Forecasting with Dynamic Factor Models: An empirical exercise

Edoardo Scalcione*

RoME Master in Economics

EIEF and LUISS

September 2022

Abstract

This paper evaluates the two Dynamic Factor Models (DFMs) introduced by Stock and Watson (2002a) (SW henceforth) and Forni, Hallin, Lippi, and Reichlin (2005) (FHLR henceforth) in an empirical exercise on Euro Area macroeconomic monthly data spanning the period from 1997 to 2018. An in-sample analysis of the estimated factors underlines the importance of the expectational variables in explaining the comovement of macroeconomic series and that the Great Recession has structurally changed the interpretation of the factors. The forecasting exercise shows that the DFMs outperform the univariate predictions for Industrial Production, Inflation and Unemployment rate but more on real variables rather than nominal ones and more during the recession periods. Following the nesting procedure of SW and FHLR introduced by D'Agostino and Giannone (2012), we provide evidence of the fact that imposing the structure of the DFMs in the forecasting equation improves the predictive performance on real variables and it is the main driver of the difference in the accuracy of the two models. Lastly, we show that the size and the selection of the variables to include in the panel affect the performance. Removing the variables that are mainly idiosyncratic or variables that have too cross-correlated idiosyncratic errors can improve the accuracy of the predictions.

*I thank Marco Lippi and Alessandro Giovannelli for their patient guidance and for having shared their data with me. Without them this work would never be possible. I also acknowledge the insightful comments that I have received by EIEF faculty during the Thesis Control Sessions.

1 Introduction

Economists and forecasters nowadays can rely on a vast number of economic time series. It seems reasonable that focusing only on a few key indicators can produce a substantial loss of useful information and, potentially, suboptimal forecasts. However, standard multivariate forecasting methods (e.g., Vector AutoRegression, VAR) can deal only with time series of low cross-sectional dimension, and they cannot be identified in situations where the number of the time series is of the same order or even larger than the length of the time series. Hence, VARs are not appropriate in this framework of large-dimensional data.

As a solution to this problem, the literature has given large space to the class of models called *dynamic factor models* (DFMs), because of their ability to model simultaneously and consistently datasets in which the number of series exceeds the number of time series observations.

Indeed, the main idea of the DFMs is that a small number of factors can capture the bulk of the variation of the series by exploiting their comovements while the remaining part is a mean-zero idiosyncratic disturbance. Therefore, the macroeconomic series, empirically characterized by strong comovement, are a natural field of application for factor models.

Geweke (1977) originally proposed the DFMs as a time-series extension of factor models previously developed for cross-sectional data. This justifies the term "dynamic" in the name to underline the time dimension of the data. In early influential work, Sargent, Sims, et al. (1977) show that two dynamic factors can explain a large fraction of the variance of important U.S. quarterly macroeconomic variables, including output, employment and prices. This central empirical finding that a few factors can explain a large fraction of the variance of many macroeconomic series is confirmed by many following studies; see for example Stock and Watson (2002a) and D'Agostino and Giannone (2012).

The DFMs literature is rich, and many different models have been presented in the last thirty years. Following Stock and Watson (2011), we can mention three main classes of models that differ by their assumptions and, consequently, by the estimation method of the factors.

The first class assumes that the cross-section dimension is finite, and the idiosyncratic term is cross-sectionally uncorrelated. The Kalman filter is used to compute the Gaussian likelihood, the parameters are estimated by maximum likelihood, and then the Kalman filter and smoother are employed to obtain efficient estimates of the factors (see Quah, Sargent, et al. (1993) for example).

The second class moves to an infinite dimension model allowing for some weak cross-correlation of the idiosyncratic term. Therefore, it considers a non-parametric estimation. The factors are estimated by the cross-sectional average of the data, typically principal components or related methods such as generalized principal components. The motivation of this estimation approach is that weighted averages of the idiosyncratic disturbances (under some assumptions) converge to zero by the weak law of large numbers, so that only linear combinations of the factors remain, and the space of the factors is consistently estimated as panel dimensions go to infinity. The models in Stock and Watson (2002a) and Forni, Hallin, Lippi, and Reichlin (2005) belong to this class.

The third class of models merges the statistical efficiency of the state space approach with the robustness and convenience of the principal components approach. The estimation procedure occurs in two steps. First, the factors are estimated by principal components or generalized principal components. In the second step, these estimated factors are used to estimate the unknown parameters of the state space representation. See Giannone et al. (2008) for more details on the estimation methodology.

In this work we focus on the second class of models, precisely on the two most popular DFMs. The first model is described by Stock and Watson (2002a) (SW henceforth) and uses principal components (PCs) to estimate the factors. The second model is introduced by Forni, Hallin, Lippi, and Reichlin (2005) (FHLR henceforth) and uses a two-step procedure of dynamic principal components and generalized principal components (GPCs) to estimate the factors.

Even if the literature has focused on the forecasting properties of the DFMs, many applications go beyond prediction exercises. Forni, Giannone, et al. (2009) and Forni, Gambetti, et al. (2020) use the DFMs in structural analysis, for example. In this framework, the dynamic factor structure is extremely useful to correct for measurement error and solves the non-fundamentelness problem that is sometimes present in structural VARs. See Forni, Gambetti, et al. (2020) for the so-called Common Component Structural VARs. Other interesting applications relate these models to structural factor-augmented vector autoregressions (see Bernanke et al. (2005)), DSGEs (see Boivin and Giannoni (2006)) but also instrumental variables (see Kapetanios and Marcellino (2010)). See Stock and Watson (2011) for a detailed overview on the literature on DFMs.

In this thesis we briefly describe the two models and how to nest them in the same framework underlying the main differences. Next, we compare the two models in an empirical exercise with both in-sample and out-of-sample analysis using a large monthly dataset of macroeconomic and financial time series for the Euro Area economy from 1997 to 2018. This exercise aims to interpret the factors and evaluate the forecasting performance of the DFMs and how is affected by the business cycle, the estimation design and the features of the dataset. The three variables of interest are industrial production, inflation and unemployment rate. The predictions span from one to twelve months ahead.

The literature comparing the predictive performance of SW and FHLR is prolific and it has reached a mixed conclusion. In a Monte Carlo study, Forni, Hallin, Lippi, and Reichlin (2005) show that FHLR is more precise than SW when there are persistent dynamics in the factors and in the idiosyncratic errors. Using the US dataset proposed by Stock and Watson (2002b), Boivin and Ng (2005) find that SW generally outperforms FHLR while D'Agostino and Giannone (2012) find that there is not an evident difference between the performance of the two models. Moreover, moving to the Euro Area, Albano (2016) and Forni, Giovannelli, et al. (2018) find that FHLR is the best model even if before the Great Recession of 2008 the DFMs perform poorly with respect to the standard univariate autoregressive predictions. Using a panel of German data Schumacher (2007) finds that FHLR is more accurate for GDP forecasts. See Eickmeier and Ziegler (2008) for a large meta-analysis of forty-six forecasting exercises using DFMs.

Our main results are that the DFMs outperform the univariate model for all the variables of interest, particularly for IP and UR. The DFMs are more precise during the recession periods and in the years after, especially for the unemployment rate. The relative performance of the DFMs is mixed but the key element that explains the different performance of FHLR and SW is the different forecasting equation. Imposing the structure of the DFMs in the forecasting equation improves the predictive performance on real variables. We interpret this result as evidence of the fact that DFM structure can be a particularly good representation for real variables. The size and the selection of the series in the panel affect the predictive performance. Removing the variables that are mainly idiosyncratic or variables that have too cross-correlated idiosyncratic errors can improve the accuracy of the predictions, particularly for SW.

Given that these results on the Euro Area data are reasonably aligned with the US literature, we have now reached the state of the art on the dynamic factor models on Euro Area data. Further research requires to go beyond and compare our factor models to alternative approach in data-rich environment. In particular, a common alternative is a penalized regression. Konzen and Ziegelmann (2016) use LASSO regression to forecast inflation while Li and Chen (2014) apply LASSO, elastic net regression and group-LASSO to Stock and Watson (2002b) US dataset. Li and Chen (2014) show that LASSO forecasts tend to perform slightly better than a DFM even if not uniformly over the sample period. The best result is obtained combining the two forecasts (linearly). As pointed out by the authors, indeed, the DFM can capture the large variations underlying multiple time series while LASSO-based approaches, on the other hand, can identify important predictors for recovering a sparse predictive model. Combining the two estimates seems to increase the predictive power.

An emerging literature applies also non-linear algorithms to macroeconomic forecasting with promising results. For example, Medeiros et al. (2021) find out that random forest model outperforms LASSO on US inflation. Therefore, further analysis on the Euro Area should include also non-linear algorithms in the comparison.

This work contributes to the vast literature on dynamic factors models along different dimensions.

Firstly, it moves the attention to the Euro Area considering a new dataset of Euro Area data that is larger and more recent than the US panel introduced by Stock and Watson (2002b).

Secondly, it extends the in-sample analysis of McCracken and Ng (2015) to the Euro Area data also introducing a dynamic analysis of the factors over time that should increase the interpretability of the estimated factors.

Thirdly, it enters the debated question about the relative performance of the two most popular DFMs (SW and FHLR) applying the nesting method described by D'Agostino and Giannone (2012) for the first time to European and more recent data (after 2000). Moreover, it also includes to the variable of interest the unemployment rate moving the debate on the relative performance of the models to a different dimension.

Finally, it extends the work on the role of data selection introduced by Boivin and Ng (2006) to Euro Area data and considers also the FHLR predictor while instead the cited paper considers only the SW model.

The work is organized as follows. In Section 2 we briefly overview the theory behind the DFMs and their identification assumptions. We also review the theoretical structure of SW and FHLR models underlying the differences and nesting the two models. In Section 3 we introduce our empirical exercise and report the results of the in-sample analysis of the factors, we compare the forecasting performance of the DFMs to the univariate forecasts and the relative performance of SW and FHLR. In the last part of the section, we consider the effect of data selection on the predictions, and we try to address two possible related concerns. Section 4 concludes and raise some question for future research. Some additional material such as figures and tables related to the empirical exercise in Section 3 is in the Appendix.

2 Dynamic Factor Models: Theory

Consider a sample of dimensions (N, T) of a covariance stationary process $\mathbf{Y}_t^N = (y_{1,t}, \ldots, y_{N,t},$ where N is of the same order of T or even greater. We are interested in forecasting some elements of $\mathbf{Y}_t^N$ h period ahead. A very simple idea could be to linearly project the series on the space spanned by the entire set of variables up to time t. However, under the assumption that $N \approx T$, this linear projection is unfeasible (curse of dimensionality). If we assume that most of the variation is explained by a few unobservable common factors, say $r \ll N$, then a possible solution is to consider a Dynamic Factor

Model (DFM). Call $\mathbf{X}_t^N = (x_{it}, \ldots, x_{Nt})$ the standardized version of $\mathbf{Y}_t^N$. Following Forni, Hallin, Lippi, and Reichlin (2000), in a DFM framework we assume that the standardized series x_{it} can be decomposed into:

$$x_{it} = \chi_{it} + \xi_{it} = b_{i1}(L)u_{1t} + b_{i2}(L)u_{2t} + \cdots + b_{iq}(L)u_{qt} + \xi_{it}, \tag{2.1}$$

where χ_{it} is the *common component*, ξ_{it} the *idiosyncratic component*, $b_{ij}(L),\ j = 1, \ldots, q$ linear filters in the lag operator L of order s and $\mathbf{u}_t = (u_{1t}, \ldots, u_{qt})$ a q-dimensional vector of orthonormal white noises that we call *common shocks*. The two components, χ_{it} and ξ_{it}, are unobservable.

Crucially, we assume that ξ_{it} is orthogonal to $\mathbf{u}_s$ with $s \in \mathbb{Z}$ so that the idiosyncratic term is orthogonal to the common shocks at any leads and lags.

Standard assumption in factors model, as in Sargent, Sims, et al. (1977), is that the idiosyncratic component is cross-sectionally orthogonal, and, consequently, the covariance matrix of the idiosyncratic term is diagonal. In this case, we call it an *exact factor structure*. As pointed out by Chamberlain and Rothschild (1983) and followed by Forni, Hallin, Lippi, and Reichlin (2000), Stock and Watson (2002a) and many studies on DFMs, this assumption is likely to be violated in the real world. We can partially relax this assumption by introducing a so-called *approximate factor structure*, where the idiosyncratic term is allowed to be "weakly" cross-correlated. The meaning of "weakly" will be clear very soon when we are going to define more precisely the asymptotic assumptions that allow the identification of the model.

A quite common assumption for DFMs is that it can be written as

$$x_{it} = a_i(L)\mathbf{f}_t + \xi_{it}, \tag{2.2}$$

where $\mathbf{f}_t$ is a $q \times 1$ vector of *dynamic factors* that follows a VAR scheme of the form $N(L)u_t$. Letting $\mathbf{F}_t = (\mathbf{f}_t', \ldots, \mathbf{f}_{t-s}')'$ and $A_i = (a_{i1}, \ldots, a_{is})$ we can write (2.2) as

$$x_{it} = A_i\mathbf{F}_t + \xi_{it}, \tag{2.3}$$

where $r = q(s+1)$ common factors have only a contemporaneous effect on the observed series. For this reason, (2.3) is typically called *static representation* and $\mathbf{F}_t$ *static factors* while we refer to (2.2) as *dynamic representation* given that the factors are loaded with their lags.

Regardless of the name, the static representation (2.3) incorporates the dynamic nature of the model. The spectral density of the common component has rank q, the number of the dynamic factors, indeed.

However, the class of DFMs that have a static representation is larger than the class described by (2.2). Indeed, there exist models that have a static representation even if they cannot be represented as in (2.2). Consider for example a model with static representation

$$x_{it} = a_1 F_{1t} + a_2 F_{2t} + \xi_{it},$$

where the static factors are

$$F_{1t} = (1 - \alpha_1 L)^{-1} u_t, \qquad F_{2t} = (1 - \alpha_2 L)^{-1} u_t.$$

Clearly this model cannot be expressed as in (2.2) but it has a static representation. Therefore, let generically express the static representation of a DFM in the following way:

$$x_{it} = \lambda_{i1} F_{1t} + \lambda_{i2} F_{2t} + \cdots + \lambda_{ir} F_{rt} + \xi_{it} = \mathbf{\Lambda}_i \mathbf{F}_t + \xi_{it}. \tag{2.4}$$

Denote by Γ_N^x the variance-covariance matrix of $\mathbf{X}_t^N$ and by Γ_N^χ and Γ_N^ξ the variance-covariance matrix of the common component and the idiosyncratic component. Analogously, we can define the spectral density of the observed series, the common component and the idiosyncratic component as $\Sigma_N^x(\theta)$, $\Sigma_N^\chi(\theta)$ and $\Sigma_N^\xi(\theta)$, where $\theta \in [-\pi, \pi]$ is the frequency. Note that, under the orthogonality assumption of the two components, both the variance-covariance and the spectral density of $\mathbf{X}_t^N$ can be decomposed into the sum of the equivalent matrices for the two components, i.e.,

$$\Gamma_N^x = \Gamma_N^\chi + \Gamma_N^\xi, \quad \text{and} \quad \Sigma_N^x(\theta) = \Sigma_N^\chi(\theta) + \Sigma_N^\xi(\theta), \quad \forall \theta \in [-\pi, \pi].$$

Now we can formally state the theorem proved by Chamberlain and Rothschild (1983) and the assumptions that guarantee the identification of the model as described in (2.4).

Under the restriction that the space spanned by the factors is finite-dimensional, the assumptions are:

Assumptions

1. For all $N \in \mathbb{N}$ the vector $\mathbf{X}_t^N$ is covariance stationary.

2. Let μ_{Nj}^x be the j-th eigenvalue of Γ_N^x and $\bar{\mu}_j^x = \sup_{N \in \mathbb{N}} \mu_{Nj}^x$. There exists a positive r such that $\bar{\mu}_r^x = \infty$ and $\bar{\mu}_{r+1}^x < \infty$.

Theorem (Chamberlain and Rothschild (1983))

Given the Assumptions above, $\mathbf{X}_t = (x_{it}, i \in \mathbb{N})$ *can be represented as in* (2.4) *where* $\mathbf{F}_t$ *is a stationary* r*-dimensional vector. Moreover,*

1. ξ_{it} *satisfies Assumption 1 and* $\bar{\mu}_1^\xi < \infty$ *where* $\bar{\mu}_1^\xi = \sup_{N \in \mathbb{N}} \mu_{N1}^\xi$.

2. χ_{it} *satisfies Assumption 1,* $\bar{\mu}_r^\chi = \infty$ *and* $\bar{\mu}_{r+j}^\chi = 0$ *for all* $j > 0$ *where analogously* $\bar{\mu}_j^\chi = \sup_{N \in \mathbb{N}} \mu_{Nj}^\chi$.

3. *The idiosyncratic term and the static factors are uncorrelated at any time* $t \in \mathbb{Z}$*.*

4. *The number of static factors, the common and the idiosyncratic component for all the series are uniquely identified.*

Also, the converse is true, i.e., if $\mathbf{X}_t$ *can be represented as in* (2.3) *with* χ_{it} *and* ξ_{it} *satisfying the points 1-3 of the Theorem, then* $\mathbf{X}_t$ *satisfies Assumption 1-2 above.*

The finite-dimension restriction can be relaxed so that the model cannot be represented as in (2.4) but only as in (2.1). A theorem similar to the one presented allowing for infinite-dimension space spanned by $(\chi_{1t}, \ldots, \chi_{Nt})$ is proved by Forni, Hallin, Lippi, and Reichlin (2001) and the estimation of the model presented in Forni, Hallin, Lippi, and Zaffaroni (2015). However, in the two DFMs that we present in this work, the finite-dimension restriction remains binding. The empirical exercise presented in the following sections already faces many possible specifications using only SW and FHLR, so we decided to focus only on finite-dimensional models leaving a similar exploration of more computationally intense models for future research[1].

In the framework described, we are going to estimate the factors taking linear combinations of the x's, in particular principal components (PCs) as in Stock and Watson (2002a) and generalized principal components (GPCs) as in Forni, Hallin, Lippi, and Reichlin (2005).

The intuition behind assumptions 1-2 is that we want the common component to survive to the aggregation while the idiosyncratic component should tend to zero in variance. Assumption 2, indeed, implies that the common component is pervasive in the sense that it is ruled out the possibility that some elements in $\mathbf{F}_t$ are loaded only by a finite number of x's.

Point 1 of the theorem gives a precise meaning to the "weak" cross-covariance that is allowed in the approximate factor structure.

In the following sections, in order to avoid a too-heavy notation the dependence on N of $\mathbf{X}_t$, Γ^x and all other quantities is not made explicit even if clearly still present.

[1] To have an idea of the forecasting performance of the infinite-dimensional model introduced by Forni, Hallin, Lippi, and Zaffaroni (2015) see the results on US data in Forni, Giovannelli, et al. (2018) and the equivalent exercise on Euro Area data in Albano (2016). Results in some Monte Carlo experiments are presented in Forni, Hallin, Lippi, and Zaffaroni (2017).

2.1 SW model

The first DFM that we consider in our empirical exercise is introduced by Stock and Watson (2002a). In this case, the static representation takes the form

$$x_{it} = \lambda_{i1}F_{1t} + \lambda_{i2}F_{2t} + \cdots + \lambda_{ir}F_{rt} + \xi_{it} = \mathbf{\Lambda}\mathbf{F}_t + \xi_{it},$$

where $\mathbf{\Lambda}$ and $\mathbf{F}_t$ are estimated using the first r principal components of the standardized series.

Define as D_r the diagonal matrix whose diagonal elements are the r largest eigenvalues of $\hat{\Gamma}_0 = T^{-1}\sum_{t=1}^{T}\mathbf{X}_t\mathbf{X}_t'$, the sample cross-covariance matrix, and V_r the $N \times r$ matrix whose columns are the corresponding r eigenvectors. Our estimated static factors are then:

$$\hat{\mathbf{F}}_t = V_r'\mathbf{X}_t, \tag{2.5}$$

and the estimated covariance matrix of the common component is:

$$\hat{\Gamma}_0^{\chi} = V_r D_r V_r'. \tag{2.6}$$

The estimated $\hat{\mathbf{F}}_t$ are then used as predictors in the following forecasting equation:

$$\hat{y}^{SW}_{i,t+h|t} = \hat{\alpha}_{ih}\hat{\mathbf{F}}_t + \hat{\beta}_{ih}(L)y_{it}, \tag{2.7}$$

i.e., the projection of $y_{i,t+h}$ on the space spanned by $(\hat{\mathbf{F}}_t, y_{it}, y_{i,t-1}, \dots)$[2]. The equation (2.7) is estimated by simple ordinary least squares (OLS) regression. The presence of lagged values of the series can be explained by the idea that they can capture possible autocorrelation of the idiosyncratic term. Note that, as underlined by D'Agostino and Giannone (2012), (2.7) does not incorporate the dynamic structure of the model that is instead embedded in the FHLR approach. We address in more detail this difference in the forecasting equation in Section 2.3.

[2] In this case we are considering only a contemporaneous dependence on the factors. This forecasting method is known as Diffusion Index - Autoregressive (DI-AR). Theoretically, it could be possible to also consider the lagged values of the factors, but Stock and Watson (2002a) and other empirical works (see for example Forni, Giovannelli, et al. (2018)) show that the effect on forecasting performance is null or even negative. Moreover, we show that in many cases even the lagged values of the series are not beneficial and so the optimal method is a simple Diffusion Index (DI). Albano (2016); D'Agostino and Giannone (2012); Forni, Giovannelli, et al. (2018) find comparable results.

2.2 FHLR model

The model introduced by Forni, Hallin, Lippi, and Reichlin (2005) moves to a frequency domain approach as in Forni, Hallin, Lippi, and Reichlin (2000) but the space spanned by the factors is still assumed to be finite. The procedure originally proposed by Forni, Hallin, Lippi, and Reichlin (2000) uses the *dynamic principal component* to obtain a consistent estimator of the common component. However, this method creates serious issues for forecasting because it is based on two-sided filters whose performance deteriorates at the margins of the sample. Forni, Hallin, Lippi, and Reichlin (2005) propose a two-step approach to overcome this problem:

1. Estimate the covariance matrix of the two orthogonal components as in Forni, Hallin, Lippi, and Reichlin (2000) using a frequency domain method.

2. Given the two covariance matrices estimate the factors using the generalized principal components.

In the first step from the sample autocovariance of order k, namely $\hat{\Gamma}^x_k = (T-k)^{-1}\sum_{t=1}^{T-k}\mathbf{X}_t\mathbf{X}'_{t-k}$, we can obtain a consistent estimator of the spectral density:

$$\hat{\Sigma}^x(\theta) = \frac{1}{2\pi}\sum_{k=-m}^{m} w_k \hat{\Gamma}^x_k e^{-i\theta k}, \tag{2.8}$$

where $w_k = 1 - \frac{|k|}{m+1}, k = -m, \dots, m$ is a triangular window of length m. In this way, we can estimate the spectra at $2m+1$ equally spaced frequencies in $[-\pi, \pi]$. Let $\mathcal{D}_q(\theta)$ be a diagonal matrix having on the diagonal the q largest eigenvalues of $\hat{\Sigma}^x(\theta)$ and $\mathcal{V}_q(\theta)$ the $N \times q$ matrix whose columns are the corresponding eigenvectors. Then, if $\mathbf{X}_t$ is driven by q dynamic factors we can estimate the spectral density of the common component as:

$$\hat{\Sigma}^\chi(\theta) = \mathcal{V}_q(\theta)\mathcal{D}_q(\theta)\mathcal{V}_q(\theta)'. \tag{2.9}$$

The estimated spectral density of the idiosyncratic component can be computed residually as:

$$\hat{\Sigma}^\xi(\theta) = \hat{\Sigma}^x(\theta) - \hat{\Sigma}^\chi(\theta).$$

In order to get $\hat{\Gamma}^\chi_k$ and $\hat{\Gamma}^\xi_k$, we use the inverse discrete Fourier transforms

$$\hat{\Gamma}^\chi_k = \frac{2\pi}{2m+1}\sum_{h=-m}^{m}\hat{\Sigma}^\chi(\theta_h)e^{-i\theta_h k}$$
$$\hat{\Gamma}^\xi_k = \frac{2\pi}{2m+1}\sum_{h=-m}^{m}\hat{\Sigma}^\xi(\theta_h)e^{-i\theta_h k},$$

with $\theta_h = \frac{2\pi}{2m+1}h,\ h = -m, \ldots, m$.

