., Diffusion Indices), dynamic factor models (e.g., FAVAR, FAVARMA), other data-rich models (e.g., Complete Subset Regression) and standard forecast combinations (simple average, trimmed average, etc.). The types of data considered include real series (Industrial Production and Employment), nominal series (Consumer Price Index), the stock market index (SP500) and exchange rates (US-UK and US-Canada).

First, we find that data-rich models and forecast combination approaches perform well in predicting real series. The two dominating techniques to forecast Industrial Production growth and Employment growth are the IP-AVRG and the CSR. The worst performing are the RW models, which simply suggests that real series are highly predictable.

Second, we find that the ARMA(1,1) model predicts inflation growth incredibly well and outperform data-rich models. This good performance of the ARMA(1,1) model is likely attributable to the fact that inflation growth is exogenous with respect to the conditioning information set available to us. As a result, data-rich models are over-parameterized and therefore have poor generalization performance for this series.

Third, the best forecasts of the SP500 returns at short horizons are obtained by taking the average of all the other forecasts. Interestingly, the RW underperforms some data-rich models, which suggest that stock returns are predictable to some extent. Moreover, several data-rich models produce a significantly higher proportion of correctly signed forecasts for the SP500 returns than RW models.

Fourth, data-rich models are mostly of no use when it comes to predicting exchange rates pointwise. Univariate models deliver the best point forecasts at very short horizons while the RW models dominate at moderate and long horizons. However, the ARMA(1,1) and several data-rich models produce a significantly higher proportion of correctly signed exchange rates forecasts than RW models.

As the hyperparameters of our models (number of lags, number of regressors or factors, etc.) are recalibrated for each series, horizon and out-of-sample period, we are able to document that the optimal structure of our forecasting equations changes much over time. Finally, we find that the dispersion of out-of-sample point forecasts is a very good predictor of some macroeconomic and financial uncertainty measures used in the literature.

References

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

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

Baker, S. R., Bloom, N. & Davis, S. J. (2015), Measuring economic policy uncertainty, Technical report, NBER Working Paper No. 21633.

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, Technical report, FEDS 2016-024, Board of Governors of the Federal Reserve System.

Bedock, N. & Stevanovic, D. (2016), 'An empirical study of credit shock transmission in a small open economy', *Canadian Journal of Economics* .

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**, 387422.

Bloom, N. (2009), 'The impact of uncertainty shocks', *Econometrica* **77**(3), 623–685.

Boivin, J., Giannoni, M. & Stevanović, D. (2013), Dynamic effects of credit shocks in a data-rich environment, Technical report, Federal Reserve Bank of New York Staff Reports 615.

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.

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 and 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', *Review of Economic Studies* **Forthcoming**.

Chevillon, G. (2007), 'Direct multi-step estimation and forecasting', *Journal of Economic Surveys* **21**(4), 746–785.

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

Diebold, F. X. & Mariano, R. S. (1995), 'Comparing predictive accuracy', *Journal of Business and Economic Statistics* **13**, 253–263.

Diebold, F. X. & Pauly, P. (1987), 'Structural change and the combination of forecasts', *Journal of Forecasting* **6**, 21–40.

Dufour, J.-M. & Stevanovic, D. (2013), 'Factor-augmented VARMA models with macroeconomic applications', *Journal of Business and 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.

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.

Galbraith, J. (2003), 'Content horizons for univariate time series forecasts', *International Journal of Forecasting* **19**(1), 43–55.

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

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, Department of Economics, Bocconi.

Hallin, M. & Liska, R. (2007), 'Determining the number of factors in the general dynamic factor model', *Journal of the American Statistical Association* **102**, 603–617.

Jurado, K., Ludvigson, S. & Ng, S. (2015), 'Measuring uncertainty', *American Economic Review* .

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.

Ludvigson, S. C. & Ng, S. (2005), 'The empirical risk-return relation: A factor analysis approach', *Journal of Financial Economics* **83**, 171–222.

Mao Takongmo, C. & Stevanovic, D. (2015), 'Selection of the number of factors in presence of structural instability: a monte carlo study', *Actualit conomique* **91**, 177–233.

Marcellino, M., Stock, J. H. & Watson, M. W. (2006), 'A comparison of direct and iterated multistep ar methods for forecasting macroeconomic time series', *Journal of Econometrics* **135**, 499–526.

McCracken, M. W. & Ng, S. (2015), 'Fred-md: A monthly database for macroeconomic research', *Journal of Business and Economic Statistics* .

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

Rossi, B. (2013), 'Exchange rate predictability', *Journal of Economic Literature* **51**(4), 1063–1119.

Rossi, B., Sekhposyan, T. & Soupre, M. (2016), Understanding the sources of macroeconomic uncertainty, Technical report, Department of Economics, Universitat Pompeu Fabra.

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.

Stevanovic, D. (2016), 'Common time variation of parameters in reduced-form macroeconomic models', *Studies in Nonlinear Dynamics & Econometrics* **20**(2), 159–183.

Stock, J. H. & Watson, M. W. (2002), 'Forecasting using principal components from a large number of predictors', *Journal of the American Statistical Association* **97**, 11671179.

Stock, J. H. & Watson, M. W. (2002b), 'Macroeconomic forecasting using diffusion indexes', *Journal of Business and 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. (2005), Implications of dynamic factor models for var analysis, Technical report, NBER WP 11467.