In the second step, from the estimated covariances we obtain a linear combination of the x's that theoretically should be more efficient than the standard principal components used in SW model. In this case, the optimal weights of the r linear combination are the first r *generalized eigenvectors*, the columns of the $N \times r$ matrix V_{rg}, associated with the first r *generalized eigenvalues*, i.e., the elements in the diagonal matrix D_{rg}. The generalized eigenvectors are recursively defined by the GPC problem:

$$\hat{\Gamma}_0^{\chi} V_{rg} = \hat{\Gamma}_0^{\xi} V_{rg} D_{rg} \text{ s.t.} \quad V'_{rg} \hat{\Gamma}_0^{\xi} V_{rg} = I_r. \tag{2.10}$$

The estimated static factors in FHLR are then the first r generalized principal components, i.e.,

$$\hat{\mathbf{F}}_t^G = V'_{rg} \mathbf{X}_t. \tag{2.11}$$

In the empirical exercise, it is important to underline that, following Forni, Hallin, Lippi, and Reichlin (2005), we force to zero the off-diagonal terms of the estimated $\hat{\Gamma}_0^{\xi}$ because it is ill-conditioned when N is large compared to T as it is in our data. Simulations show that this procedure significantly improves the forecasting performance not affecting the consistency of the estimator. Indeed, Forni, Hallin, Lippi, and Reichlin (2005) prove that replacing $\hat{\Gamma}_0^{\xi}$ with any symmetric positive semi-definite matrix with bounded eigenvalues does not affect the consistency results.

Generally, it can be shown that the generalized principal components of $\mathbf{X}_t$ are equal to the standard principal components of the transformed vector $\widetilde{\mathbf{X}}_t = (\hat{\Gamma}_0^{\xi})^{-1/2} \mathbf{X}_t$. Given the forced diagonal structure of $\hat{\Gamma}_0^{\xi}$, the generalized principal components are equal to the standard principal components of the series weighted by the inverse of the estimated idiosyncratic variance. In this way, we give more weight to variables that are well explained by the common factors and less importance to the series that are mainly idiosyncratic. It is now clear why the FHLR method should provide a more efficient weighting scheme in the estimation of the static factors.

The FHLR approach is different from SW also in the forecasting projection because it explicitly considers the dynamic structure imposed by the model, using the estimated $\hat{\Gamma}_0^{\chi}$ that conveys information contained in the whole covariance sequence $\{\hat{\Gamma}_k^x,\ k \in \mathbb{Z}\}$. In this case, following Forni, Hallin, Lippi, and Reichlin (2005) we predict the series only using forecasts of the common component. There is some attempt of idiosyncratic forecast in the literature (see D'Agostino and Giannone (2012) for example) but the results suggest that the idiosyncratic term is unpredictable. The forecast of the common component

h step ahead is, in vector, the projection:

$$\hat{\chi}_{t+h|t} = \text{Proj}(\chi_{i,t+h}|\hat{\mathbf{F}}_t^G) = \hat{\Gamma}_h^{\chi} V_{rg}(V_{rg}'\hat{\Gamma}_0^x V_{rg})^{-1} V_{rg}' \mathbf{X}_t. \tag{2.12}$$

Given the prediction of the common component, we compute the prediction of the series i h step ahead, namely $\hat{y}_{i,t+h|t}^{FHLR}$, reattributing the mean and the variance as the factors estimation is computed on the standardized data $\mathbf{X}_t$, i.e.,

$$\hat{y}_{i,t+h|t}^{FHLR} = \hat{\sigma}_i \hat{\chi}_{t+h|t} + \hat{\mu}_i.$$

2.3 SW vs FHLR: Nesting the models

In order to properly compare the two different DFMs is useful to nest them. For this reason, we follow the nesting of the two approaches described by D'Agostino and Giannone (2012) and here briefly presented. Note that FHLR introduces two main new elements:

1. FHLR uses GPCs instead of PCs and, consequently, the weighting scheme according to the signal-to-noise ratio;
2. FHLR performs a constrained (CON) projection using the estimated autocovariance of the two components while SW uses an unconstrained (UNC) OLS regression.

Therefore, we can define two hybrid models that incorporate only one single difference as described by the following scheme:

	UNC	**CON**
PC	SW	PC/CON
GPC	GPC/UNC	FHLR

In this way, if we want to evaluate the effect on the forecasting performance of the weighting scheme associated with the generalized principal components, we can compare PC/UNC (*alias* SW) and GPC/UNC (or equivalently PC/CON and GPC/CON). If instead we are interested in understanding what is the impact of the constrained projection, we can compare the predictive performance of PC/CON (*alias* SW) and PC/UNC (or equivalently of GPC/CON and GPC/UNC).

The GPC/UNC is computed by replacing the static factors estimated by the principal components $\hat{\mathbf{F}}_t$ with the equivalent factors estimated using the generalized principal components $\hat{\mathbf{F}}_t^G$ in (2.7). The forecasting equation becomes:

$$\hat{y}_{i,t+h|t}^{GPC/UNC} = \hat{\alpha}_{ih}\hat{\mathbf{F}}_t^G + \hat{\beta}_{ih}(L)y_{it}, \tag{2.13}$$

where the coefficients are estimated by OLS.

The PC/CON model, instead, is obtained by replacing in (2.12) the matrix V_{rg} with the matrix V_r whose columns are the first r standard eigenvectors of $\hat{\Gamma_0^x}$. The projected common component is then

$$\hat{\chi}_{t+h|t}^{PC/CON} = \hat{\Gamma}_h^{\chi} V_r (V_r' \hat{\Gamma}_0^x V_r)^{-1} V_r' \mathbf{X}_t. \tag{2.14}$$

It follows that the predicted series is

$$\hat{y}_{i,t+h|t}^{PC/CON} = \hat{\sigma}_i \hat{\chi}_{t+h|t}^{PC/CON} + \hat{\mu}_i.$$

Note that as the number of common shocks q increases, the dynamic structure restriction in (2.12) becomes less stringent; if as an extreme exercise we assume $q = n$ and we use a rectangular window, i.e., $w_k = 1,\ k = -m, \ldots, m$, the estimated autocovariance of the common component coincides with the sample autocovariance matrix and the equation (2.14) coincides with the OLS regression (without lagged values of the predicted series) in (2.7). In Section 3.3.2 we compare the predictive performance of these four nested models trying to understand which of the two new elements drives the difference (when evident) between the performance of SW and FHLR.

3 The empirical exercise

In this section we introduce an empirical exercise with the DFMs introduced above. The exercise includes both an in-sample and out-of-sample analysis.

The aim of the in-sample analysis is to interpret the estimated factors and how they evolve over time. The out-of-sample exercise, instead, firstly compares the forecasting performance of the DFMs with a univariate predictor and then among themselves using the nesting procedure in Section 2.3. In the last part of the out-of-sample exercise, we evaluate the effect of the size and features of the panel used in the exercise on the predictive performance of the two models.

3.1 Data description

The dataset employed in this empirical exercise is composed of 309 macroeconomic monthly series spanning the period from January 1997 to December 2018. Data, therefore, include the Great Recession originated in 2007 and its spillover effect in the Euro Area from the second quarter of 2008 to the second quarter of 2009. It also includes the so-called Euro Area Sovereign Debt crisis in 2012 while the

COVID-19 pandemic period is left out.

The series refer both to the Euro Area (EA 19) and to the four main European economies: France, Germany, Italy and Spain. The variables can be grouped by their measurement domain: Building Permits & Civil Engineering (BP), Consumer Survey based Confidence Indicators (CS), Harmonized Consumer Price Indices (CPI), Harmonized Unemployment Rates (UR), Industrial Production (IP), Industry & Construction Surveys (ICS), Money & Interest Rates (MIR), Producer Price Index (PPI), Service Surveys (SS) and Turnover & Retail Sales (TRS).

Figure 3.1 displays the composition of the dataset along the two dimensions just presented. The distribution of the series is quite heterogeneous both across countries and categories. Industrial Production and Industry & Construction surveys are the most present in particular for Germany and France while the latter is almost absent for Italy. In the Appendix, the complete set of variables with their denomination is listed in Table A.1.

To our knowledge, this is the first attempt to use SW and FHLR on this dataset which is also fairly large as compared to the US dataset considered by Stock and Watson (2002b) and Forni, Hallin, Lippi, and Reichlin (2005), 150 series, but also to the EA dataset used by Albano (2016); Forni, Giovannelli, et al. (2018), 176 series. On the contrary, the time dimension is quite small. Section 3.3 addresses this issue in the description of the design of the forecasting exercise.

In order to achieve stationarity, as required by the DFMs estimation, the series are converted in the first difference (some survey variables), the first difference of the logarithm (mainly real variables and unemployment) and the second difference of the logarithm (mainly nominal variables). After the transformation, we reject the unit root presence in the Augmented Dickey-Fuller test for all the series considered. To have a more detailed frame of the filter applied to each series, see Table A.1 in the Appendix. As usual in the DFMs framework, no treatment for outliers is applied.

The main motivation behind the factor representation is the strong comovement among macroeconomic series which is a well-known empirical result. This provides empirical evidence in favor of the idea that a small number of shocks drive the entire economy. At first glance, the comovement in our dataset is evident from the share of the variance explained by the estimated factors. Formally, the explained variance is computed for the static factors as the ratio $\text{trace}\, D_r / \text{trace}\, \hat{\Gamma}_0^x$ while for the dynamic factors as the ratio $\text{trace}\, \hat{\Gamma}_0^\chi / \text{trace}\, \hat{\Gamma}_0^x$ where $\hat{\Gamma}_0^\chi$ is computed using the first q dynamic principal components as described in Section 2.2. These two measures are reported in Table 3.1 for q and r ranging from 1 to 10.

Table 3.1 shows that a small number of factors is able to capture almost the bulk of the variance of the entire set of variables. Two dynamic factors and ten static factors can explain almost half of the

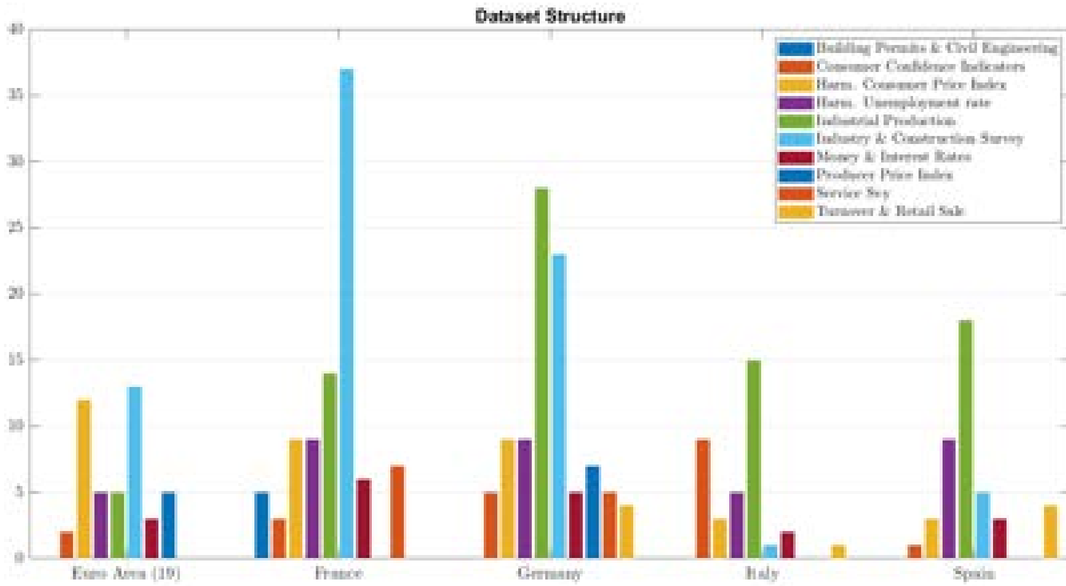

Figure 3.1: Dataset Composition by area and measurement domain

total variance.

Dynamic factors account for at least the same variance explained by the static factors by construction; however, it seems clear that there is evidence of a dynamic structure in the dataset. Remember that $r = q(s + 1)$, so given that $q = 2$ and $r = 10$ factors explained the same amount of variance then it should be $s \sim 5$ suggesting some dynamics.

Table 3.1: Variance explained by the estimated factors

	1	2	3	4	5	6	7	8	9	10
q *dynamic factors*	0.34	0.48	0.59	0.67	0.74	0.80	0.84	0.88	0.90	0.92
r *static factors*	0.14	0.22	0.28	0.31	0.34	0.37	0.39	0.41	0.43	0.45

3.2 In-sample Analysis

The first aim of our empirical exercise is to analyze in-sample the estimated factors trying to understand which series are better explained by the factors and how this depends on the business cycle.

For this in-sample analysis, we fix $r = 5$ in SW model and $q = 3, r = 5$ in FHLR. We consider these values because they are exactly in the middle of the set of possible parameters that we are going to consider in the forecasting exercise, i.e., $r = 1, \dots, 10$ and $q = 1, \dots, 5$. Moreover, they also result to be the best specifications available or close to them in many situations in the following forecasting exercise.

As in McCracken and Ng (2015), the main statistic that we want to consider is the explanatory power of each factor, namely $R_i^2(k)$ for $i = 1, \dots, N$ and $k = 1, \dots, r$. Once estimated the r factors by

SW or FHLR, we regress each factor k on each series i to obtain the explanatory power $R_i^2(k)$. Averaging on the series we compute the average explanatory power of each factor, i.e., $R^2(k) = N^{-1}\sum_{i=1}^{N} R_i^2(k)$.

Table 3.2 presents the ten variables mostly explained by the first five factors estimated by SW, i.e., the 10 variables with the highest $R_i^2(k)$ for $k = 1, \dots, 5$. The factors are estimated using the entire sample. The same table is in the Appendix for the first five factors estimated by FHLR, see Table A.2.

The first factor in both cases, captures mostly the variation in survey variables, in particular variables related to the personal outlook in industrial production. It is not surprising then that the first factor seems to trace the business cycle quite precisely, see Fig. A.1 in the Appendix. As in McCracken and Ng (2015), we also consider as a rough business cycle indicator the cumulative sum of the first factor, i.e., $\hat{\mathbf{F}}_t^{cum} = \sum_{h=1}^{t} \hat{\mathbf{F}}_j$. There are indeed some examples of business cycle indexes based on DFMs even with a more elaborated methodology. It is worth mentioning on this topic *Eurocoin* as introduced by Altissimo et al. (2001).

Note that the first factor is by construction the most informative in the sense that explained the largest share of variance (almost 14%), so it is interesting to point out that these expectational variables explain a substantial share of variation in the macroeconomic series.

In SW, the second factor is still related to surveys while factors 3 and 4 are mainly related to prices and factor 5 to real activities. In FHLR, instead, the second and third factors are mainly real factors while the fifth is a nominal factor. Factor 4 is more mixed. $R_i^2(4)$ are low indeed.

Table 3.2: Explanatory power Factors 1-5 SW

$R_i^2(1)$	Variable	$R_i^2(2)$	Variable	$R_i^2(3)$	Variable	$R_i^2(4)$	Variable	$R_i^2(5)$	Variable
0.562	ICS-23	0.493	SS-305	0.349	CPI-213	0.297	CPI-218	0.243	IP-148
0.561	ICS-48	0.456	ICS-28	0.310	CPI-204	0.263	CPI-235	0.230	IP-165
0.545	ICS-26	0.455	SS-307	0.306	CPI-234	0.260	CPI-220	0.226	IP-144
0.536	IP-109	0.453	ICS-44	0.267	CPI-216	0.235	IP-167	0.224	IP-164
0.534	ICS-45	0.446	SS-308	0.224	IP-163	0.221	CPI-221	0.223	IP-145
$R^2(1) =$	0.139	$R^2(2) =$	0.083	$R^2(3) =$	0.054	$R^2(4) =$	0.036	$R^2(5) =$	0.032

To understand how the factors change over time, we consider the following exercise. We estimate the model recursively from January 2004 up to December 2018 extending the estimation sample month by month. At each period we compute our measure of the explanatory power of the factors. Figure 3.2 and Figure A.2 in the Appendix present the results. On the x-axis there is the estimation period, on the y-axis the series, grouped by category. The color is assigned according to $R_{it}^2(k),\ k = 1, \dots, 5$ where t is the estimation period so that the sample used for the estimation is $[1, t]$.

Figure 3.2 and Figure A.2 show that there is indeed some dependence over the business cycle. The Great Recession in 2008-2009 seems to be a permanent shock in the explanatory power of the factors. The high correlation with expectational variables stands only after 2009. It is interesting to note that when the factors are computed following the FHLR approach, see Figure A.2, the Great Recession shock seems to be transitory, at least for the first two factors. In Factor 3 the permanency of the shock is most evident. In both approaches, the third factor starts as an expectational factor and then changes into a mainly nominal factor in SW and a more mixed factor in FHLR.

FHLR's Factors 4 and 5 also show a shock in 2011-2012 after the Debt Crisis and it is also clear here that these factors have a more mixed nature. In the Appendix, the same Figures are reported for the factor loadings, see Figures A.3 and A.4. See the Appendix for more details on the definition of the loadings.

Lastly, we want to directly compare the factors estimated according to the two different DFMs considered. In order to do so, using the same recursive exercise, we compute the correlation across the same factors in the two approaches, namely $\rho_{kt} = \text{corr}(\hat{\mathbf{F}}_{kt}, \hat{\mathbf{F}}^G_{kt})$, $k = 1, \ldots, 5$ and t denotes the estimation period. Remember that here $\hat{\mathbf{F}}_{kt}$ and $\hat{\mathbf{F}}^G_{kt}$ have dimension $t \times 1$ given the recursive scheme.

Figure 3.3 displays the correlations in a heatmap with an interpretation similar to 3.2, but on the y-axis there are now the five factors. It is evident that the first three factors are highly correlated, particularly after the Great Recession where the correlation is close to 1. The last two factors, instead, are substantially different after 2009. The last factors, indeed, are almost orthogonal in the period between the two crises.

3.3 Forecasting exercise

The aim of this section is to evaluate the performance of the two DFMs in forecasting Industrial Production (IP), Consumer Price Index (CPI) and Unemployment rate (UR) at different horizons. The forecasting period is from February 2007 up to the end of the sample, December 2018. For the predictions, we use a rolling window of ten years[3] so that at time t we use the sample $[t-119, t]$ to predict the series h step ahead, $h = 1, 3, 6, 12$ months. This means that we start with the ten years sample from March[4] 1997 to February 2007 to predict March 2007, May 2007, August 2007 and February 2008. The last estimation period is December 2017 given that the sample ends in December 2018, 12 months after.

It is common in the literature to divide the sample into a *pre-sample* and a *sample proper*. The first one is used to calibrate the model, so choosing the optimal parameters, and then in the latter, only the best specifications of each model are estimated and compared. This is the golden rule to properly compare two different models. See as an example Forni, Giovannelli, et al. (2018). However,

[3] We repeat the same exercise also with a rolling window of 5 years to assess how the window size affects the predictions. See Section 3.3.1.

[4] Note that we lose the first two months of the sample because prices are transformed in the second difference of the logarithm to achieve stationarity.

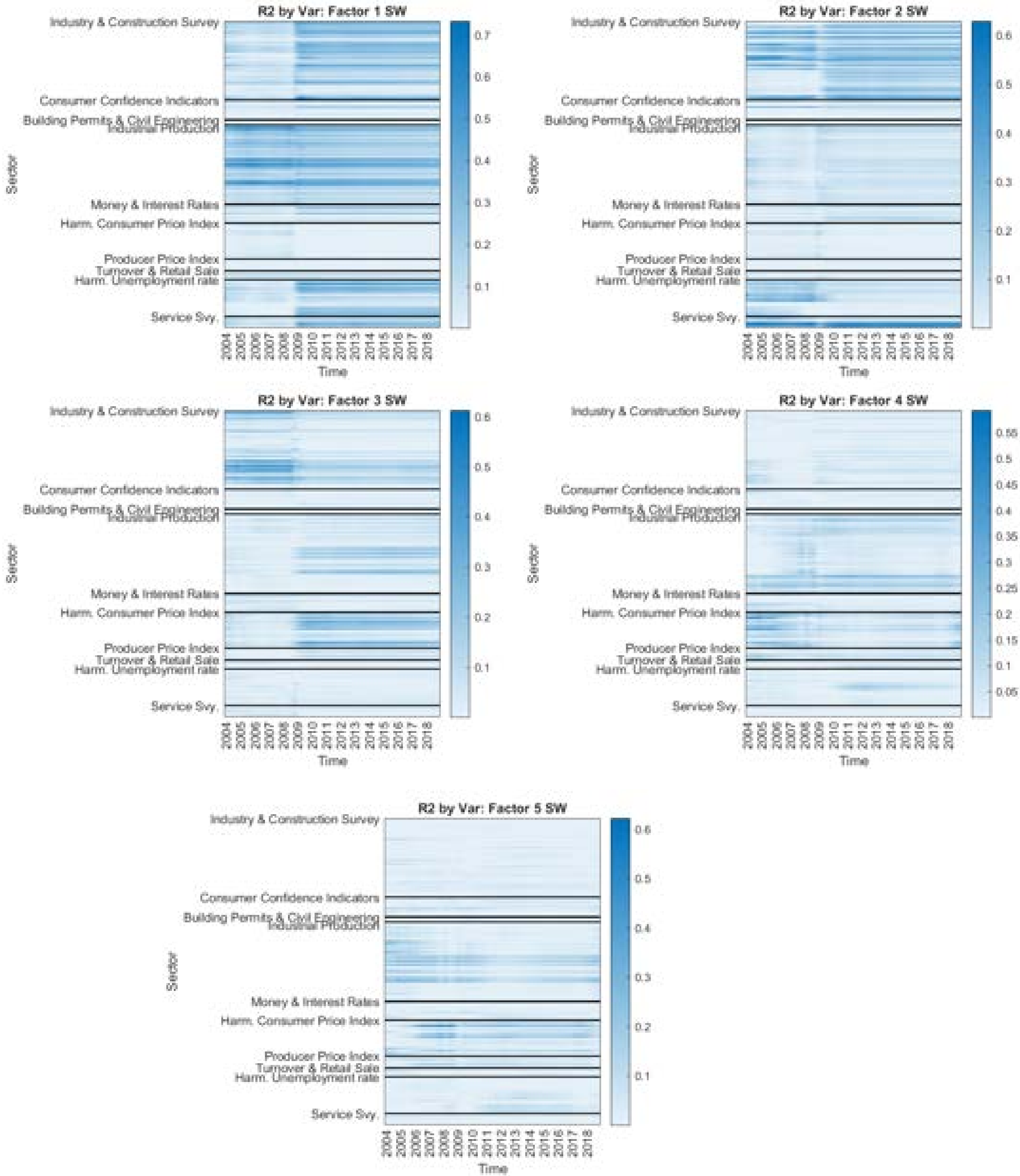

Figure 3.2: $R^2_{it}(k),\ k = 1, \ldots, 5$, SW Factors

this methodology requires a fairly large sample. This is not our case. For this reason, we have decided to use the entire sample available as *sample proper*. As in D'Agostino and Giannone (2012), we present

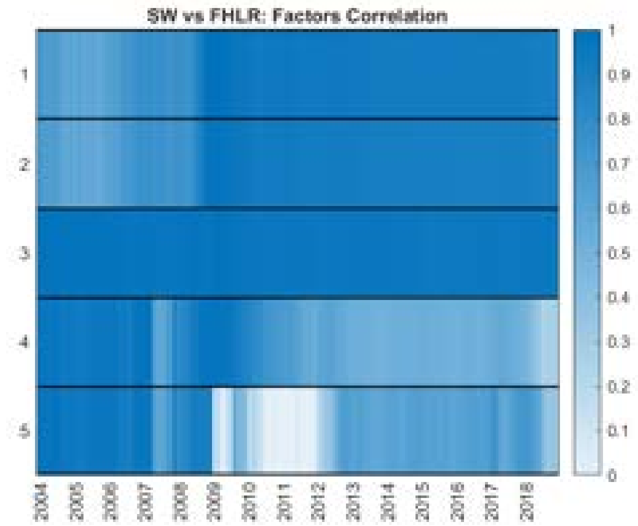

Figure 3.3: $\rho_{kt},\ k = 1, \ldots, 5$, SW vs FHLR: Factors correlation

a battery of specifications. Note that if only a particular specification of a model is better than a specification of another model it is not a conclusive result on the relative performance of the models. It may be possible that the optimal parameters would have been disregarded *ex-ante* in the calibration and so it would not have been the specification used to evaluate the relative performance of the model. Nonetheless, we still think that this forecasting exercise is informative, especially if the predictions are similar for a set of parameters (at least a subset of the space) and/or the relative performance of different specifications is quite uniform over the sample period. Figure A.5 in the Appendix seems to be quite reassuring on this.

Following (2.7),(2.12),(2.13) and (2.14), we consider only direct forecasting at each horizon h and not iterating one-step-ahead.

In the exercise we also compute as a benchmark the predictions of the univariate autoregressive (AR) model that are given by the forecasting equation $\hat{y}^{AR}_{i,t+h|t} = \hat{\beta}_{ih}(L)y_{it}$, where the coefficients are estimated by OLS.

Let define the target of our predictions as the level of the EA log of industrial production index and of the harmonized unemployment rate and the EA first difference of the logarithm of the harmonized CPI, i.e., EA monthly inflation. These definitions are aligned with the literature, see D'Agostino and Giannone (2012) and Forni, Giovannelli, et al. (2018).

Note that the literature has mainly focused on IP and CPI. We decided to also include UR because we think it may be of great interest to evaluate the predictive performance on this crucial monthly macroeconomic series.

As the three variables are treated in $\mathbf{Y}_t$ to achieve stationarity, we need some transformations to obtain the forecasting target. Let in general x_t be a time series and $y_t = \log x_t - \log x_{t-1}$ be the first difference of the logs and $z_t = \log x_t - 2\log x_{t-1} + \log x_{t-2} = y_t - y_{t-1}$ be the second difference of the logs and, recursively, the first difference of the log first difference. Note that $x_{t+h} = y_{t+1} + \cdots + y_{t+h} + x_t$ and that $y_{t+h} = z_{t+1} + \cdots + z_{t+h} + y_t$. Clearly we can compute the target prediction by cumulating the forecasts of the stationary variables from 1 up to h step ahead. Therefore, define the target[5] at time $t+h$ generically as $Z_{i,t+h} = Y_{i,t+1} + \cdots + Y_{i,t+h}$. For the three variables of interest the target is then:

$$Z_{i,t+h} = \begin{cases} \log IP_{t+h} - \log IP_t & \text{for i=IP} \\ (1-L)\log CPI_{t+h} - (1-L)\log CPI_t & \text{for i=CPI} \\ \log UR_{t+h} - \log UR_t. & \text{for i=UR} \end{cases} \tag{3.1}$$

For all the variables the forecast is:

$$\hat{Z}_{i,t+h|t} = \hat{Y}_{i,t+1|t} + \cdots + \hat{Y}_{i,t+h|t}, \tag{3.2}$$

and the prediction error of the model m, normalized for the horizon's length is:

$$FE^m_{ith} = \frac{1}{h}\left(Z_{i,t+h} - \hat{Z}_{i,t+h|t}\right) = \frac{1}{h}\left((Y_{i,t+1} - \hat{Y}_{i,t+1|t}) + \cdots + (Y_{i,t+h} - \hat{Y}_{i,t+h|t})\right). \tag{3.3}$$

Note that for the models with an unconstrained regression as forecasting equation we can also directly regress the target at time $t+h$ $Z_{i,t+h} =$ on the factors $\hat{\mathbf{F}}_t$ and the lagged values of the target. This is the option described by Stock and Watson (2002b). Empirically, this procedure tends to outperform the theoretical predictions obtained using cumulated forecasts as in (3.2). In the Appendix, Figures A.6, A.7 and A.8 compare the two methods and show that the cumulated prediction is outperformed for all the variables considered. For each model m, we can now define the Mean Square Forecast Error (MSFE) as

$$MSFE^m_{ih} = \frac{1}{(T_1 - h) - T_0 + 1}\sum_{t=T_0}^{T_1-h}\left(FE^m_{ith}\right)^2, \tag{3.4}$$

where T_0 and T_1 are the first and last dates of the forecasts. The optimal specification of each model is determined according to the MSFE. We consider the average MSFE across all horizons. In order to compare two different models m_1 and m_2, we use the ratio of their MSFEs, known as Relative Mean Square Forecast Error (RMSFE), i.e.,

$$RMSFE^{m_1/m_2}_{ih} = \frac{MSFE^{m_1}_{ih}}{MSFE^{m_2}_{ih}}.$$

[5] Note that here we are not including the level at time t in the target because it is clearly known at time t and it simplifies in the forecasting error.

If the RMSFE is below one it means that m_1 outperforms m_2 on the series i at the horizon h, if it is above the unit the opposite is true.

In the figures, instead, to evaluate the performance over time, we introduce two other measures: the cumulative sum of the squared forecasting errors (CSSFE) and the cumulative sum of the squared forecasting errors differences (CSSFED) between models m_1 and m_2. The two measures are defined at time t respectively as $\sum_{\tau=T_0}^{t} \left(FE_{i\tau h}^{m}\right)^2$ and $\sum_{\tau=T_0}^{t} \left(FE_{i\tau h}^{m_1}\right)^2 - \left(FE_{i\tau h}^{m_2}\right)^2$.

At any point in time, a CSSFED above zero means that the model m_2 outperforms the m_1 model, while positive (negative) changes in the slope of the CSSFED suggest that there is an increase (decrease) in the relative performance of the m_2 model with respect to the m_1 model. Model m_1 is the benchmark model while m_2 is the proposed model.

The exercise has the following structure. Firstly, we compare the DFMs to the benchmark autoregressive (AR) model, then we properly compare SW and FHLR using the nesting procedure explained in Section 2.3. In this way, we want to assess the benefit (if there is) of each of the two novelty elements introduced by FHLR.

3.3.1 DFMs vs AR

We report the results for a fairly large space of the parameters and compare the performance of the two DFMs with respect to the benchmark AR using the RMSFE on the entire sample, the cumulative FE^2 and CSSFED with respect to AR. We also test against the null of global equal performance of two predictors with the Diebold-Mariano test (see Diebold and Mariano (1995)). The p-values for both the full sample (2007-2018) and post Great Recession (2010-2018) are in Table A.9 in the Appendix.

For SW model we allow the number of static factors r to vary from 1 to 10 and find out that the best specification, according to the MSFE, is $r = 5$ for both IP and CPI and $r = 6$ for UR. As previously stressed, Figure A.5 provides two important elements of evidence that the identified best specifications can be informative about the performance of the model: i) the model is quite robust (at least on short-horizon predictions) to the number of factors given that the values close to the optimal r tend to have very similar performance, ii) the cumulative FE^2 of the optimal r often not only stochastically dominates the others but also tends to have a parallel or even flatter trend on the entire sample. Given this evidence, we are confident that these specifications can be representative of the relative performance of the models. See Table A.3 in the Appendix for more details on the RMSFE with different choice of the parameters.

Following the forecasting equation in (2.7), we introduce also p lagged values in the regression. The degree p of the polynomial $\hat{\beta}_{ih}(L)$ is chosen using the Bayesian information criterion (BIC) in the range from 0 to 6. The so-called DI-AR forecast, however, very slightly outperformed the DI, i.e., $p = 0$, only for CPI at short horizon while having a substantial negative effect on the other variables. As in previous examples in the literature, see Albano (2016); Forni, Giovannelli, et al. (2018), it seems that, once controlled for the factors, the remaining variation is unpredictable. This suggests that the idiosyncratic component is unforecastable and that only the common component can be predicted; a similar result is found in D'Agostino and Giannone (2012) also for the FHLR method.

For FHLR we consider a wide range of parameter specifications. The number q of dynamic factors ranges from 1 to 5 while the number r of static factors from q to 10. The Bartlett lag window for the estimation of the spectral density of the observed series $\hat{\Sigma}^x(\theta)$ is fixed[6] at $B = 35$ to avoid a too-heavy computational cost for a method that is *per se* more elaborated than SW. The dimensionality of the parameters space is already considerable indeed. At each period, for each target variable, 40 different combinations of q and r and 480 projections are considered. The best specification is $q = 3$ and $r = 10$ for IP, $q = 5$ and $r = 5$ for CPI and $q = 3$ and $r = 6$ for UR. In the Appendix the RMSFE for all the specifications is reported in Table A.5. The order of the univariate AR model is decided following the BIC criterion at each period. Tables A.3 and A.5 in the Appendix show that averaging over the horizon the two DFMs clearly outperform the standard AR almost independently of the specification considered. The difference is substantial for IP and UR while more reduced for CPI.

The Diebold-Mariano test leads to the same conclusions. The null of equal performance with AR is rejected for FHLR at the 5% significance level for IP and UR at all the horizons but IP at $h = 12$ and $h = 6$, the latter case is rejected, however, at the 10% significance level. Regarding CPI we can reject the equal performance at the 10% significance level only at $h = 1, 3$. Focusing on SW, instead, we can reject the null at the 10% significance level for IP at $h = 1, 3, 6$ and UR at all horizons. We cannot reject the equal forecasting performance on CPI between AR and SW at any horizon. See Panel A of Table A.9, for more details on the p-values.

Given that the performance of the models on the entire sample can be considerably affected by the Great Recession, we evaluate the models also restricting the attention to the period following the first recession. Even on this subsample, for IP and UR at $h = 1$, both DFMs outperform the AR at the 1% significance level. At $h = 12$, FHLR is more precise than AR at the 5% significance level while SW only at the 10%.

Figures 3.4, 3.5 and 3.6 display at each horizon the target predictions, the cumulative FE^2 and the CSSFED with the benchmark AR of the best specifications for AR, SW and FHLR. These figures

[6] See Albano (2016) and Forni, Giovannelli, et al. (2018) for a detailed analysis of the impact of different window size using Euro Area data.

allow to understand how the predictive performance evolves over time and how the business cycle affects it. The shaded areas represent the recessionary periods according to CEPR.

As expected, we observe a clear jump during the Great Recession for all the variables at almost all the horizons. For the AR, the jump is substantially greater than for the DFMs in the prediction error of IP and UR while the difference with the DFMs is lower for CPI. On the contrary, the EA Debt crisis started in 2011, i.e., the second shaded area, does not dramatically affect the performance of the models. The cumulative FE^2 is quite smooth for the three variables and keeps almost a constant trend after the first recession up to the end of the sample, i.e., January 2018 for the 1-step ahead forecast up to December 2018 for the 12-step ahead.

Let summarize the main results that we obtain by the graphical inspection for the three variables in the following way:

IP. The clear improvement of the DFMs over the AR is driven by the different performances during the Great Recession. The CSSFED of the two-factors model with respect to the AR is indeed almost flat after the first recession. At $h = 1$, however, the DFMs outperform the AR also after 2010.

CPI. The results are more controversial. As stressed before, the large amount of information introduced with the DFMs does not improve forecastability significantly. The cumulative FE^2 of both the two DFMs and AR are remarkably close and change the slope in the sample period several times. This means that no model clearly outperforms the other uniformly on the entire sample. The CSSFED crosses the zero-line and changes the slope at different points over the forecasting period. At the longest horizon, where the CSSFED is smoother, AR performs better after the Great Recession up to 2015 when the performance starts to slightly deteriorate with respect to SW.

UR. Both the DFMs outperform the benchmark AR at all horizons. At the shortest and the longest horizon, the CSSFED is increasing in the entire sample, meaning that the DFMs uniformly outperform the AR and particularly in the last five years. At $h = 3, 6$, instead, the better performance of DMFs is mainly due to the Great Recession, given that after the recession the CSSFED is almost flat as for IP.

This exercise provides evidence that the DFMs perform better during recession periods. One plausible interpretation is that downturn periods are characterized by increased comovements, a situation in which factor-based forecasts are likely to be more accurate.

Moreover, there is also evidence that the DFMs improve the forecasting performance more on real variables rather than nominal ones. This is a frequent result in the literature, see for example Forni, Giovannelli, et al. (2018) where instead relaxing the finite-dimension constraint on the factor space helps the predictability of nominal variables.

Lastly, we want to analyze the effect of the size of the rolling window on the forecasting performance. There is empirical evidence that the predictive performance is sensitive to the size of the window. If a time series is affected by structural breaks, it may be possible that a shorter window size improves the performance because it considers only more recent observations. See, for example, Inoue et al. (2017) for a more detailed presentation. Following this idea, we repeat the same forecasting exercise using a 5-year rolling window. These specifications of the DFMs are called from now on SW-5y and FHLR-5y. The results are presented in Tables A.4 and A.6 and in Figures A.9, A.10 and A.11 in the Appendix. In the last two columns of Table A.9, the p-values of the Diebold-Mariano test against the null of equal performance of the two window sizes are reported.

The main result is that the shorter window, averaging over the horizons, produces worse predictions on the entire sample uniformly for all the specifications and even for the AR model. Comparing the best specifications, the prediction accuracy of the two windows is significantly different at the 10% significance level for SW on IP at $h = 6$, CPI at $h = 1$ and UR at $h = 6, 12$ while for FHLR on IP at $h = 3$, CPI at $h = 3, 6, 12$ and UR at $h = 3, 6, 12$.

The shortest window produces less accurate forecasts during the Recession periods, especially using SW and on CPI. However, focusing only on the post-Great-Recession period the 5-year window still produces significantly worse predictions in the same cases except for CPI with SW and UR with FHLR at $h = 1$ where instead the shortest window is beneficial. The CSSFED of the 5-year with respect to the 10-year window is in many cases monotonically decreasing after the Great Recession for all the variables except for the shortest horizon and CPI. It is worth noting that the better relative performance that we observe for these cases in the last five years reflects exactly the idea that the shorter window allows the predictions to be less influenced by structural breaks that occurred in the past such as the Great Recession.

Moreover, Figures A.9, A.10 and A.11 clearly show that the performance of the shortest window deteriorates with respect to the 10-year window as the horizon increases. This result should not be surprising given that at a longer horizon the shortness of the estimation sample may be a limit. As noted above, if we consider the shortest horizon $h = 1$, the cumulative FEs^2 are close for all the variables and, in some time intervals after the Great Recession, the CSSFED is even increasing. Table A.11 confirms these conclusions. The correlation of the forecasts decreases with the horizon, in particular for FHLR. The predictions tend to be highly correlated for IP and even more for UR. Correlation is lower for CPI.

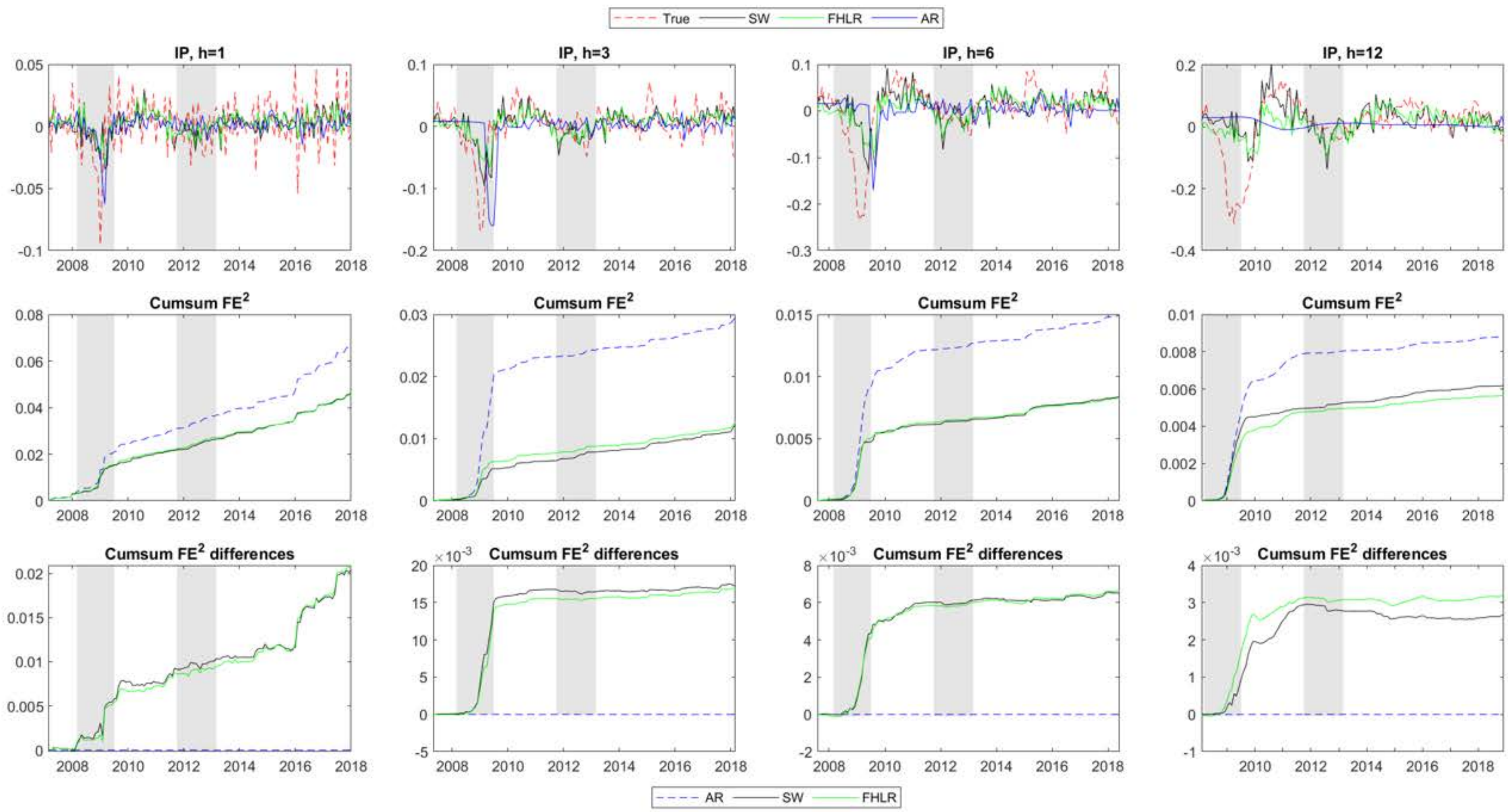

Figure 3.4: Forecasts (panel above), cumulative sum of the squared forecasting errors (FE^2, center panel) and cumulative FE^2 differences with respect to AR (CSSFED, panel below): IP.

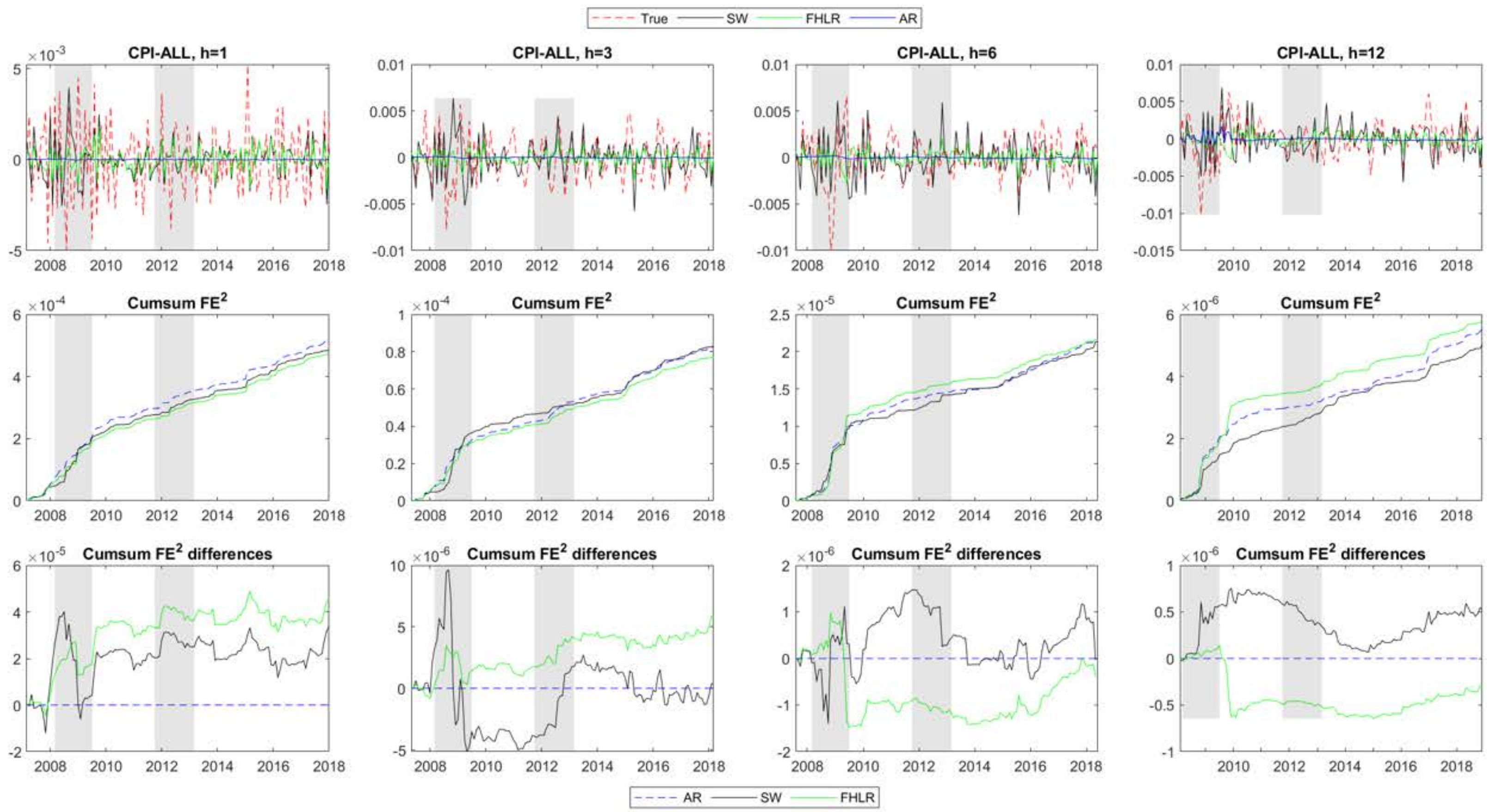

Figure 3.5: Forecasts (panel above), cumulative sum of the squared forecasting errors (FE^2, center panel) and cumulative FE^2 differences with respect to AR (CSSFED, panel below): CPI.

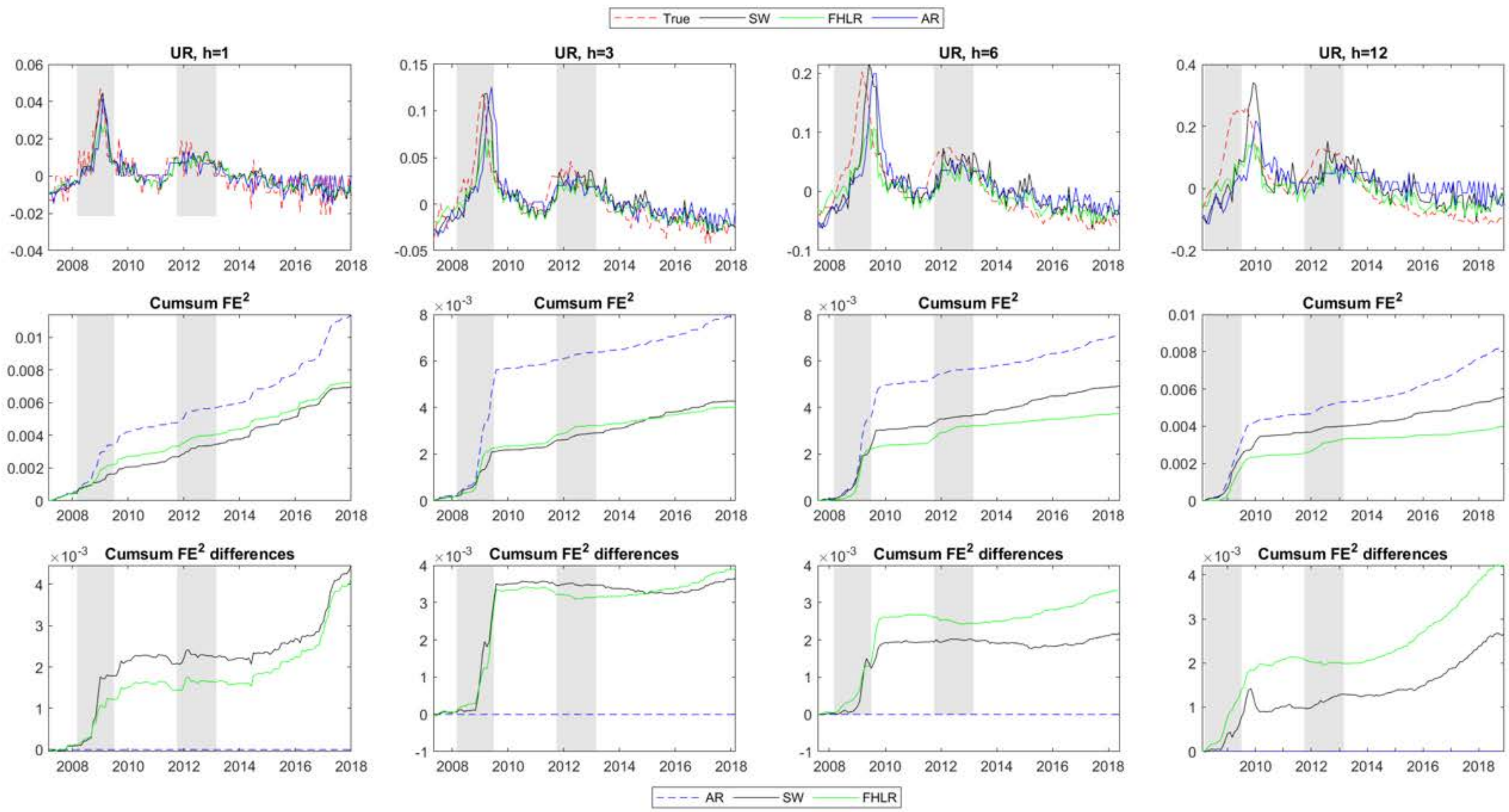

Figure 3.6: Forecasts (panel above), cumulative sum of the squared forecasting errors (FE^2, center panel) and cumulative FE^2 differences with respect to AR (CSSFED, panel below): UR.

3.3.2 SW vs FHLR: Nested models

It is interesting to note that the two models produce almost collinear predictions for IP and UR at all horizons but $h = 12$. The correlation of CPI's forecasts is instead lower and decreasing with the horizon. This is first evidence of the fact that the effect of the frequency-domain approach especially emerges as we increase the horizon. One possible explanation is that FHLR better captures the long-run variation exploiting the dynamic structure of the panel. However, as stressed before, in order to properly compare the two DFMs we nest the two models. In this way, we can isolate the effect of the distinctive features of the two models and find a rationale for dissimilar performance.

This section follows the nesting procedure described in Section 2.3 and evaluates the performance of the two factor models and of the two hybrid models GPC/UNC and PC/CON for the optimal specification of FHLR found in Section 3.3.1. This means $r = 10, 5, 6$ and $q = 3, 5, 3$ respectively for IP, CPI and UR. Even when these are not the best choices for SW, i.e., in the case of IP and CPI[7], however, Table A.3 shows that the RMSFE is close to the best specification available.

Table 3.3 reports the average RMSFE for the four models. Interestingly, the two hybrid models perform slightly better on average at least for IP and UR. Figures 3.7, 3.8 and 3.9 report the cumulative FE^2 at each horizon for the four models and the CSSFED using as benchmark SW. The main results are as follows:

IP. FHLR outperforms SW on 1-step ahead forecast, particularly in the period 2010-2014, due to the generalized principal components. GPC/UNC has the best performance indeed. At the other horizons, instead, the key element is the constrained forecasting equation. It seems that using the covariance of the common component estimated by imposing rank q on the spectral density increases the predictability of IP at longer horizons. The generalized principal components, instead, do not improve the forecasting performance. FHLR is more precise than SW at $h = 6, 12$ while the opposite is true for $h = 3$. At $h = 12$ more accuracy in the predictions of the constrained forecasting equation is evident starting from the second recession.

CPI. At $h = 1$ which model performs better is quite unclear and the results are driven by very small sub-samples. The relative performance changes over time. SW basically outperforms FHLR up to the end of the second recession, in the following period the latter is more accurate. In this case what feature drives the relative performance is not evident.

At longer horizons, the CSSFED is smoother, and the situation is clearer. SW outperforms FHLR at $h = 6, 12$ while the opposite is true for $h = 3$. The relative accuracy is substantially affected by the

[7]Note that here we are also setting $p = 0$, i.e., no lags of the series in the forecasting equation. CPI's best specification includes the lags. However, the effect on the RMSFE of $p = 0$ is marginal.

Great Recession. Even for CPI, on longer horizons, the key element that explains the difference in the performance of the two factors models is the constraint in the forecasting equation. In this case, however, the effect is negative at $h = 6, 12$ even if at $h = 12$ FHLR is slightly catching up with SW after the crisis. This result is coherent with the idea that the structure of a DFM is more appropriate for real variables rather than prices as already noted by the relative performance of the DFMs with respect to the baseline AR model.

UR. The difference between the two models increases with the horizon. At $h = 1$ SW performs better than FHLR while at the other horizons SW is outperformed even if it is more accurate during the second recession. However, in the last five years, the generalized principal components drive the FHLR's better predictions at all the horizons. UR confirms that the most influential feature of FHLR is the constrained forecasting equation. The third column of Table A.9 shows that UR is the only variable for which at $h = 6, 12$ we can reject the null of equal performance on the entire sample at the 10% of significance level. Moreover, confirming the results of Figure 3.9, excluding the Great Recession, FHLR significantly outperforms SW at $h = 1, 3$ at the 5% of significance level.

Therefore, it seems that independently of the variable considered as we increase the horizon as more the role of the constrained rank of the common component spectral density becomes important and drives the different performance of FHLR and SW. This provides evidence of the fact that indeed the structure underlined by a DFM could be a good representation for real variables while the negative contribution of the constrained forecasting equation on inflation predictions confirms that the benefit of using a DFM is limited.

The weighting scheme introduced using the GPC, instead, is more important at the shortest horizon even if the effect seems to be of second order with respect to the constrained forecasting equation. As pointed out by D'Agostino and Giannone (2012), this indicates that simple principal components approximate the factor space quite well and efficiency improvements achieved by weighting for the signal-to-noise ratio do not have a major impact on the forecasting accuracy.

Table 3.3: RMSFE with respect to AR for the best specifications of each model.

	SW	*PC/CON*	*GPC/UNC*	*FHLR*
IP	0.611	0.607	0.602	0.621
CPI	0.918	0.915	0.919	0.923
UR	0.626	0.585	0.585	0.635

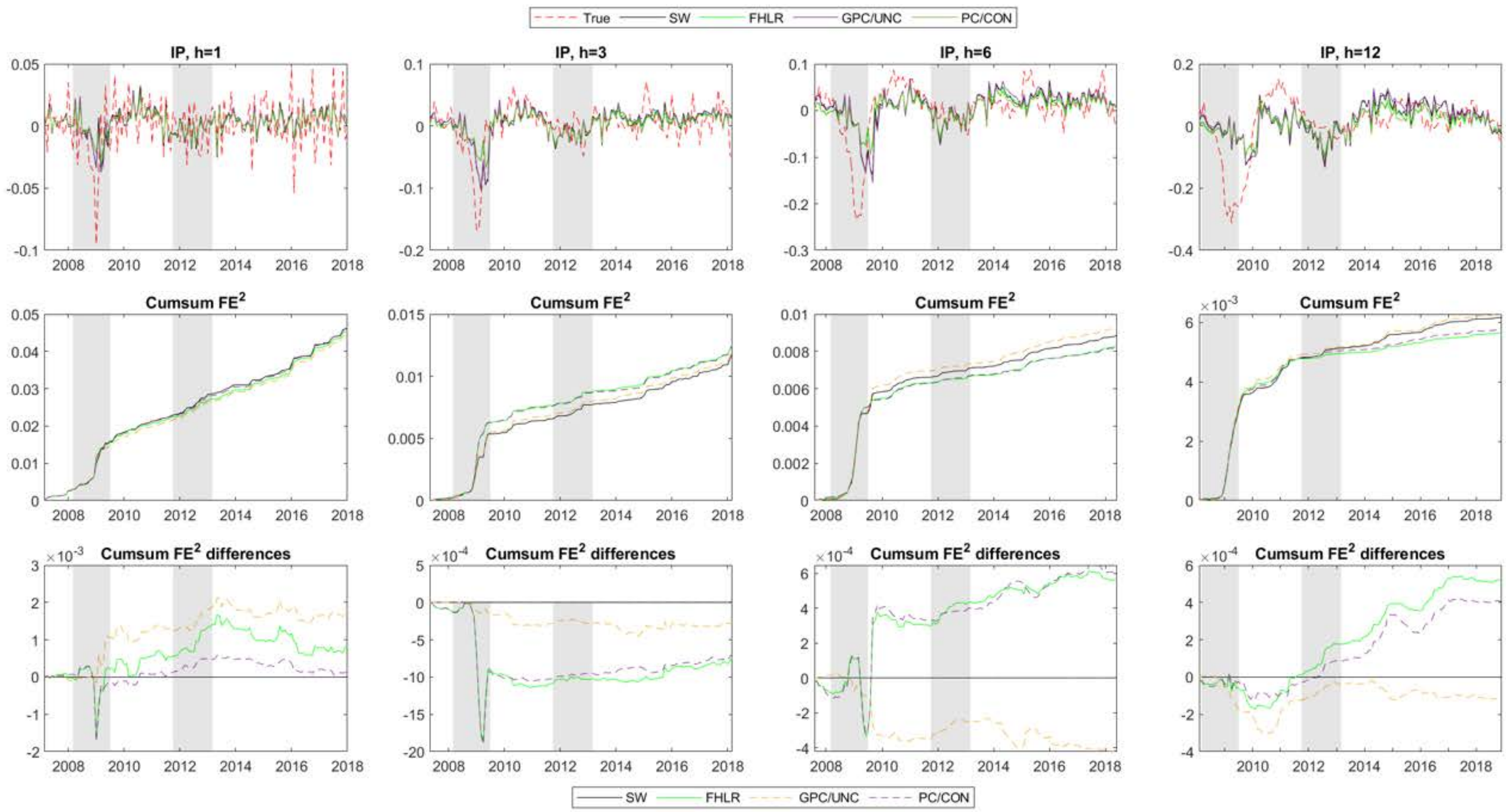

Figure 3.7: Nested Models: Forecasts (panel above), cumulative sum of the squared forecasting errors (FE^2, center panel) and cumulative FE^2 differences with respect to SW (CSSFED, panel below), IP.

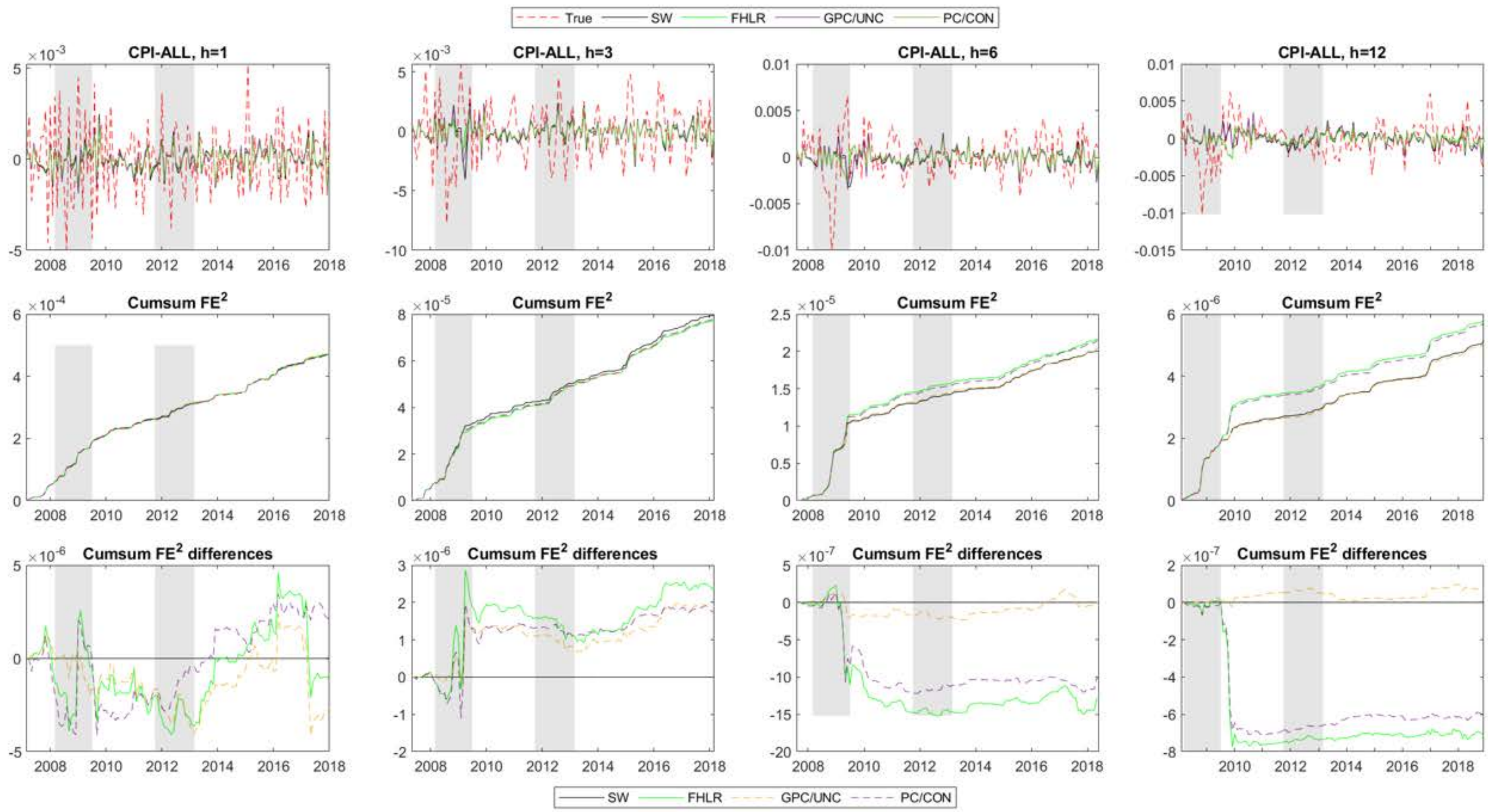

Figure 3.8: Nested Models: Forecasts (panel above), cumulative sum of the squared forecasting errors (FE^2, center panel) and cumulative FE^2 differences with respect to SW (CSSFED, panel below), CPI.

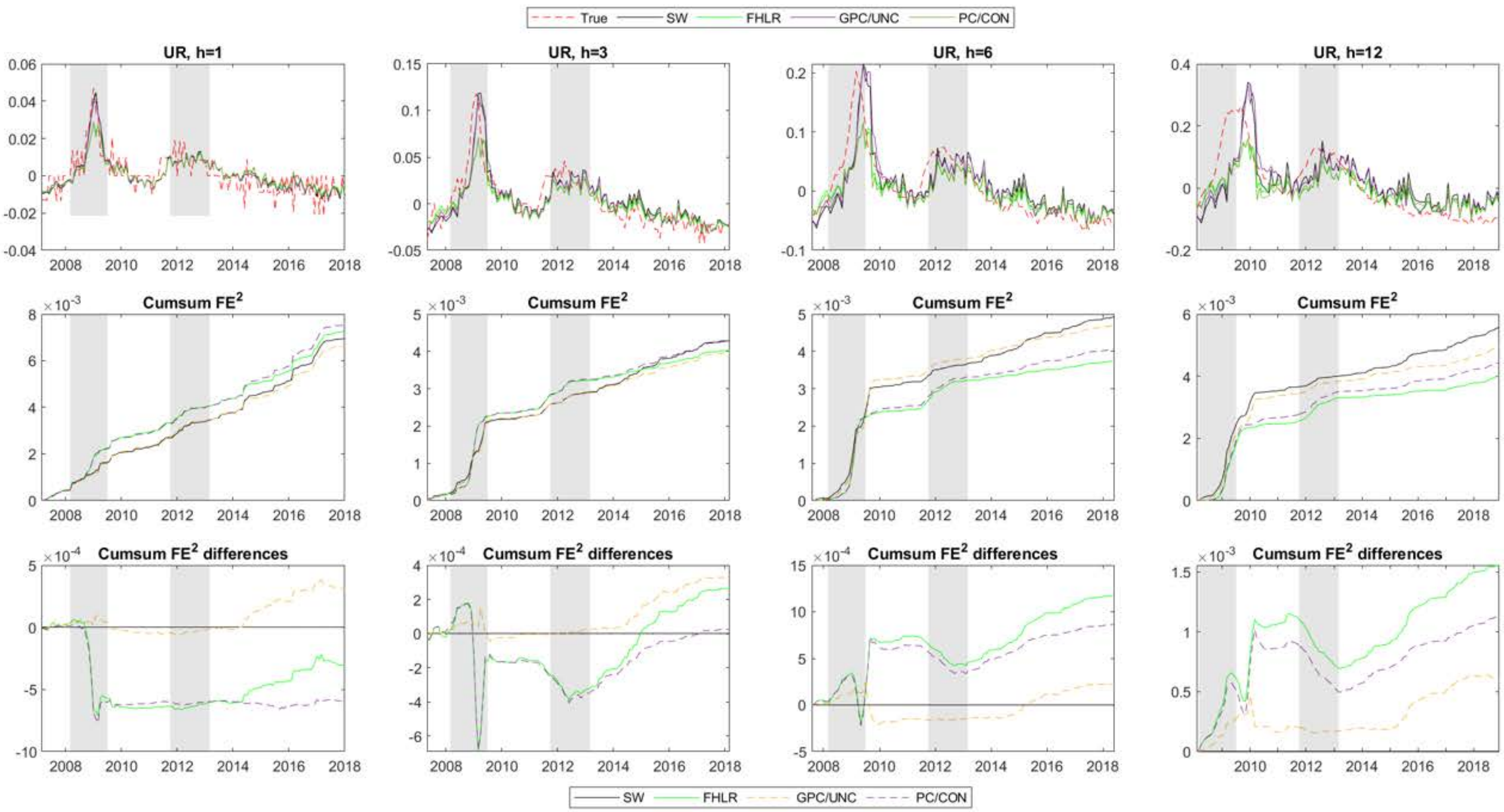

Figure 3.9: Nested Models: Forecasts (panel above), cumulative sum of the squared forecasting errors (FE^2, center panel) and cumulative FE^2 differences with respect to SW (CSSFED, panel below), UR.

3.4 The role of data selection

In the last part of this empirical exercise, we want to focus on how the size and the composition of the dataset affect the factor estimates and, consequently, the forecasting performance of the two factor models.

In the DFMs' setting, Boivin and Ng (2006) shed light on the idea that not all the data improve the forecasting performance of the models. It may also be the case that adding more series to the panel decreases the accuracy of the predictions. This problem can be quite relevant in the framework of factor models where the series to introduce in the large dataset are drawn arbitrarily by the user from a small set of categories as in our case.

At first glance, the asymptotic properties of the estimators suggest that the factor estimates are more efficient the larger the cross-section N. Indeed, a quite simple intuition about the large sample properties of the factor estimates can come from a model with one factor and identical loadings normalized to 1, i.e., $x_{it} = F_t + \xi_{it}$. Aggregating over the panel of cross-sectional dimension we get $\hat{F}_t^N = F_t + \sum_{i=1}^{N} \xi_{it}$; it follows that $\text{Var}(\hat{F}_t^N) = \text{Var}(\sum_{i=1}^{N} \xi_{it})$. Suppose that the ξ_{it} are iid with variance σ^2, then $\text{Var}(\hat{F}_t^N) = \sigma^2/N$ that would decrease with N increasing the efficiency as we increase the dimension of the data.

However, whether more data yield more efficient estimates depend very much on the properties of the additional series. Consider the special case when a researcher unintentionally includes the N series m times. The total number of the series is now mN, but N sets of m idiosyncratic errors are perfectly correlated. When the errors are iid, it can be shown that $\text{Var}(\hat{F}_t^N) = \sigma^2/N$ which is independent of m but depends only on N. The efficiency gain associated with the duplicated data is zero because the second version of the data does not enlarge the information set. Moreover, it can be shown that in cases of series that have errors so strongly correlated with others, adding them can even be detrimental and decrease the efficiency of the factor estimates. So, the intuition is that it is not the amount of data that matters but the new information that each series encompasses.

To get a first idea, let us consider this simple exercise reported in Boivin and Ng (2006). Think of ordering the series within a category by the importance of its common component and putting together a dataset comprising of high-ranked series from each category. Now expand this dataset by adding the lower-ranked, say "noisy" series. Two effects occur simultaneously. The average size of the common component falls as more series are added, and the possibility of correlated errors increases as more series (even more "noisy") from the same category are included. This may lead to a situation where more data might not be desirable.

Therefore, following Boivin and Ng (2006) we have just presented two problems that may lead factors extracted by more series to be less useful for forecasting:

1. The presence of noisy series in the panel.

2. Substantial cross-correlation in the idiosyncratic term.

In the first case, the presence of the noisy series reduces the variation explained by the common component. As we have seen in Section 2 one of the identification assumptions of the DFMs is that the common component should be pervasive. Therefore, including mainly noisy series we implicitly reduce the importance of the common component that is the core element to forecast.

The second problem relates to the assumption of weakly cross-correlated idiosyncratic terms. If the idiosyncratic terms of the series are highly correlated, then it may be possible that the correlation is not so "weak", i.e., the asymptotic assumption on the eigenvalues of the idiosyncratic covariance may be violated.

With these ideas in mind, we consider two new datasets, each of which tries to address one of the two problems. In the next paragraph, we first explain how these new panels are built. Next, we compare the forecasting performance of the two selected datasets to the original one. In this way, we show how the accuracy of the predictions of the two DFMs is affected by the selection of the variables.

3.4.1 Noisy series and cross-correlation in the idiosyncratic term

Let define as the commonality ratio of a series the share of total variance explained by variation in the common component, namely $\text{Var}(\chi_{it})/\text{Var}(x_{it})$. Therefore, we can identify as noisy series those that present a very low level of commonality ratio, i.e., that are mainly explained by the idiosyncratic error.

Figure A.12 presents the distribution of the share of variance accounted by the idiosyncratic term that is, given the orthogonality of the two components, the complementary of the commonality ratio of the series. As the number of dynamic factors increases, the variance left unexplained decreases and the distribution shifts to the left. If we consider $q = 3$ factors, there is no series explained by the idiosyncratic term for more than 80% and almost 75% of the series has a commonality ratio above 50%. To be more precise, the first quartile of the commonality ratio is 0.49. Following Van Nieuwenhuyze (2006), we use the first quartile as a cut-off and drop all the series below. This procedure defines a new dataset that includes the 75% best-explained series and therefore contains 233 variables.

In the rest of the paragraph, we detect with NS the models obtained by the dataset that deals with the presence of too noisy series; hence we can define SW-NS and FHLR-NS as the predictions obtained using SW and FHLR on this dataset.

The second new dataset tries to address the problem of substantial cross-correlation in the idiosyncratic term. The method used to select the series, in this case. is in the spirit (even if slightly modified) of the so-called "Rule 1" introduced by Boivin and Ng (2006). For this reason, we label with Rule1 the predictions of the models estimated using this new dataset.

Let define the highest idiosyncratic cross-correlation of variable j as

$$\tau_j^\star = \max_{i \neq j} |\tau_{ij}| \tag{3.5}$$

where $|\tau_{ij}|$ is the cross-correlation, in modulus, of the idiosyncratic term of the variables i and j.

Figure A.13 plots in a heatmap the correlation matrix of the idiosyncratic error estimated assuming $r = 5$ static factors. Following Boivin and Ng (2006), we estimate the idiosyncratic covariance matrix from the residuals sample covariance using SW approach. In theory, a different solution could be to obtain $\hat{\Gamma}^\xi$ directly from the spectral density but, as mentioned above in Section 2.2, the estimation of the off-diagonal element of $\hat{\Gamma}^\xi$ is ill-conditioned[8]. It is quite evident that our concern about the idiosyncratic cross-correlation could be well-founded. Indeed, there are clusters of cross-correlation off-the-diagonal mainly within the categories. This means that, as likely to be, the correlation among series within the categories may go beyond the correlation induced by the common component. This is also empirical evidence of how unrealistic the assumption of a diagonal idiosyncratic covariance matrix is. Note that, in the framework of approximate factor structure, i.e., under the assumption of weakly correlated idiosyncratic errors, these clusters of highly correlated residuals are not an issue *per se* but only if asymptotically they lead the eigenvalues of the idiosyncratic covariance to infinity. This is clearly untestable, and the finite-sample properties are unclear.

For this reason, Boivin and Ng (2006) using the so-called "Rule 1" build a new dataset removing for each series j the series with cross-correlation of the idiosyncratic term $\tau_j^\star$. In this way, the authors reduce the dataset from 147 series to 71. Applying this rule to the dataset we drop 206, leaving us with 103 series.

We decided to consider a softer version of Rule 1. The procedure works as follows:

1. Sort (in descending order) the series by commonality ratio.

2. Starting from the top, remove sequentially from the list of variables the series with maximum cross-correlation of the idiosyncratic term.

[8] A third possibility could be to use the sample covariance matrix of residuals of FHLR. However, both these two alternatives did not produce substantially different results. This motivates us to focus here on the method already applied in the literature.

Therefore, once a variable has been deleted because it is the residually most cross-correlated with a series with higher commonality, its own most correlated series is not deleted. The intuition behind this modification of Rule1 is that once the problematic series is no longer in the panel there is no need to also remove a series that is no more highly idiosyncratically correlated with a series in the panel.

Using this method the final dataset contains 179 series, so we have 76 more series than resulting from the standard Rule 1.

For SW, the results of the best specifications are presented in Figures 3.10, 3.11 and 3.12 while for FHLR the figures are in the Appendix, see A.14, A.15 and A.16. The RMSFEs for all the possible combinations of parameters of the two models are in Tables A.7 and A.8 in the Appendix.

The first result is that the new datasets produce almost collinear predictions for all the variables at all the horizons except FHLR on the dataset produced by Rule1 where the correlation is around 0.8. This is evidence that the series dropped are globally not particularly informative given that there is not a substantial loss of information. This happens even if the series show globally strong comovement as in our dataset where the first quartile of the commonality ratio is 0.5 and no series has a commonality below 20%.

Focusing on the forecasting performance, the main result is that on the entire sample using the best specification available Rule1 dataset slightly outperformed the baseline for IP and UR using SW while FHLR predictions are almost equally accurate. Regarding the NS dataset, the performance is similar in all the cases but UR using SW where is slightly better.

These two corrections do not help the performance of CPI forecasts. Hence, the worse performance of DFMs on nominal variables cannot be explained by the two concerns exposed in this paragraph and is still an unanswered question.

Dividing by the variable of interest, the results can be summarized in the following way:

IP. The Rule1 dataset performs better during the Great Recession at $h = 6, 12$ but after that CSSFED with respect to SW is almost flat meaning that the two datasets produce almost equally accurate forecasts. For FHLR, instead, the improvement is concentrated in the period between the two recessions and for $h = 12$ we can reject the null of equal performance on the entire sample at the 10% significance level. See Table A.10 for the p-values of the Diebold-Mariano test on the new datasets. NS dataset, instead, outperforms the original dataset only for SW predictions at $h = 3$ at the 10% significance level while producing results similar to the benchmark at all other horizons. NS dataset does not improve FHLR performance which is negatively affected especially in the period between the recessions because, differently from the Rule1 dataset, as noted above, it poorly captures the rebound of IP after the Great Recession.

CPI. The relative performance of the three datasets is strongly influenced by the Great Recession where is very volatile for SW. The three datasets produce not significantly different performance of both SW and FHLR predictors on the entire sample. If we focus on the post-Great-Recession period, instead, the Rule1 dataset outperforms the original dataset using SW on the two longest horizons. At $h = 12$ also NS performs very well but only in the period 2010-2014. Using FHLR prediction, however, the two new datasets improve the performance in the first recession for $h = 3, 6, 12$ while remaining close to the original dataset in the rest of the sample. At $h = 1$ NS is outperformed up to 2015 but in the last three years the CSSFED increases very rapidly.

UR. In this case, for both SW and FHLR, the NS dataset helps predictability in the last four years. Using SW, the same is also true for Rule1 dataset except $h = 1$. This pattern allows to reject the equal performance of the original dataset and NS dataset at the 5% significance level for both SW and FHLR method at $h = 1, 3$. Rule1 dataset, instead, improves the performance only using SW, significantly in the post-recession period at $h = 3, 6$, while it is dominated by the original dataset with the FHLR method.

Table 3.4 recaps the RMSFE for the best specification of the two models using the original dataset and the two created datasets. Therefore, the table shows concisely that looking at the best specification the two new datasets improve the performance more of SW than FHLR and particularly for UR.

Table 3.4: RMSFE with respect to AR for the best specifications of each model.

	SW	*SW-Rule1*	*SW-NS*	*FHLR*	*FHLR-Rule1*	*FHLR-NS*
IP	0.608	0.600	0.610	0.621	0.622	0.624
CPI	0.914	0.920	0.926	0.923	0.917	0.921
UR	0.626	0.605	0.597	0.635	0.643	0.629

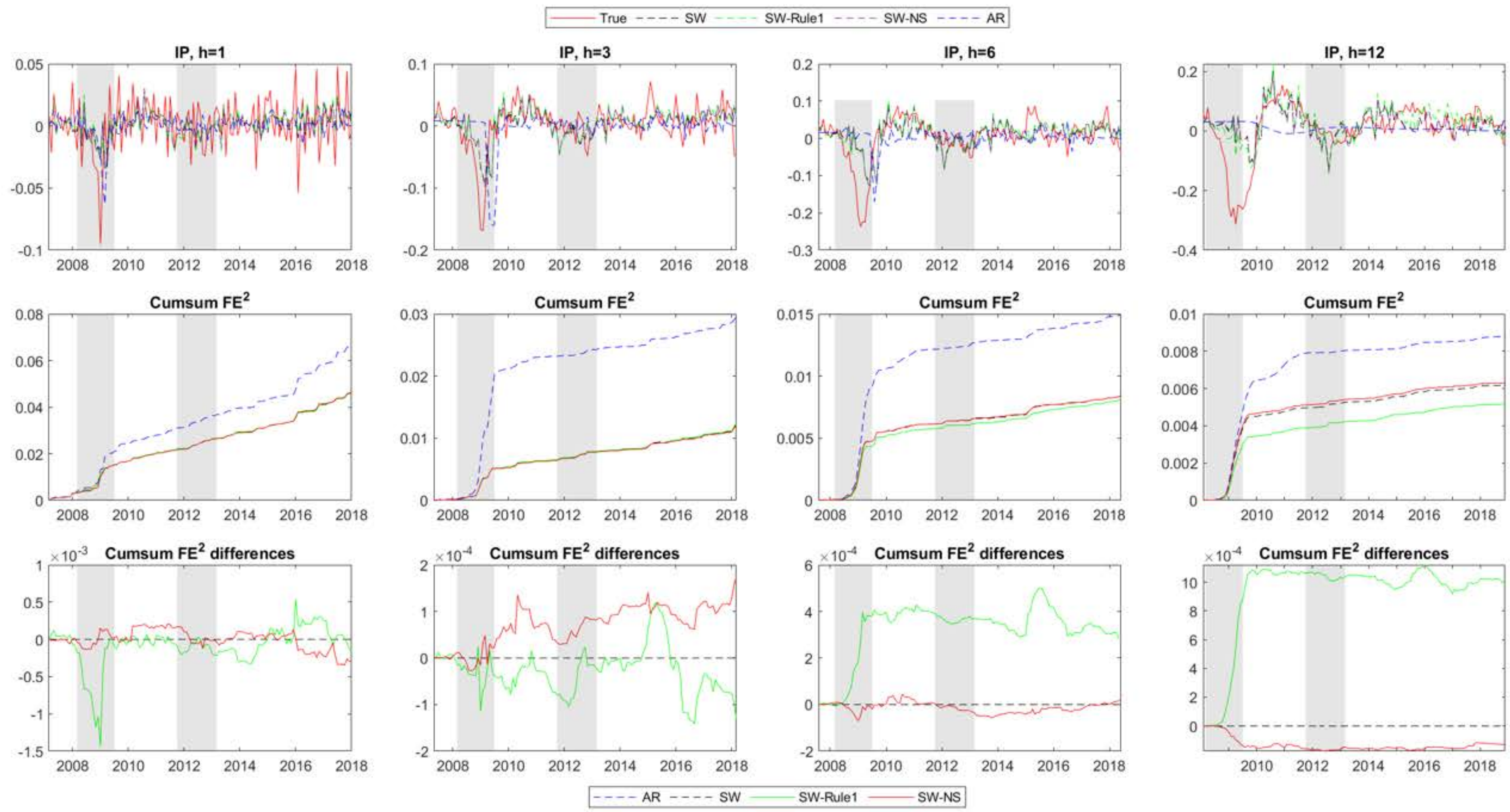

Figure 3.10: Forecasts (panel above) and cumulative sum of the squared forecasting errors (FE^2, panel below) for the two new datasets (NS and Rule1): IP.

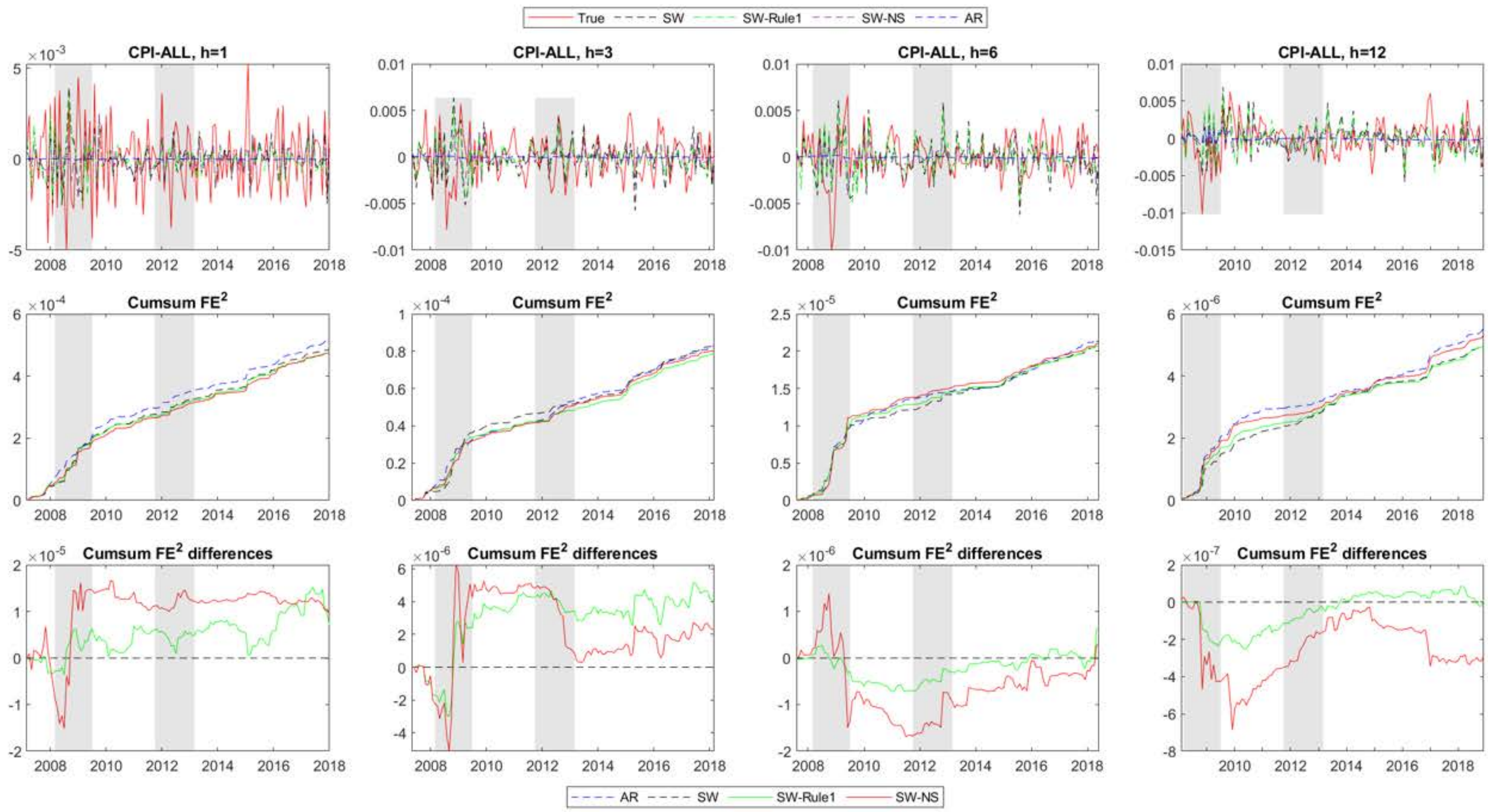

Figure 3.11: Forecasts (panel above) and cumulative sum of the squared forecasting errors (FE^2, panel below) for the two new datasets (NS and Rule1): CPI.

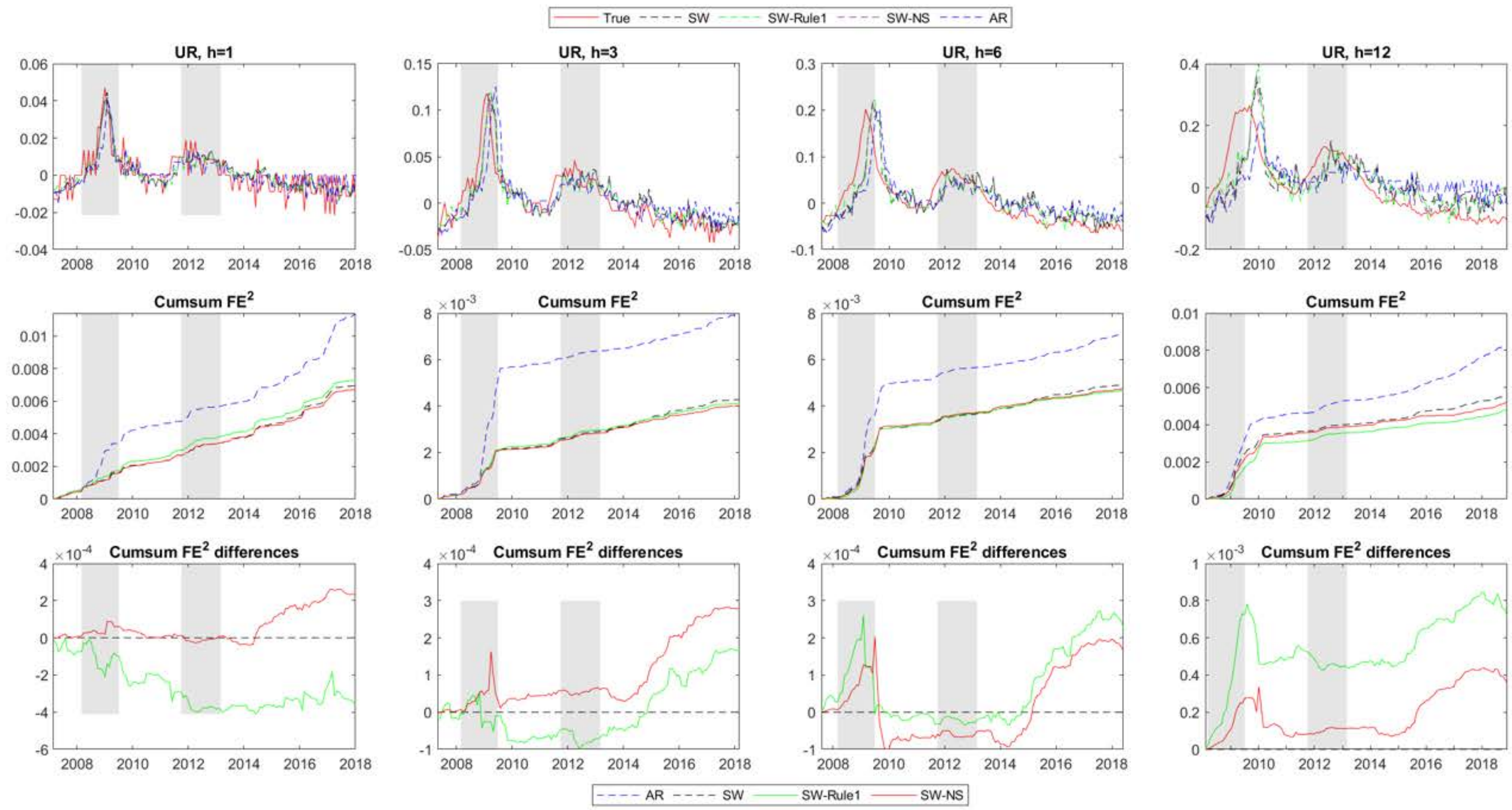

Figure 3.12: Forecasts (panel above) and cumulative sum of the squared forecasting errors (FE^2, panel below) for the two new datasets (NS and Rule1): UR.

4 Conclusions

This paper has compared SW, FHLR and AR in an empirical exercise for EA data including the Great Recession, the Debt Crisis and the following period.

The in-sample analysis of the estimated factors has underlined the importance that the expectational variables have in explaining the comovement of macroeconomic series. Moreover, it has also emerged that the Great Recession has structurally changed the explanatory power of the factors suggesting a structural change in the comovements of the main macroeconomic variables.

The out-of-sample analysis has led to some relevant conclusions about the forecasting ability of the dynamic factor models.

The DFMs outperform the univariate predictions for IP, CPI and UR but more on real variables rather than nominal ones. This result is confirmed by the effect of the constrained forecasting equation as disentangled in the comparison of the two DFMs using the nesting procedure. Embedding the structure of a DFM even in the forecasting equation improves the predictive performance on real variables suggesting that the DFM could be a good representation particularly for real variables. This exercise provides evidence that the DFMs perform better during recession periods. One plausible interpretation is that downturn periods are characterized by strong comovements, which is indeed the key element exploited by factor-based forecasts. Therefore, it should not be surprising that DFMs are in this case more accurate. For this reason, it would be of great interest to evaluate the forecasting performance of the DFMs also on the recent recession caused by the COVID-19 pandemic. As described in Section 1, further research on EA data should also move in the direction of a comparison of the DFMs to other data-intensive models such as penalized regressions, random forest and neural networks.

Regarding the relative performance of SW and FHLR, nesting the models, we observed that independently of the variable considered as we increase the horizon as more the role of the constrained forecasting equation becomes important and drives the different performance of FHLR and SW. The weighting scheme introduced using the generalized principal components, instead, is more important at the shortest horizon even if the effect seems to be of second order. This indicates that standard PCs already provide a good approximation of the space spanned by the factors and the efficient weighting scheme based on the signal-to-noise ratio associated with the generalized PCs does not have a major impact on the forecasting accuracy. Imposing the structure of the DFMs in the forecasting equation improves the predictive performance on real variables. We interpret this result as evidence of the fact that DFM structure can be a particularly good representation for real variables.

The predictive performance of the models is affected by the size and the features of the panel used to estimate the factors. Removing the variables that are mainly idiosyncratic or variables that have too cross-correlated idiosyncratic errors can improve the accuracy of the predictions, particularly for SW predictions and for UR.

References

Albano, D. (2016). Latent factors in large economic datasets: forecasting and data analysis using factor models. An application to the insurance sector. *Doctoral Dissertation, Università degli Studi di Napoli Federico II, Napoli.*

Altissimo, F., Bassanetti, A., Cristadoro, R., Forni, M., Hallin, M., Lippi, M., Reichlin, L., & Veronese, G. (2001). EuroCOIN: a real time coincident indicator of the euro area business cycle. *Available at SSRN 296860.*

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

Boivin, J., & Giannoni, M. (2006). DSGE models in a data-rich environment.

Boivin, J., & Ng, S. (2005). Understanding and comparing factor-based forecasts.

Boivin, J., & Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, *132(1)*, 169–194.

Chamberlain, G., & Rothschild, M. (1983). Arbitrage, factor structure and mean-variance analysis in large asset markets. *Econometrica*, *51*, 1305–1324.

D'Agostino, A., & Giannone, D. (2012). Comparing alternative predictors based on large-panel factor models. *Oxford Bulletin of Economics and Statistics*, *74(2)*, 306–326.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, *13(3)*, 253–263.

Eickmeier, S., & Ziegler, C. (2008). How successful are dynamic factor models at forecasting output and inflation? A meta-analytic approach. *Journal of Forecasting*, *27(3)*, 237–265.

Forni, M., Gambetti, L., Lippi, M., & Sala, L. (2020). Common component structural VARs. *Available at SSRN 3753929.*

Forni, M., Giannone, D., Lippi, M., & Reichlin, L. (2009). Opening the black box: Structural factor models with large cross sections. *Econometric Theory*, *25(05)*, 1319–1347.

Forni, M., Giovannelli, A., Lippi, M., & Soccorsi, S. (2018). Dynamic factor model with infinite-dimensional factor space: Forecasting. *Journal of Applied Econometrics*, *33*, 625–642.

Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2000). The generalized dynamic- factor model: Identification and estimation. *Review of Economics and Statistics*, *82(4)*, 540–554.

Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2001). The generalized dynamic- factor model: Representation theory. *Econometric Theory*, *17*, 1113–1141.

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*, 830–840.

Forni, M., Hallin, M., Lippi, M., & Zaffaroni, P. (2015). Dynamic factor model with infinite-dimensional factor space: Representation. *Journal of Econometrics*, *185*, 359–371.

Forni, M., Hallin, M., Lippi, M., & Zaffaroni, P. (2017). Dynamic factor model with infinite-dimensional factor space: Asymptotic analysis. *Journal of Econometrics*, *199(1)*, 72–92.

Geweke, J. (1977). The dynamic factor analysis of economic time series. *Latent variables in socio-economic models*.

Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, *55*(4), 665–676.

Inoue, A., Jin, L., & Rossi, B. (2017). Rolling window selection for out-of-sample forecasting with time-varying parameters. *Journal of Econometrics*, *196*(1), 55–67.

Kapetanios, G., & Marcellino, M. (2010). Factor-GMM estimation with large sets of possibly weak instruments. *Computational Statistics & Data Analysis*, *54*(11), 2655–2675.

Konzen, E., & Ziegelmann, F. A. (2016). Lasso-type penalties for covariate selection and forecasting in time series. *Journal of Forecasting*, *35*(7), 592–612.

Li, J., & Chen, W. (2014). Forecasting macroeconomic time series: Lasso-based approaches and their forecast combinations with dynamic factor models. *International Journal of Forecasting*, *30*(4), 996–1015. https://doi.org/https://doi.org/10.1016/j.ijforecast.2014.03.016

McCracken, M., & Ng, S. (2015). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, *34(4)*, 574–589.

Medeiros, M. C., Vasconcelos, G. F., Veiga, Á., & Zilberman, E. (2021). Forecasting inflation in a data-rich environment: The benefits of machine learning methods. *Journal of Business & Economic Statistics*, *39*(1), 98–119.

Quah, D., Sargent, T. J., et al. (1993). A dynamic index model for large cross sections. *Business cycles, indicators, and forecasting*, *28*, 285–306.

Sargent, T. J., Sims, C. A., et al. (1977). Business cycle modeling without pretending to have too much a priori economic theory. *New methods in business cycle research*, *1*, 145–168.

Schumacher, C. (2007). Forecasting German GDP using alternative factor models based on large datasets. *Journal of Forecasting*, *26*(4), 271–302.

Stock, J. H., & Watson, M. W. (2011). Dynamic Factor Models. In Oxford Handbook on Economic Forecasting. *eds. Michael P. Clements and David F. Hendry.*, *Oxford University Press*.

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

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

Van Nieuwenhuyze, C. (2006). A Generalised Dynamic Factor Model for the Belgian Economy: Identification of the Business Cycle and GDP Growth Forecasts. *Journal of Business Cycle Measurement and Analysis*, *2(2)*, 213–247.

A Appendix

List of the series

Categories are coded in the following way:

1. Industry & Construction Surveys (ICS),
2. Consumer Survey based Confidence Indicators (CS),
3. Building Permits & Civil Engineering (BP),
4. Industrial Production (IP),
5. Money & Interest Rates (MIR),
6. Harmonized Consumer Price Indices (CPI),
7. Producer Price Index (PPI),
8. Turnover & Retail Sales (TRS)
9. Harmonized Unemployment Rates (UR),
10. Service Surveys (SS).

Let W_t be the raw data. The filter code defines the following transformation function:

$$Y_t = \begin{cases} W_t & \text{if Filter code} = 1 \\ (1-L)W_t & \text{if Filter code} = 2 \\ (1-L)^2 W_t & \text{if Filter code} = 3 \\ \log W_t & \text{if Filter code} = 4 \\ (1-L)\log W_t & \text{if Filter code} = 5 \\ (1-L)^2 \log W_t & \text{if Filter code} = 6 \end{cases}$$

Table A.1: List of series

	Name	Category	Area	Filter code
1	IND SVY: EMPLOYMENT EXPECTATIONS	ICS	Euro Zone	2
2	IND SVY: EXPORT ORDER-BOOK LEVELS	ICS	Euro Zone	2
3	IND SVY: ORDER-BOOK LEVELS	ICS	Euro Zone	2
4	IND SVY: MFG - SELLING PRICE EXPECTATIONS	ICS	Euro Zone	2
5	IND SVY: PRODUCTION EXPECTATIONS	ICS	Euro Zone	2
6	IND SVY: PRODUCTION TREND	ICS	Euro Zone	1
7	IND SVY: MFG - STOCKS OF FINISHED PRODUCTS	ICS	Euro Zone	2
8	CONSTR. SVY: PRICE EXPECTATIONS	ICS	Euro Zone	2

Table A.1 – *Continued from previous page*

	Name	Category	Area	Filter code
9	IND SVY: EXPORT ORDER BOOK POSITION	ICS	Euro Zone	2
10	IND SVY: PRODUCTION TRENDS IN RECENT MTH.	ICS	Euro Zone	1
11	IND SVY: SELLING PRC. EXPECT. MTH. AHEAD	ICS	Euro Zone	2
12	RET. SVY: EMPLOYMENT	ICS	Euro Zone	1
13	RET. SVY: ORDERS PLACED WITH SUPPLIERS	ICS	Euro Zone	1
14	CONSTR. SVY: SYNTHETIC BUS. INDICATOR	ICS	France	2
15	BUS. SVY: CONSTR. SECTOR - CAPACITY UTILISATION RATE	ICS	France	2
16	CONSTR. SVY: ACTIVITY EXPECTATIONS	ICS	France	1
17	CONSTR. SVY: PRICE EXPECTATIONS	ICS	France	1
18	CONSTR. SVY: UNABLE TO INCREASE CAPACITY	ICS	France	2
19	CONSTR. SVY: WORKFORCE CHANGES	ICS	France	1
20	CONSTR. SVY: WORKFORCE FORECAST CHANGES	ICS	France	2
21	SVY: MFG OUTPUT - ORDER BOOK & DEMAND	ICS	France	2
22	SVY: MFG OUTPUT - ORDER BOOK & FOREIGN DEMAND	ICS	France	2
23	SVY: MFG OUTPUT - PERSONAL OUTLOOK	ICS	France	1
24	SVY: AUTO IND - ORDER BOOK & DEMAND	ICS	France	1
25	SVY: AUTO IND - PERSONAL OUTLOOK	ICS	France	1
26	SVY: BASIC & FAB MET PDT EX MACH & EQ - PERSONAL OUTLOOK	ICS	France	1
27	SVY: ELE & ELEC EQ, MACH EQ - ORDER BOOK & DEMAND	ICS	France	2
28	SVY: ELE & ELEC EQ, MACH EQ - ORDER BOOK & FOREIGN DEMAND	ICS	France	1
29	SVY: ELE & ELEC EQ, MACH EQ - PERSONAL OUTLOOK	ICS	France	1
30	SVY: MFG OUTPUT - PRICE OUTLOOK	ICS	France	1
31	SVY: MFG OF CHEMICALS & CHEMICAL PDT - ORDER BOOK & DEMAND	ICS	France	1
32	SVY: MFG OF CHEMICALS & CHEMICAL PDT - PERSONAL OUTLOOK	ICS	France	1
33	SVY: MFG OF FOOD PR & BEVERAGES - ORDER BOOK & DEMAND	ICS	France	1
34	SVY: MFG OF FOOD PR & BEVERAGES - ORDER BOOK & FOREIGN DEMAND	ICS	France	1
35	SVY: MFG OF TRSP EQ - FINISHED GOODS INVENTORIES	ICS	France	1
36	SVY: MFG OF TRSP EQ - ORDER BOOK & DEMAND	ICS	France	1
37	SVY: MFG OF TRSP EQ - ORDER BOOK & FOREIGN DEMAND	ICS	France	1
38	SVY: MFG OF TRSP EQ - PERSONAL OUTLOOK	ICS	France	1
39	SVY: OTH MFG, MACH & EQ RPR & INSTAL - ORD BOOK & DEMAND	ICS	France	1
40	SVY: OTH MFG, MACH & EQ RPR & INSTAL - ORD BOOK & FGN DEMAND	ICS	France	1
41	SVY: OTH MFG, MACH & EQ RPR & INSTAL - PERSONAL OUTLOOK	ICS	France	1
42	SVY: OTHER MFG - ORDER BOOK & DEMAND	ICS	France	2
43	SVY: RUBBER, PLASTIC & NON MET PDT - ORDER BOOK & DEMAND	ICS	France	2
44	SVY: RUBBER, PLASTIC & NON MET PDT - ORDER BOOK & FGN DEMAND	ICS	France	1
45	SVY: RUBBER, PLASTIC & NON MET PDT - PERSONAL OUTLOOK	ICS	France	1
46	SVY: TOTAL IND - ORDER BOOK & DEMAND	ICS	France	2
47	SVY: TOTAL IND - ORDER BOOK & FOREIGN DEMAND	ICS	France	2
48	SVY: TOTAL IND - PERSONAL OUTLOOK	ICS	France	1
49	SVY: TOTAL IND - PRICE OUTLOOK	ICS	France	1
50	SVY: WOOD & PAPER, PRINT & MEDIA - ORD BOOK & FGN DEMAND	ICS	France	2
51	TRD. &IND: BUS SIT	ICS	Germany	2
52	TRD. &IND: BUS EXPECT IN 6MO	ICS	Germany	2
53	TRD. &IND: BUS SIT	ICS	Germany	2
54	TRD. &IND: BUS CLIMATE	ICS	Germany	2
55	CNSTR IND: BUS CLIMATE	ICS	Germany	2
56	MFG: BUS CLIMATE	ICS	Germany	2
57	MFG: BUS CLIMATE	ICS	Germany	2
58	MFG CONS GDS: BUS CLIMATE	ICS	Germany	2
59	MFG (EXCL FBT): BUS CLIMATE	ICS	Germany	2
60	WHSLE (INCL MV): BUS CLIMATE	ICS	Germany	2
61	MFG: BUS SIT	ICS	Germany	2
62	MFG: BUS SIT	ICS	Germany	2
63	MFG (EXCL FBT): BUS SIT	ICS	Germany	2
64	MFG (EXCL FBT): BUS SIT	ICS	Germany	2
65	CNSTR IND: BUS EXPECT IN 6MO	ICS	Germany	2
66	CNSTR IND: BUS EXPECT IN 6MO	ICS	Germany	2
67	MFG: BUS EXPECT IN 6MO	ICS	Germany	2
68	MFG: BUS EXPECT IN 6MO	ICS	Germany	2
69	MFG CONS GDS: BUS EXPECT IN 6MO	ICS	Germany	2
70	MFG (EXCL FBT): BUS EXPECT IN 6MO	ICS	Germany	2
71	MFG (EXCL FBT): BUS EXPECT IN 6MO	ICS	Germany	2
72	RT (INCL MV): BUS EXPECT IN 6MO	ICS	Germany	2
73	WHSLE (INCL MV): BUS EXPECT IN 6MO	ICS	Germany	2
74	BUS. CONF. INDICATOR	ICS	Italy	2

Table A.1 – *Continued from previous page*

	Name	Category	Area	Filter code
75	ORDER BOOK LEVEL: IND	ICS	Spain	2
76	ORDER BOOK LEVEL: FOREIGN - IND	ICS	Spain	1
77	ORDER BOOK LEVEL: INVESTMENT GOODS	ICS	Spain	1
78	ORDER BOOK LEVEL: INT. GOODS	ICS	Spain	1
79	PRODUCTION LEVEL - IND	ICS	Spain	1
80	CONS. CONFIDENCE INDICATOR	CS	Euro Zone	2
81	CONS. SVY: ECONOMIC SITUATION LAST 12 MTH. - EMU 11/12	CS	Euro Zone	2
82	CONS. SVY: POSSIBLE SAVINGS OPINION	CS	France	1
83	CONS. SVY: FUTURE FINANCIAL SITUATION	CS	France	2
84	SVY - HOUSEHOLDS, ECONOMIC SITUATION NEXT 12M	CS	France	2
85	CONS. CONFIDENCE INDICATOR - GERMANY	CS	Germany	2
86	CONS. CONFIDENCE INDEX	CS	Germany	5
87	GFK CONS. CLIMATE SVY - BUS. CYCLE EXPECTATIONS	CS	Germany	1
88	CONS.S CONFIDENCE INDEX	CS	Germany	5
89	CONS. CONFIDENCE CLIMATE (BALANCE)	CS	Germany	2
90	CONS. SVY: ECONOMIC CLIMATE INDEX (N.WEST IT)	CS	Italy	5
91	CONS. SVY: ECONOMIC CLIMATE INDEX (SOUTHERN IT)	CS	Italy	5
92	CONS. SVY: GENERAL ECONOMIC SITUATION (BALANCE)	CS	Italy	2
93	CONS. SVY: PRICES IN NEXT 12 MTHS. - LOWER	CS	Italy	5
94	CONS. SVY: UNEMPLOYMENT EXPECTATIONS (BALANCE)	CS	Italy	2
95	CONS. SVY: UNEMPLOYMENT EXPECTATIONS - APPROX. SAME	CS	Italy	5
96	CONS. SVY: UNEMPLOYMENT EXPECTATIONS - LARGE INCREASE	CS	Italy	5
97	CONS. SVY: UNEMPLOYMENT EXPECTATIONS - SMALL INCREASE	CS	Italy	5
98	CONS. SVY: GENERAL ECONOMIC SITUATION (BALANCE)	CS	Italy	2
99	CONS. SVY: HOUSEHOLD BUDGET - DEPOSITS TO/WITHDRAWALS	CS	Spain	5
100	CONS. SVY: HOUSEHOLD ECONOMY (CPY) - MUCH WORSE	BP	France	5
101	CONS. SVY: ITALIAN ECON.IN NEXT 12 MTHS.- MUCH WORSE	BP	France	5
102	CONS. SVY: MAJOR PURCHASE INTENTIONS - BALANCE	BP	France	2
103	CONS. SVY: MAJOR PURCHASE INTENTIONS - MUCH LESS	BP	France	5
104	CONS. SVY: HOUSEHOLDS FIN SITUATION - BALANCE	BP	France	2
105	INDL. PROD. - EXCLUDING CONSTR.	IP	Euro Zone	5
106	INDL. PROD. - CAP. GOODS	IP	Euro Zone	5
107	INDL. PROD. - CONS. NON-DURABLES	IP	Euro Zone	5
108	INDL. PROD. - CONS. DURABLES	IP	Euro Zone	5
109	INDL. PROD. - CONS. GOODS	IP	Euro Zone	5
110	INDL. PROD.	IP	France	5
111	INDL. PROD. - MFG	IP	France	5
112	INDL. PROD. - MFG (2010=100)	IP	France	5
113	INDL. PROD. - MANUF. OF MOTOR VEHICLES, TRAILERS, SEMITRAILERS	IP	France	5
114	INDL. PROD. - INT. GOODS	IP	France	5
115	INDL. PROD. - INDL. PROD. - CONSTR.	IP	France	5
116	INDL. PROD. - MANUF. OF WOOD AND PAPER PRODUCTS	IP	France	5
117	INDL. PROD. - MANUF. OF COMPUTER, ELECTRONIC AND OPTICAL PROD	IP	France	5
118	INDL. PROD. - MANUF. OF ELECTRICAL EQUIPMENT	IP	France	5
119	INDL. PROD. - MANUF. OF MACHINERY AND EQUIPMENT	IP	France	5
120	INDL. PROD. - MANUF. OF TRANSPORT EQUIPMENT	IP	France	5
121	INDL. PROD. - OTHER MFG	IP	France	5
122	INDL. PROD. - MANUF. OF CHEMICALS AND CHEMICAL PRODUCTS	IP	France	5
123	INDL. PROD. - MANUF. OF RUBBER AND PLASTICS PRODUCTS	IP	France	5
124	INDL. PROD. - INVESTMENT GOODS	IP	Italy	5
125	INDL. PROD.	IP	Italy	5
126	INDL. PROD.	IP	Italy	5
127	INDL. PROD. - CONS. GOODS - DURABLE	IP	Italy	5
128	INDL. PROD. - INVESTMENT GOODS	IP	Italy	5
129	INDL. PROD. - INT. GOODS	IP	Italy	5
130	INDL. PROD. - CHEMICAL PRODUCTS & SYNTHETIC FIBRES	IP	Italy	5
131	INDL. PROD. - MACHINES & MECHANICAL APPARATUS	IP	Italy	5
132	INDL. PROD. - MEANS OF TRANSPORT	IP	Italy	5
133	INDL. PROD. - METAL & METAL PRODUCTS	IP	Italy	5
134	INDL. PROD. - RUBBER ITEMS & PLASTIC MATERIALS	IP	Italy	5
135	INDL. PROD. - WOOD & WOOD PRODUCTS	IP	Italy	5
136	INDL. PROD.	IP	Italy	5
137	INDL. PROD. - COMPUTER, ELECTRONIC AND OPTICAL PRODUCTS	IP	Italy	5
138	INDL. PROD. - BASIC PHARMACEUTICAL PRODUCTS	IP	Italy	5
139	INDL. PROD. - CONSTR. (SA)	IP	Germany	5
140	INDL. PROD. - IND INCL CNSTR	IP	Germany	5

Table A.1 – *Continued from previous page*

	Name	Category	Area	Filter code
141	INDL. PROD. - MFG	IP	Germany	5
142	INDL. PROD. - REBASED TO 1975=100	IP	Germany	5
143	INDL. PROD. - CHEMS & CHEM PRDS	IP	Germany	5
144	INDL. PROD. - IND EXCL CNSTR	IP	Germany	5
145	INDL. PROD. - IND EXCL ENERGY & CNSTR	IP	Germany	5
146	INDL. PROD. - MINING & QUAR	IP	Germany	5
147	INDL. PROD. - CMPTR, ELECCL & OPT PRDS, ELECL EQP	IP	Germany	5
148	INDL. PROD. - INTERM GOODS	IP	Germany	5
149	INDL. PROD. - CAP. GOODS	IP	Germany	5
150	INDL. PROD. - DURABLE CONS GOODS	IP	Germany	5
151	INDL. PROD. - TEX & WEARING APPAREL	IP	Germany	5
152	INDL. PROD. - PULP, PAPER&PRDS, PUBSHG&PRINT	IP	Germany	5
153	INDL. PROD. - CHEM PRDS	IP	Germany	5
154	INDL. PROD. - RUB&PLAST PRDS	IP	Germany	5
155	INDL. PROD. - BASIC MTLS	IP	Germany	5
156	INDL. PROD. - CMPTR, ELECCL & OPT PRDS, ELECL EQP	IP	Germany	5
157	INDL. PROD. - MOTOR VEHICLES, TRAILERS&SEMI TRAIL	IP	Germany	5
158	INDL. PROD. - TEX & WEARING APPAREL	IP	Germany	5
159	INDL. PROD. - PAPER & PRDS, PRINT, REPROD OF RECRD MEDIA	IP	Germany	5
160	INDL. PROD. - CHEMS & CHEM PRDS	IP	Germany	5
161	INDL. PROD. - BASIC MTLS, FAB MTL PRDS, EXCL MACH&EQP	IP	Germany	5
162	INDL. PROD. - REPAIR & INSTALL OF MACH & EQP	IP	Germany	5
163	INDL. PROD. - MFG EXCL CNSTR & FBT	IP	Germany	5
164	INDL. PROD. - MINING & IND EXCL FBT	IP	Germany	5
165	INDL. PROD. - IND EXCL FBT	IP	Germany	5
166	INDL. PROD. - INTERM & CAP. GOODS	IP	Germany	5
167	INDL. PROD. - FAB MTL PRDS EXCL MACH & EQP	IP	Spain	5
168	INDL. PROD.	IP	Spain	5
169	INDL. PROD. - CONS. GOODS	IP	Spain	5
170	INDL. PROD. - CAP. GOODS	IP	Spain	5
171	INDL. PROD. - INT. GOODS	IP	Spain	5
172	INDL. PROD. - ENERGY	IP	Spain	5
173	INDL. PROD. - CONS. GOODS, NON-DURABLES	IP	Spain	5
174	INDL. PROD. - MINING	IP	Spain	5
175	INDL. PROD. - MFG IND	IP	Spain	5
176	INDL. PROD. - OTHER MINING & QUARRYING	IP	Spain	5
177	INDL. PROD. - TEXTILE	IP	Spain	5
178	INDL. PROD. - CHEMICALS & CHEMICAL PRODUCTS	IP	Spain	5
179	INDL. PROD. - PLASTIC & RUBBER PRODUCTS	IP	Spain	5
180	INDL. PROD. - OTHER NON-METAL MINERAL PRODUCTS	IP	Spain	5
181	INDL. PROD. - METAL PROCESSING IND	IP	Spain	5
182	INDL. PROD. - METAL PRODUCTS EXCL. MACHINERY	IP	Spain	5
183	INDL. PROD. - ELECTRICAL EQUIPMENT	IP	Spain	5
184	INDL. PROD. - AUTOMOBILE	IP	Spain	5
185	EURO INTERBANK OFFERED RATE - 3-MONTH (MEAN)	MIR	Euro Zone	5
186	MONEY SUPPLY: LOANS TO OTHER EA RESIDENTS EXCL. GOVT.	MIR	Euro Zone	5
187	MONEY SUPPLY: M3	MIR	Euro Zone	5
188	EURO SHORT TERM REPO RATE	MIR	France	5
189	DATASTREAM EURO SHARE PRICE INDEX (MTH. AVG.)	MIR	France	2
190	EURIBOR: 3-MONTH (MTH. AVG.)	MIR	France	5
191	MFI LOANS TO RESIDENT PRIVATE SECTOR	MIR	France	5
192	MONEY SUPPLY - M1	MIR	France	5
193	MONEY SUPPLY - M3	MIR	France	5
194	SHARE PRICE INDEX - SBF 250	MIR	Germany	2
195	FIBOR - 3 MONTH (MTH.AVG.)	MIR	Germany	5
196	MONEY SUPPLY - M3	MIR	Germany	5
197	MONEY SUPPLY - M2	MIR	Germany	5
198	BANK PRIME LENDING RATE / ECB MARGINAL LENDING FACILITY	MIR	Germany	5
199	DAX SHARE PRICE INDEX, EP	MIR	Italy	1
200	INTERBANK DEPOSIT RATE-AVERAGE ON 3-MONTHS DEPOSITS	MIR	Italy	5
201	OFFICIAL RESERVE ASSETS	MIR	Spain	5
202	MONEY SUPPLY: M3 - SPANISH	MIR	Spain	5
203	MADRID S.E - GENERAL INDEX	MIR	Spain	5
204	HICP - Overall index, Index	CPI	Euro Zone	6
205	HICP - All-items excluding energy, Index	CPI	Euro Zone	6
206	HICP - Food incl. alcohol and tobacco, Index	CPI	Euro Zone	6

Table A.1 – *Continued from previous page*

	Name	Category	Area	Filter code
207	HICP - Processed food incl. alcohol and tobacco, Index	CPI	Euro Zone	6
208	HICP - Unprocessed food, Index	CPI	Euro Zone	6
209	HICP - Goods, Index	CPI	Euro Zone	6
210	HICP - Industrial goods, Index	CPI	Euro Zone	6
211	HICP - Industrial goods excluding energy, Index	CPI	Euro Zone	6
212	HICP - Services, Index	CPI	Euro Zone	6
213	HICP - All-items excluding tobacco, Index	CPI	Euro Zone	6
214	HICP - All-items excluding energy and food, Index	CPI	Euro Zone	6
215	HICP - All-items excluding energy and unprocessed food, Index	CPI	Euro Zone	6
216	All-items HICP	CPI	Germany	6
217	All-items HICP	CPI	Spain	6
218	All-items HICP	CPI	France	6
219	All-items HICP	CPI	Italy	6
220	Goods (overall index excluding services)	CPI	Germany	6
221	Goods (overall index excluding services)	CPI	France	6
222	Processed food including alcohol and tobacco	CPI	Germany	6
223	Processed food including alcohol and tobacco	CPI	Spain	6
224	Processed food including alcohol and tobacco	CPI	France	6
225	Processed food including alcohol and tobacco	CPI	Italy	6
226	Unprocessed food	CPI	Germany	6
227	Unprocessed food	CPI	Spain	6
228	Unprocessed food	CPI	France	6
229	Unprocessed food	CPI	Italy	6
230	Non-energy industrial goods	CPI	Germany	6
231	Non-energy industrial goods	CPI	France	6
232	Services (overall index excluding goods)	CPI	Germany	6
233	Services (overall index excluding goods)	CPI	France	6
234	Overall index excluding tobacco	CPI	Germany	6
235	Overall index excluding tobacco	CPI	France	6
236	Overall index excluding ENERGY	CPI	Germany	6
237	Overall index excluding ENERGY	CPI	France	6
238	Overall index excluding energy and unprocessed food	CPI	Germany	6
239	Overall index excluding energy and unprocessed food	CPI	France	6
240	PPI: IND EXCLUDING CONSTR. & ENERGY	PPI	Euro Zone	6
241	PPI: CAP. GOODS	PPI	Euro Zone	6
242	PPI: NON-DURABLE CONS. GOODS	PPI	Euro Zone	6
243	PPI: INT. GOODS	PPI	Euro Zone	6
244	PPI: NON DOM. - MINING, MFG & QUARRYING	PPI	Euro Zone	6
245	PPI: NON DOM. MFG	PPI	Germany	6
246	PPI: INT. GOODS EXCLUDING ENERGY	PPI	Germany	6
247	PPI: CAP. GOODS	PPI	Germany	6
248	PPI: CONS. GOODS	PPI	Germany	6
249	PPI: FUEL	PPI	Germany	6
250	PPI: INDL. PRODUCTS (EXCL. ENERGY)	PPI	Germany	6
251	PPI: MACHINERY	PPI	Germany	6
252	DEFLATED T/O: RET. SALE IN NON-SPCLD STR WITH FOOD, BEV & TOB	TRS	Germany	5
253	DEFLATED T/O: OTH RET. SALE IN NON-SPCLD STR	TRS	Germany	5
254	DEFLATED T/O: SALE OF MOTOR VEHICLE PTS & ACCES	TRS	Germany	5
255	DEFLATED T/O: WHOLESALE OF AGL RAW MATLS & LIVE ANIMALS	TRS	Germany	5
256	DEFLATED T/O: WHOLESALE OF HOUSEHOLD GOODS	TRS	Italy	5
257	T/O: RET. TRD, EXC OF MV , MOTORCYLES & FUEL	TRS	Spain	5
258	T/O: RET. SALE OF CLTH & LEATH GDS IN SPCLD STR	TRS	Spain	5
259	T/O: RET. SALE OF NON-FOOD PRDS (EXC FUEL)	TRS	Spain	5
260	T/O: RET. SALE OF INFO, HOUSEHLD & REC EQP IN SPCLD STR	TRS	Spain	5
261	EK UNEMPLOYMENT: ALL	UR	Euro Zone	5
262	EK UNEMPLOYMENT: PERSONS OVER 25 YEARS OLD	UR	Euro Zone	5
263	EK UNEMPLOYMENT: WOMEN UNDER 25 YEARS OLD	UR	Euro Zone	5
264	EK UNEMPLOYMENT: WOMEN OVER 25 YEARS OLD	UR	Euro Zone	5
265	EK UNEMPLOYMENT: MEN OVER 25 YEARS OLD	UR	Euro Zone	5
266	FR HUR ALL PERSONS (ALL AGES)	UR	France	5
267	FR HUR FEMMES (AGES 15-24)	UR	France	5
268	FR HUR FEMMES (ALL AGES)	UR	France	5
269	FR HUR HOMMES (AGES 15-24)	UR	France	5
270	FR HUR HOMMES (ALL AGES)	UR	France	5
271	FR HUR ALL PERSONS (AGES 15-24)	UR	France	5
272	FR HURALL PERSONS(AGES 25 AND OVER)	UR	France	5

Table A.1 – *Continued from previous page*

	Name	Category	Area	Filter code
273	FR HUR FEMALES (AGES 25 AND OVER)	UR	France	5
274	FR HUR MALES (AGES 25 AND OVER)	UR	France	5
275	BD HUR ALL PERSONS (ALL AGES)	UR	Germany	5
276	BD HUR FEMMES (AGES 15-24)	UR	Germany	5
277	BD HUR FEMMES (ALL AGES)	UR	Germany	5
278	BD HUR HOMMES (AGES 15-24)	UR	Germany	5
279	BD HUR HOMMES (ALL AGES)	UR	Germany	5
280	BD HUR ALL PERSONS (AGES 15-24)	UR	Germany	5
281	BD HURALL PERSONS(AGES 25 AND OVER)	UR	Germany	5
282	BD HUR FEMALES (AGES 25 AND OVER)	UR	Germany	5
283	BD HUR MALES (AGES 25 AND OVER)	UR	Germany	5
284	IT HUR ALL PERSONS (ALL AGES)	UR	Italy	5
285	IT HUR FEMMES (ALL AGES)	UR	Italy	5
286	IT HUR HOMMES (ALL AGES)	UR	Italy	5
287	IT HUR ALL PERSONS (AGES 15-24)	UR	Italy	5
288	IT HURALL PERSONS(AGES 25 AND OVER)	UR	Italy	5
289	ES HUR ALL PERSONS (ALL AGES)	UR	Spain	5
290	ES HUR FEMMES (AGES 16-24)	UR	Spain	5
291	ES HUR FEMMES (ALL AGES)	UR	Spain	5
292	ES HUR HOMMES (AGES 16-24)	UR	Spain	5
293	ES HUR HOMMES (ALL AGES)	UR	Spain	5
294	ES HUR ALL PERSONS (AGES 16-24)	UR	Spain	5
295	ES HURALL PERSONS(AGES 25 AND OVER)	UR	Spain	5
296	ES HUR FEMALES (AGES 25 AND OVER)	UR	Spain	5
297	ES HUR MALES (AGES 25 AND OVER)	UR	Spain	5
298	DE - Service Confidence Indicator	SS	Germany	2
299	DE Services - Buss. Dev. Past 3 months	SS	Germany	1
300	DE Services - Evol. Demand Past 3 months	SS	Germany	2
301	DE Services - Exp. Demand Next 3 months	SS	Germany	1
302	DE Services - Evol. Employ. Past 3 months	SS	Germany	2
303	FR - Service Confidence Indicator	SS	France	2
304	FR Services - Buss. Dev. Past 3 months	SS	France	1
305	FR Services - Evol. Demand Past 3 months	SS	France	1
306	FR Services - Exp. Demand Next 3 months	SS	France	1
307	FR Services - Evol. Employ. Past 3 months	SS	France	1
308	FR Services - Exp. Employ. Next 3 months	SS	France	1
309	FR Services - Exp. Prices Next 3 months	SS	France	1

In-sample analysis

In this section we report the figures of the exercise in Section 3.2 for the squared loadings on the factors. See Figures A.3 and A.4. The loadings are defined in the following way:

SW. From (2.5) the loadings of the variables on the factor k are defined as the k-th eigenvector, i.e., the k-th column of the $N \times r$ matrix V_r, divided by the square root of N.

FHLR. From (2.11) the loadings of the variables on the factor k, obtained using the generalized principal components, are defined as the k-th generalized eigenvector, i.e., the k-th column of the $N \times r$ matrix V_{rg}.

Table A.2: Explanatory power Factors 1-5 FHLR

$R_i^2(1)$	Variable	$R_i^2(2)$	Variable	$R_i^2(3)$	Variable	$R_i^2(4)$	Variable	$R_i^2(5)$	Variable
0.779	ICS-23	0.410	IP-109	0.274	ICS-56	0.262	ICS-30	0.516	CPI-218
0.778	ICS-48	0.397	ICS-56	0.268	IP-148	0.261	ICS-49	0.479	CPI-235
0.742	ICS-45	0.367	IP-147	0.265	IP-163	0.183	UR-289	0.460	CPI-204
0.731	ICS-6	0.358	IP-139	0.260	IP-164	0.170	IP-167	0.440	CPI-221
0.716	ICS-26	0.358	IP-141	0.257	ICS-59	0.166	UR-295	0.437	CPI-220
$R^2(1) =$	0.133	$R^2(2) =$	0.087	$R^2(3) =$	0.053	$R^2(4) =$	0.031	$R^2(5) =$	0.034

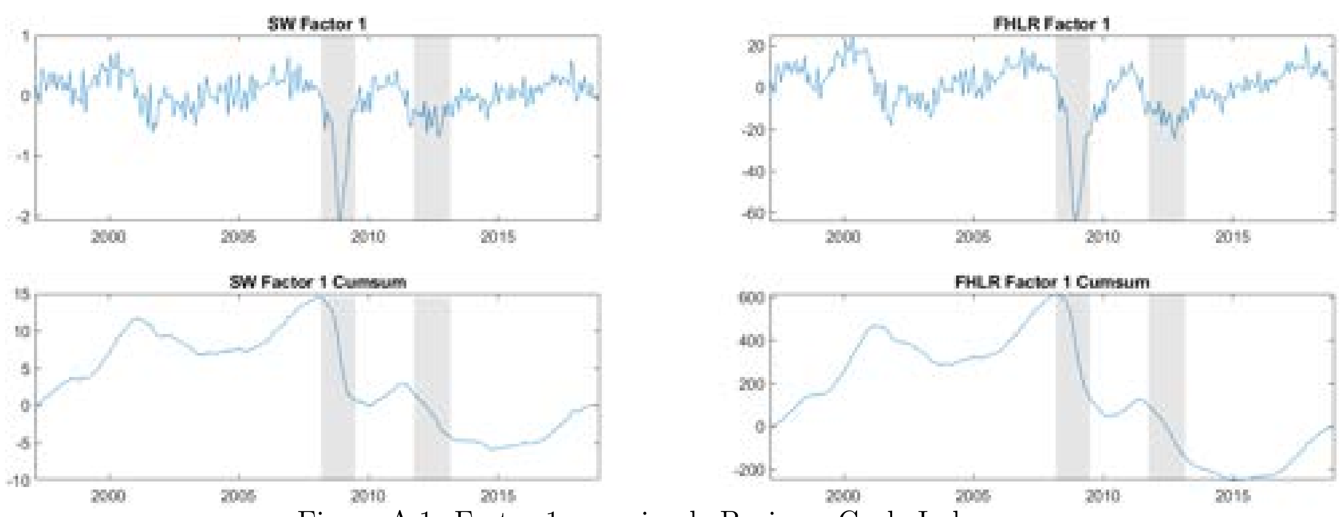

Figure A.1: Factor 1 as a simple Business Cycle Index

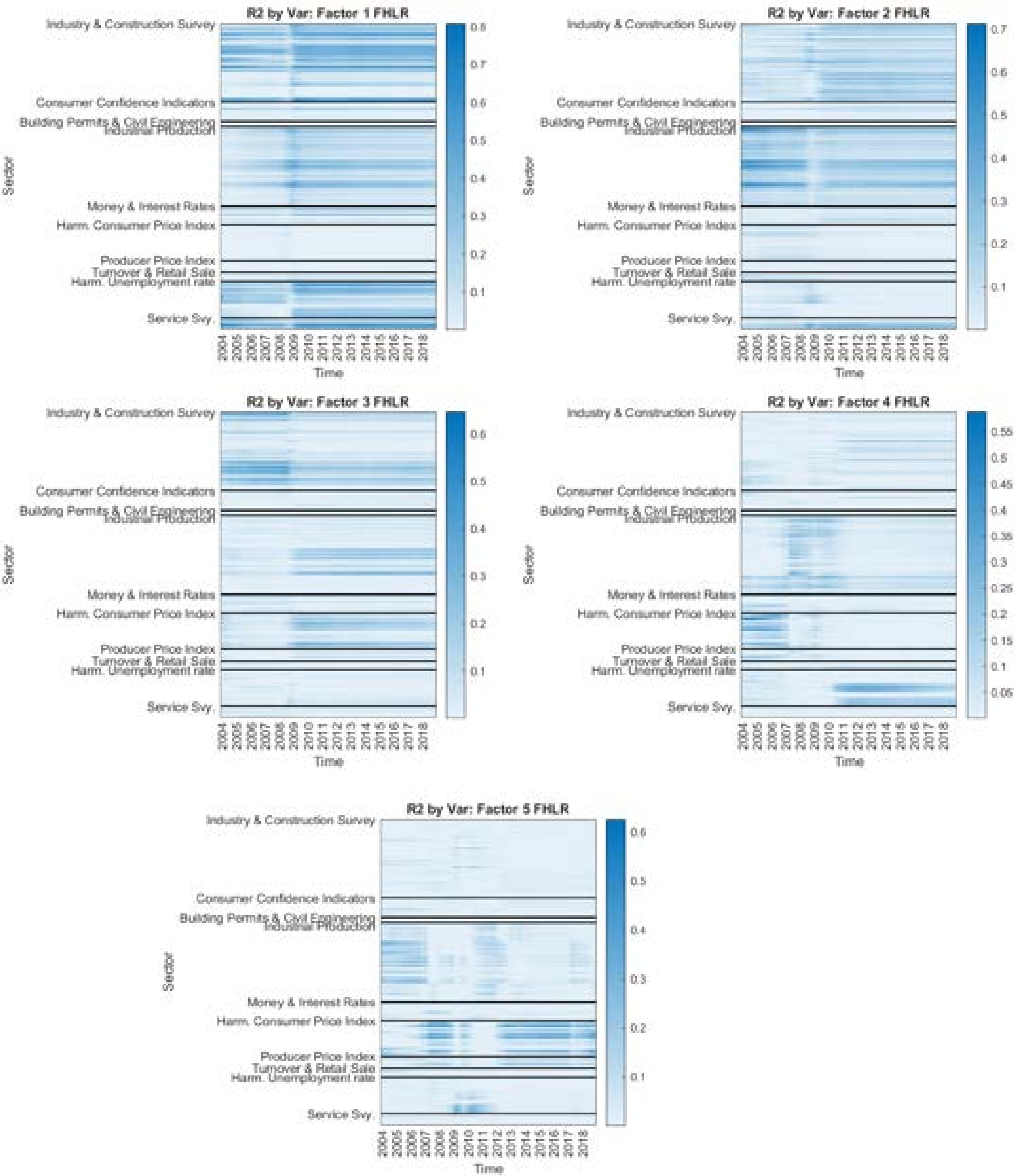

Figure A.2: $R^2_{it}(k),\ k = 1, \ldots, 5$, FHLR Factors

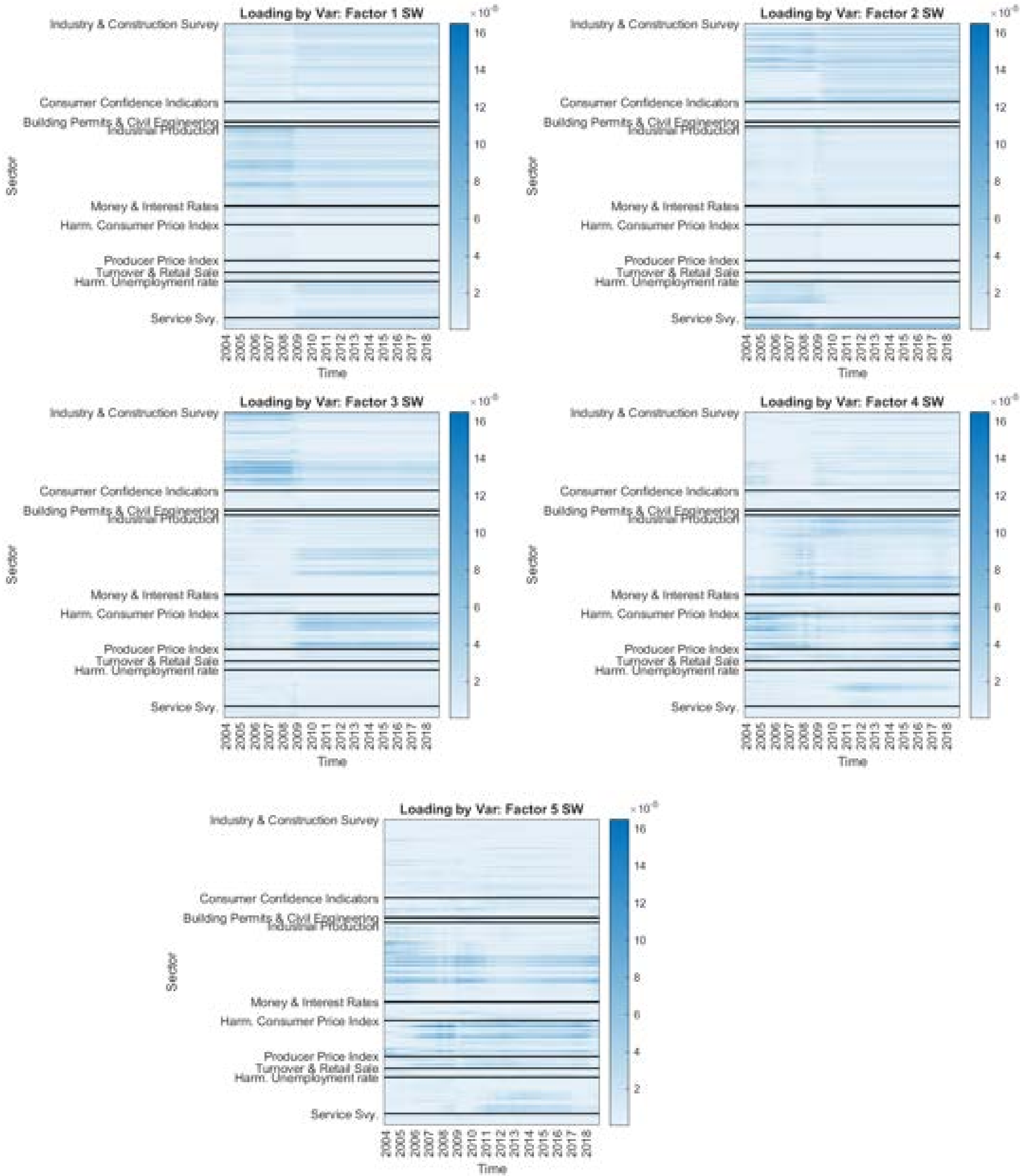

Figure A.3: Loadings on the k-th factor, $k = 1, \ldots, 5$: SW Factors

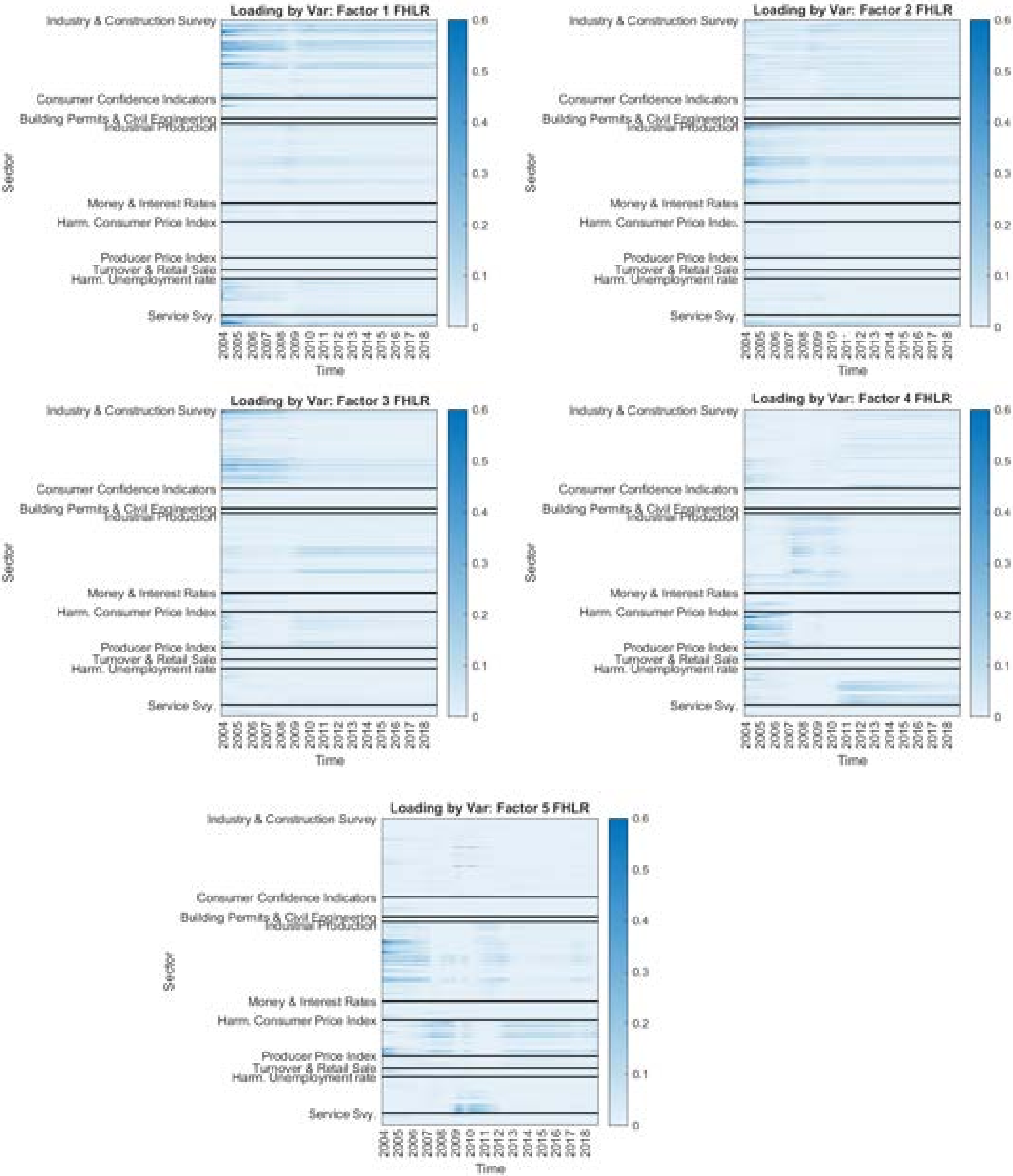

Figure A.4: Loadings on the k-th factor, $k = 1, \ldots, 5$: FHLR Factors

Forecasting exercise and the role of data selection

RMSFE Tables

Table A.3: RMSFE SW vs AR

$p > 0$	r=1	2	3	4	5	6	7	8	9	10
IP	0.845	0.841	0.674	0.658	0.656	0.663	0.691	0.727	0.730	0.734
CPI	1.013	1.011	0.987	0.931	**0.914**	1.019	1.011	1.022	1.044	1.058
UR	0.901	0.814	0.786	0.771	0.756	0.704	0.738	0.718	0.74	0.758
$p = 0$	r=1	2	3	4	5	6	7	8	9	10
IP	0.807	0.779	0.628	0.610	**0.608**	0.613	0.631	0.622	0.614	0.611
CPI	1.028	1.028	1.008	0.978	0.918	0.961	0.971	0.977	0.991	0.989
UR	0.883	0.949	0.895	0.884	0.830	**0.626**	0.629	0.632	0.639	0.647

Note: $h = 1, 3, 6, 12$ average RMSFE of SW with respect to AR with optimal order given by BIC. Panel above includes lagged of the series in the forecasting equation, $p > 0$, panel below not.

Table A.4: RMSFE SW-5y vs AR

$p > 0$	r=1	2	3	4	5	6	7	8	9	10	AR-5y
IP	1.034	0.874	0.787	0.772	0.841	0.865	0.878	0.873	0.911	0.991	1.511
CPI	1.124	1.115	**1.084**	1.232	1.263	1.229	1.339	1.336	1.339	1.500	1.005
UR	0.987	1.108	0.998	1.000	1.150	1.000	0.928	0.957	0.955	0.970	1.362
$p = 0$	r=1	2	3	4	5	6	7	8	9	10	AR-5y
IP	0.849	0.742	0.685	**0.673**	0.727	0.734	0.737	0.740	0.753	0.781	1.511
CPI	1.125	1.111	1.090	1.145	1.179	1.103	1.158	1.185	1.194	1.254	1.005
UR	1.040	1.207	1.086	1.027	1.074	1.000	0.909	0.918	**0.903**	0.943	1.362

Note: $h = 1, 3, 6, 12$ average RMSFE of SW with respect to AR with optimal order given by BIC. SW-5y is estimated using a 5-year rolling window, AR with a 10-year one; in this way, this table can be compared to Table A.3. Panel above includes lagged of the series in the forecasting equation, $p > 0$, panel below not.

Table A.5: RMSFE FHLR vs AR

IP	r =1	2	3	4	5	6	7	8	9	10
q =1	0.825	0.808	0.664	0.657	0.657	0.647	0.650	0.652	0.654	0.657
2	-	0.789	0.654	0.660	0.663	0.653	0.654	0.648	0.640	0.642
3	-	-	0.638	0.640	0.640	0.634	0.638	0.625	0.624	**0.621**
4	-	-	-	0.645	0.644	0.643	0.651	0.639	0.642	0.636
5	-	-	-	-	0.634	0.64	0.647	0.639	0.641	0.626
CPI	r =1	2	3	4	5	6	7	8	9	10
q =1	1.024	1.024	1.003	0.985	0.924	0.933	0.942	0.947	0.948	0.956
2	-	1.035	1.021	1.012	0.967	0.942	0.948	0.959	0.967	0.996
3	-	-	1.025	1.020	0.971	0.935	0.94	0.937	0.939	0.965
4	-	-	-	0.998	**0.923**	0.932	0.945	0.950	0.947	0.974
5	-	-	-	-	**0.923**	0.939	0.952	0.964	0.958	0.972
UR	r =1	2	3	4	5	6	7	8	9	10
q =1	1.007	0.979	0.939	0.919	0.901	0.877	0.88	0.881	0.885	0.878
2	-	0.946	0.884	0.805	0.79	0.670	0.685	0.716	0.722	0.731
3	-	-	0.880	0.805	0.789	**0.635**	0.661	0.686	0.694	0.691
4	-	-	-	0.808	0.806	0.640	0.667	0.696	0.701	0.707
5	-	-	-	-	0.824	0.642	0.667	0.693	0.707	0.712

Note: $h = 1, 3, 6, 12$ average RMSFE of FHLR with respect to AR with optimal order given by BIC.

Table A.6: RMSFE FHLR-5y vs AR

IP	r =1	2	3	4	5	6	7	8	9	10
q =1	0.840	0.747	0.689	0.677	0.671	0.673	0.683	0.679	0.698	0.699
2	-	0.738	0.675	0.671	0.658	0.663	0.669	0.684	0.686	0.707
3	-	-	0.742	0.715	0.708	0.699	0.706	0.715	0.707	0.714
4	-	-	-	**0.646**	0.653	0.660	0.671	0.682	0.701	0.727
5	-	-	-	-	0.708	0.708	0.700	0.699	0.727	0.743
CPI	r =1	2	3	4	5	6	7	8	9	10
q =1	1.066	1.063	1.037	1.041	1.025	1.002	0.986	1.012	1.004	1.014
2	-	1.060	1.047	1.067	1.040	0.999	0.997	1.010	1.032	1.017
3	-	-	1.104	1.066	1.100	1.082	1.100	1.063	1.048	1.029
4	-	-	-	1.089	1.102	1.045	1.021	1.037	1.038	1.123
5	-	-	-	-	1.065	1.030	**0.972**	1.027	1.030	1.038
UR	r =1	2	3	4	5	6	7	8	9	10
q =1	1.080	1.075	1.019	0.968	0.911	0.894	0.903	0.906	0.958	0.976
2	-	1.058	0.966	0.896	0.869	0.843	**0.828**	0.834	0.861	0.896
3	-	-	1.014	1.000	1.016	0.963	0.940	0.925	0.921	0.940
4	-	-	-	0.852	0.835	0.887	0.873	0.900	0.916	0.926
5	-	-	-	-	0.908	0.907	0.900	0.908	0.924	0.933

Note: $h = 1, 3, 6, 12$ average RMSFE of FHLR-5y with respect to AR with optimal order given by BIC. FHLR-5y is estimated using a 5-year rolling window, AR with a 10-year one; in this way, this table can be compared to Table A.5.

Table A.7: RMSFE FHLR-NS and SW-NS vs AR

IP	r =1	2	3	4	5	6	7	8	9	10
q =1	0.828	0.823	0.666	0.662	0.658	0.656	0.652	0.654	0.664	0.674
2	-	0.806	0.653	0.658	0.664	0.656	0.655	0.647	0.639	0.641
3	-	-	0.643	0.643	0.652	0.638	0.639	0.628	**0.624**	0.630
4	-	-	-	0.65	0.649	0.644	0.643	0.636	0.642	0.647
5	-	-	-	-	0.639	0.642	0.641	0.637	0.638	0.641
SW-NS	0.811	0.788	0.627	0.615	**0.610**	0.626	0.631	0.626	0.628	0.637
CPI	r =1	2	3	4	5	6	7	8	9	10
q =1	1.027	1.027	1.008	0.985	**0.921**	0.936	0.935	0.954	0.941	0.936
2	-	1.037	1.029	1.017	0.949	0.945	0.961	0.975	0.957	0.957
3	-	-	1.033	1.028	0.931	0.945	0.961	0.963	0.969	0.945
4	-	-	-	0.992	0.923	0.946	0.968	0.975	0.967	0.955
5	-	-	-	-	0.927	0.954	0.974	0.988	0.992	0.963
SW-NS	1.031	1.033	1.016	1.020	**0.926**	0.967	0.988	1.001	0.996	0.981
UR	r =1	2	3	4	5	6	7	8	9	10
q =1	1.008	0.991	0.951	0.929	0.899	0.897	0.898	0.899	0.888	0.891
2	-	0.971	0.901	0.811	0.805	0.665	0.679	0.726	0.737	0.738
3	-	-	0.902	0.812	0.796	**0.629**	0.665	0.703	0.705	0.710
4	-	-	-	0.828	0.811	0.631	0.660	0.713	0.712	0.719
5	-	-	-	-	0.819	0.631	0.651	0.707	0.708	0.720
SW-NS	0.895	0.968	0.910	0.863	0.828	**0.597**	0.600	0.666	0.657	0.668

Note: $h = 1, 3, 6, 12$ average RMSFE of SW-NS, FHLR-NS, SW-Rule1 and FHLR-Rule1 with respect to AR with optimal order given by BIC. SW-NS and FHLR-NS are estimated removing the first quartile of the data for variance explained by the idiosyncratic term as described in Section 3.4.1.

Table A.8: RMSFE FHLR-Rule1 and SW-Rule1 vs AR

IP	r =1	2	3	4	5	6	7	8	9	10
q =1	0.821	0.795	0.656	0.651	0.650	0.645	0.633	0.638	0.645	0.654
2	-	0.783	0.646	0.648	0.655	0.643	0.637	0.640	0.645	0.646
3	-	-	0.631	0.632	0.639	0.631	0.632	0.628	0.629	**0.622**
4	-	-	-	0.640	0.646	0.63	0.628	0.634	0.641	0.633
5	-	-	-	-	0.642	0.625	0.625	0.638	0.637	0.625
SW-Rule1	0.810	0.770	0.624	0.619	0.610	**0.600**	0.602	0.618	0.617	0.614
CPI	r =1	2	3	4	5	6	7	8	9	10
q =1	1.024	1.017	0.988	0.963	0.936	0.923	0.926	**0.917**	0.941	0.938
2	-	1.028	1.005	0.962	0.939	0.934	0.94	0.943	0.935	0.958
3	-	-	1.008	0.971	0.927	0.937	0.931	0.93	0.926	0.93
4	-	-	-	0.956	0.921	0.933	0.936	0.938	0.938	0.938
5	-	-	-	-	0.922	0.944	0.938	0.936	0.944	0.954
SW-Rule1	1.024	1.013	0.988	0.953	**0.920**	0.945	0.951	0.951	0.954	0.964
UR	r =1	2	3	4	5	6	7	8	9	10
q =1	1.023	1.008	0.975	0.971	0.947	0.929	0.891	0.879	0.876	0.9
2	-	0.971	0.916	0.848	0.844	0.73	0.692	0.704	0.711	0.717
3	-	-	0.902	0.833	0.831	**0.643**	0.661	0.664	0.682	0.692
4	-	-	-	0.849	0.835	0.655	0.668	0.672	0.686	0.716
5	-	-	-	-	0.842	0.659	0.661	0.672	0.679	0.717
SW-Rule1	0.927	0.957	0.905	0.905	0.888	0.652	0.644	0.609	0.609	**0.605**

Note: $h = 1, 3, 6, 12$ average RMSFE of SW-NS, FHLR-NS, SW-Rule1 and FHLR-Rule1 with respect to AR with optimal order given by BIC. SW-Rule1 and FHLR-Rule1 are estimated removing for each series the series with the most correlated idiosyncratic error as described in Section 3.4.1.

Table A.9: Diebold-Mariano test: p-values.

Panel A: Full sample (2007:3-2018:12)

IP	*AR vs SW*	*AR vs FHLR*	*SW vs FHLR*	*SW vs SW-5y*	*FHLR vs FHLR-5y*
h=1	0.000	0.000	0.388	0.550	0.459
3	0.051	0.044	0.580	0.739	0.939
6	0.080	0.072	0.380	0.932	0.818
12	0.117	0.117	0.277	0.897	0.778
CPI	*AR vs SW*	*AR vs FHLR*	*SW vs FHLR*	*SW vs SW-5y*	*FHLR vs FHLR-5y*
h=1	0.232	0.061	0.333	0.929	0.604
3	0.488	0.097	0.317	0.895	0.969
6	0.501	0.560	0.559	0.838	0.973
12	0.248	0.693	0.800	0.843	0.982
UR	*AR vs SW*	*AR vs FHLR*	*SW vs FHLR*	*SW vs SW-5y*	*FHLR vs FHLR-5y*
h=1	0.000	0.000	0.747	0.869	0.559
3	0.059	0.041	0.392	0.872	0.991
6	0.073	0.044	0.031	0.935	0.980
12	0.030	0.033	0.067	0.917	0.960

Panel B: Post Great Recession (2010:1-2018:12)

IP	*AR vs SW*	*AR vs FHLR*	*SW vs FHLR*	*SW vs SW-5y*	*FHLR vs FHLR-5y*
h=1	0.002	0.000	0.140	0.602	0.517
3	0.072	0.005	0.061	0.607	0.881
6	0.054	0.024	0.432	0.969	0.717
12	0.240	0.150	0.645	0.827	0.445
CPI	*AR vs SW*	*AR vs FHLR*	*SW vs FHLR*	*SW vs SW-5y*	*FHLR vs FHLR-5y*
h=1	0.298	0.272	0.536	0.648	0.549
3	0.155	0.056	0.583	0.810	0.980
6	0.435	0.143	0.297	0.393	0.989
12	0.661	0.139	0.090	0.390	0.981
UR	*AR vs SW*	*AR vs FHLR*	*SW vs FHLR*	*SW vs SW-5y*	*FHLR vs FHLR-5y*
h=1	0.001	0.000	0.018	0.349	0.096
3	0.280	0.049	0.043	0.650	0.983
6	0.146	0.046	0.128	0.784	0.988
12	0.088	0.048	0.111	0.829	0.976

Table A.10: Data selection: Diebold-Mariano test: p-values.

Panel A: Full sample (2007:3-2018:12)

IP	*SW vs SW-NS*	*SW vs SW-Rule1*	*FHLR vs FHLR-NS*	*FHLR vs FHLR-Rule1*
h=1	0.700	0.542	0.428	0.683
3	0.062	0.690	0.733	0.909
6	0.397	0.263	0.889	0.138
12	0.797	0.176	0.92	0.094
CPI	*SW vs SW-NS*	*SW vs SW-Rule1*	*FHLR vs FHLR-NS*	*FHLR vs FHLR-Rule1*
h=1	0.357	0.325	0.461	0.497
3	0.415	0.262	0.529	0.216
6	0.434	0.274	0.388	0.224
12	0.702	0.477	0.313	0.204
UR	*SW vs SW-NS*	*SW vs SW-Rule1*	*FHLR vs FHLR-NS*	*FHLR vs FHLR-Rule1*
h=1	0.049	0.864	0.397	0.995
3	0.03	0.174	0.032	0.933
6	0.287	0.207	0.131	0.879
12	0.189	0.134	0.211	0.603

Panel B: Post Great Recession (2010:1-2018:12)

IP	*SW vs SW-NS*	*SW vs SW-Rule1*	*FHLR vs FHLR-NS*	*FHLR vs FHLR-Rule1*
h=1	0.742	0.541	0.446	0.695
3	0.151	0.602	0.779	0.993
6	0.354	0.657	0.773	0.400
12	0.384	0.698	0.811	0.152
CPI	*SW vs SW-NS*	*SW vs SW-Rule1*	*FHLR vs FHLR-NS*	*FHLR vs FHLR-Rule1*
h=1	0.781	0.399	0.069	0.473
3	0.8	0.493	0.662	0.294
6	0.159	0.092	0.746	0.476
12	0.185	0.084	0.901	0.673
UR	*SW vs SW-NS*	*SW vs SW-Rule1*	*FHLR vs FHLR-NS*	*FHLR vs FHLR-Rule1*
h=1	0.049	0.646	0.491	0.971
3	0.01	0.028	0.113	0.950
6	0.062	0.076	0.306	0.949
12	0.312	0.375	0.315	0.937

Table A.11: Predictions correlation

	SW vs SW-5y				FHLR vs FHLR-5y				SW vs FHLR			
	h=1	3	6	12	h=1	3	6	12	h=1	3	6	12
IP	0.812	0.948	0.917	0.867	0.848	0.871	0.799	0.654	0.910	0.955	0.921	0.740
CPI	0.505	0.640	0.503	0.537	0.595	0.667	0.628	0.591	0.823	0.720	0.597	0.513
UR	0.937	0.958	0.967	0.726	0.960	0.956	0.916	0.754	0.981	0.961	0.938	0.904

Table A.12: Predictions correlation: NS and Rule1

	SW vs SW-NS				SW vs SW-Rule1			
	h=1	3	6	12	h=1	3	6	12
IP	0.991	0.994	0.993	0.986	0.964	0.987	0.976	0.905
CPI	0.818	0.796	0.830	0.857	0.902	0.909	0.947	0.962
UR	0.996	0.995	0.99	0.971	0.982	0.988	0.983	0.950

	FHLR vs FHLR-NS				FHLR vs FHLR-Rule1			
	h=1	3	6	12	h=1	3	6	12
IP	0.956	0.984	0.989	0.983	0.954	0.985	0.986	0.971
CPI	0.892	0.920	0.912	0.892	0.832	0.817	0.782	0.802
UR	0.999	0.999	0.998	0.995	0.997	0.996	0.993	0.982

Figures

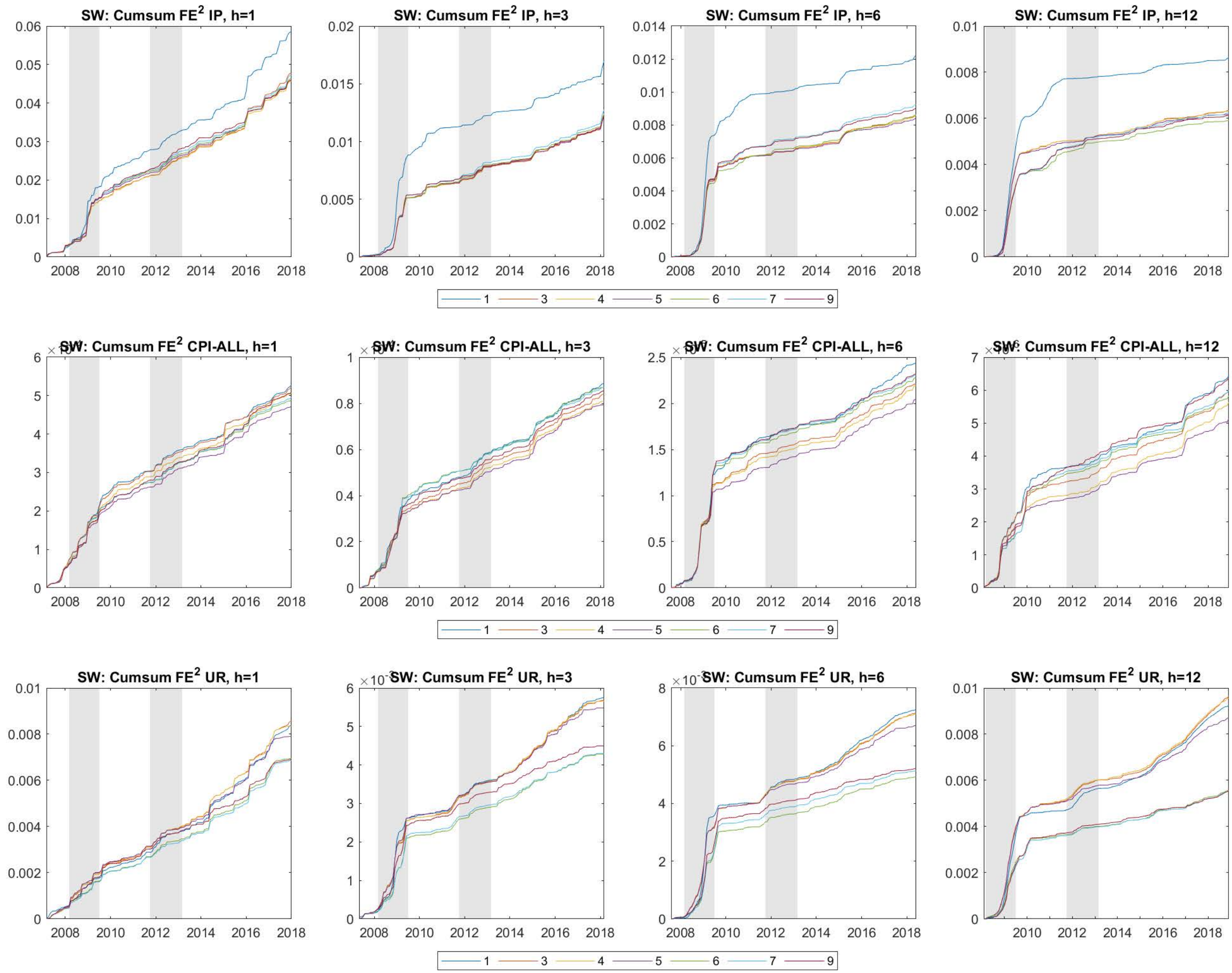

Figure A.5: Cumulative sum of the squared forecasting errors (FE^2) in SW with different choice of the number of factors r: IP, CPI and UR.

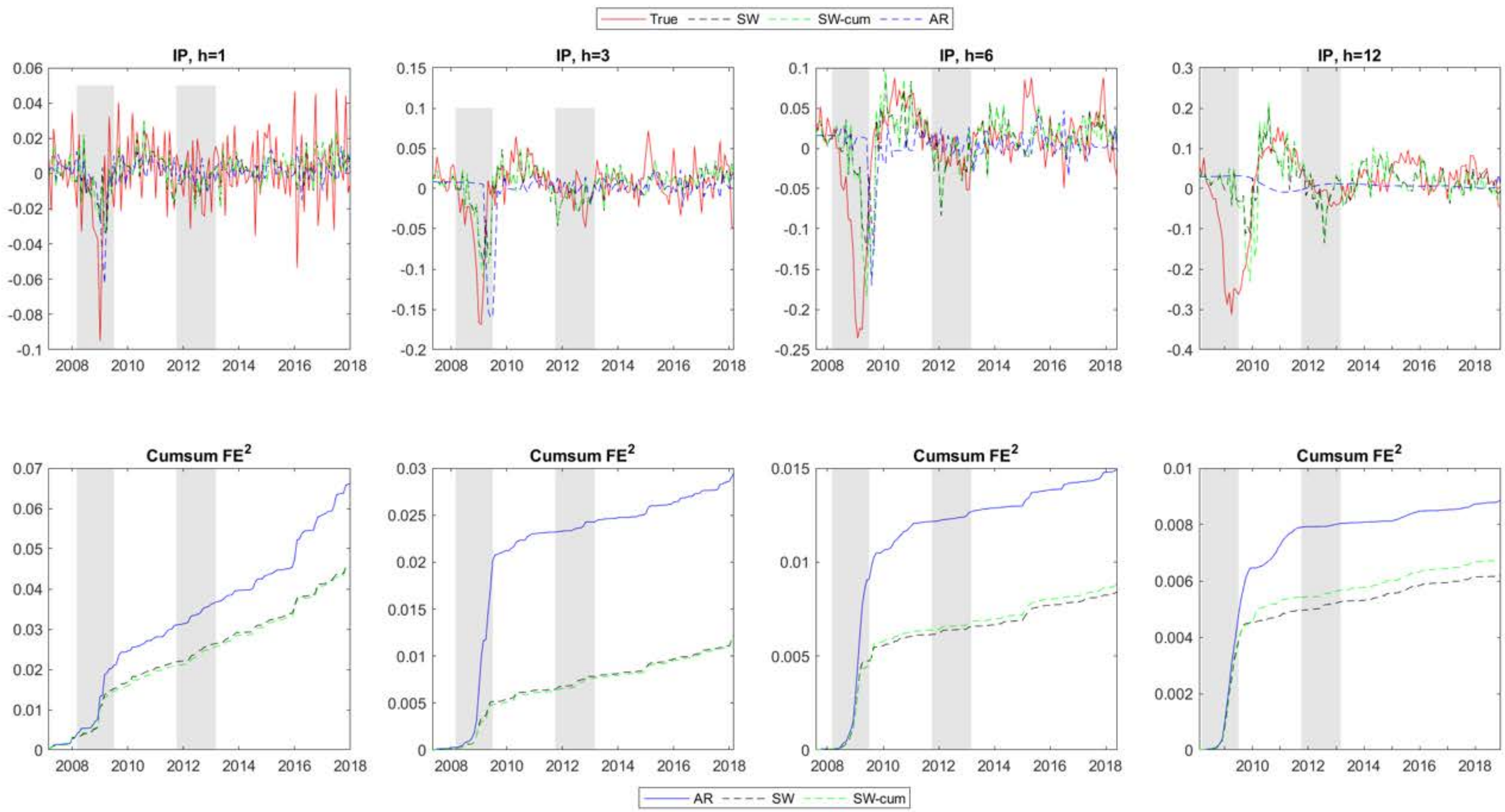

Figure A.6: Forecasts (panel above) and cumulative sum of the squared forecasting errors (FE^2, panel below): IP.
Note: SW predictions are obtained by directly forecasting the target on the factors estimated at time t. “SW-cum” uses the cumulated predictions as described in (3.2).

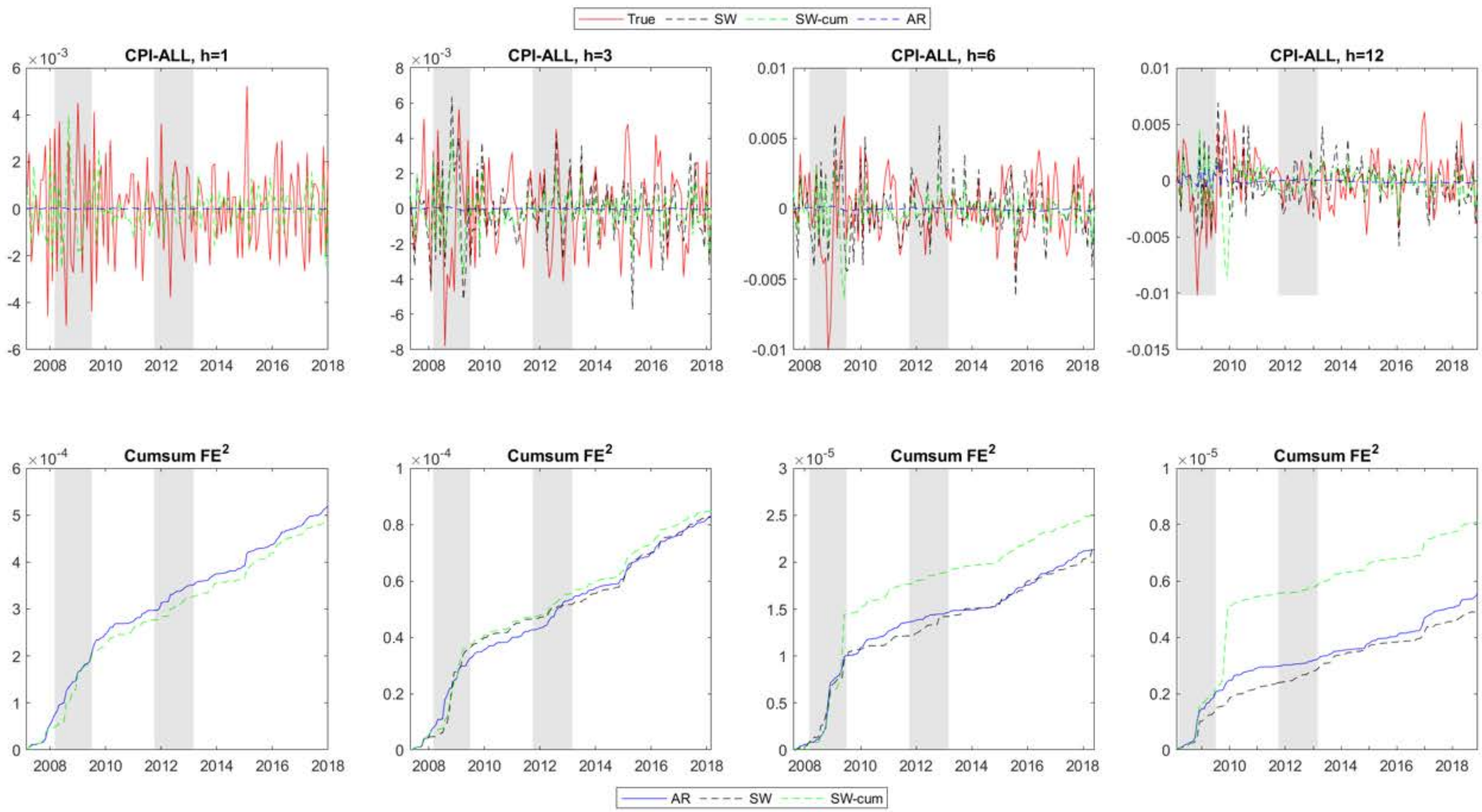

Figure A.7: Forecasts (panel above) and cumulative sum of the squared forecasting errors (FE^2, panel below): CPI.
Note: SW predictions are obtained by directly forecasting the target on the factors estimated at time t. "SW-cum" uses the cumulated predictions as described in (3.2).

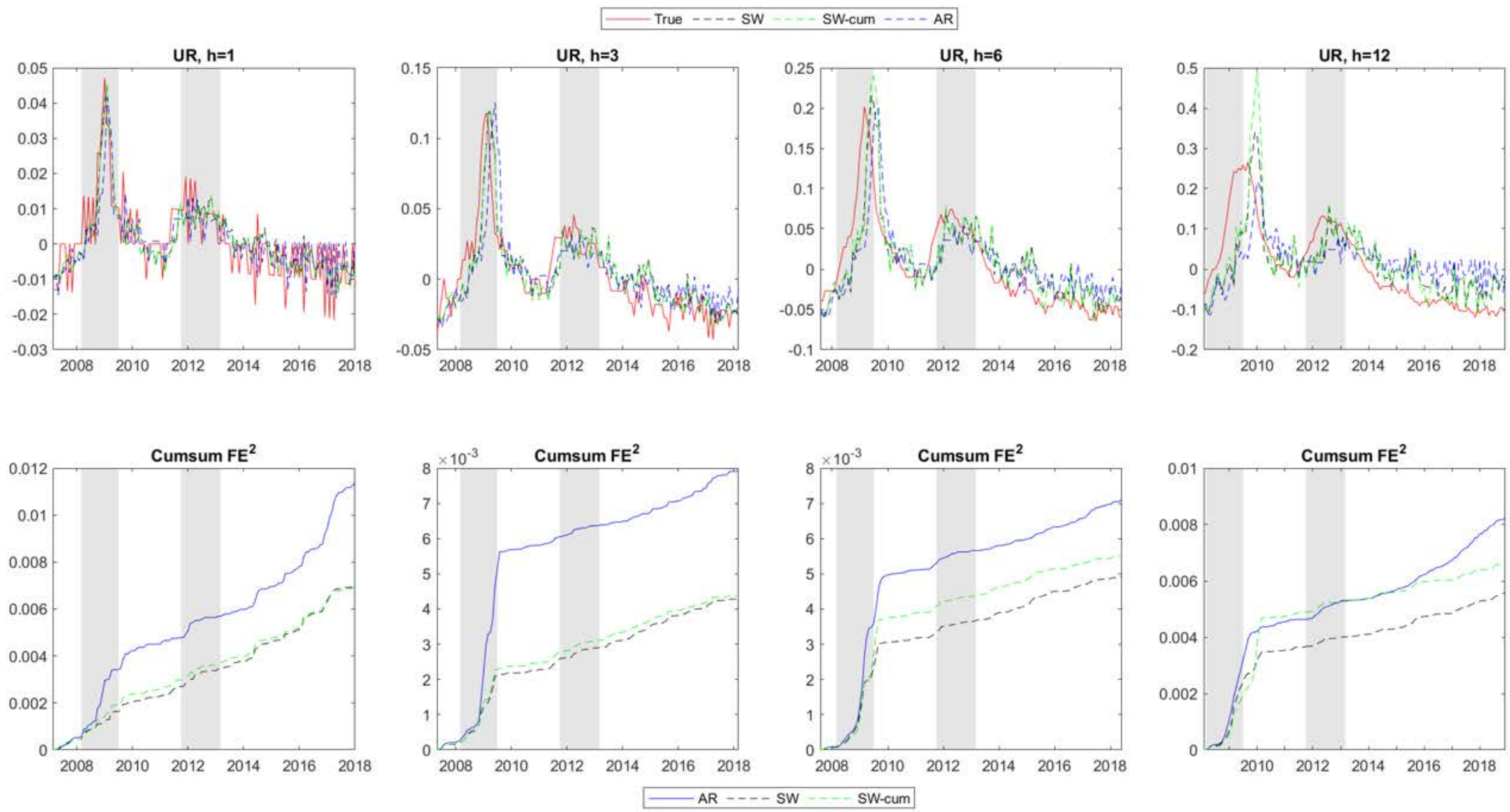

Figure A.8: Forecasts (panel above) and cumulative sum of the squared forecasting errors (FE^2, panel below): UR.
Note: SW predictions are obtained by directly forecasting the target on the factors estimated at time t. "SW-cum" uses the cumulated predictions as described in (3.2).

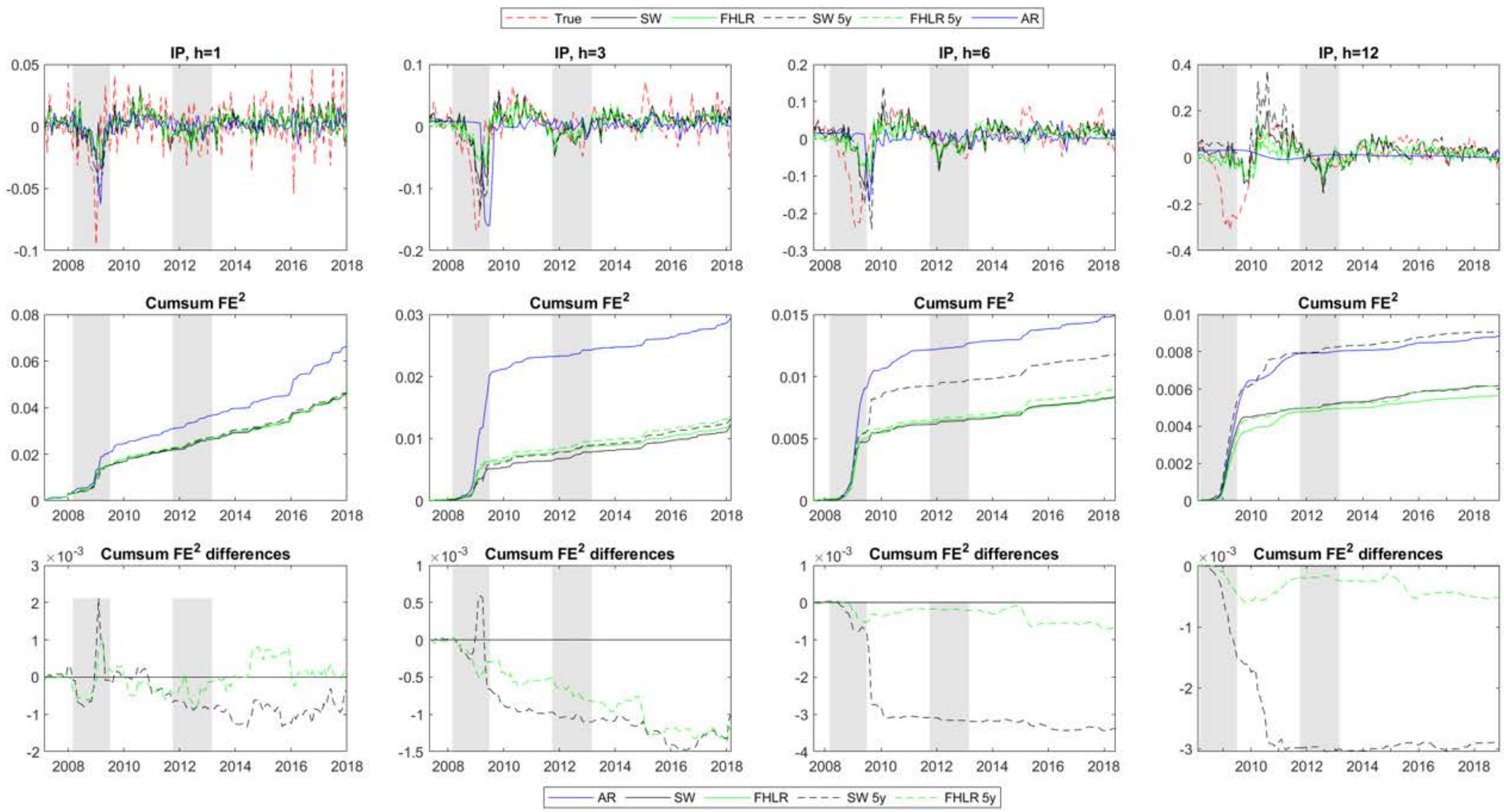

Figure A.9: Forecasts (panel above), cumulative sum of the squared forecasting errors (FE^2, center panel) and cumulative FE^2 differences with respect to the relative 10-year model (CSSFED, panel below): IP.
Note: "SW-5y" and "FHLR-5y" use a 5-year window. The estimation sample at time t is $[t-59, t]$.

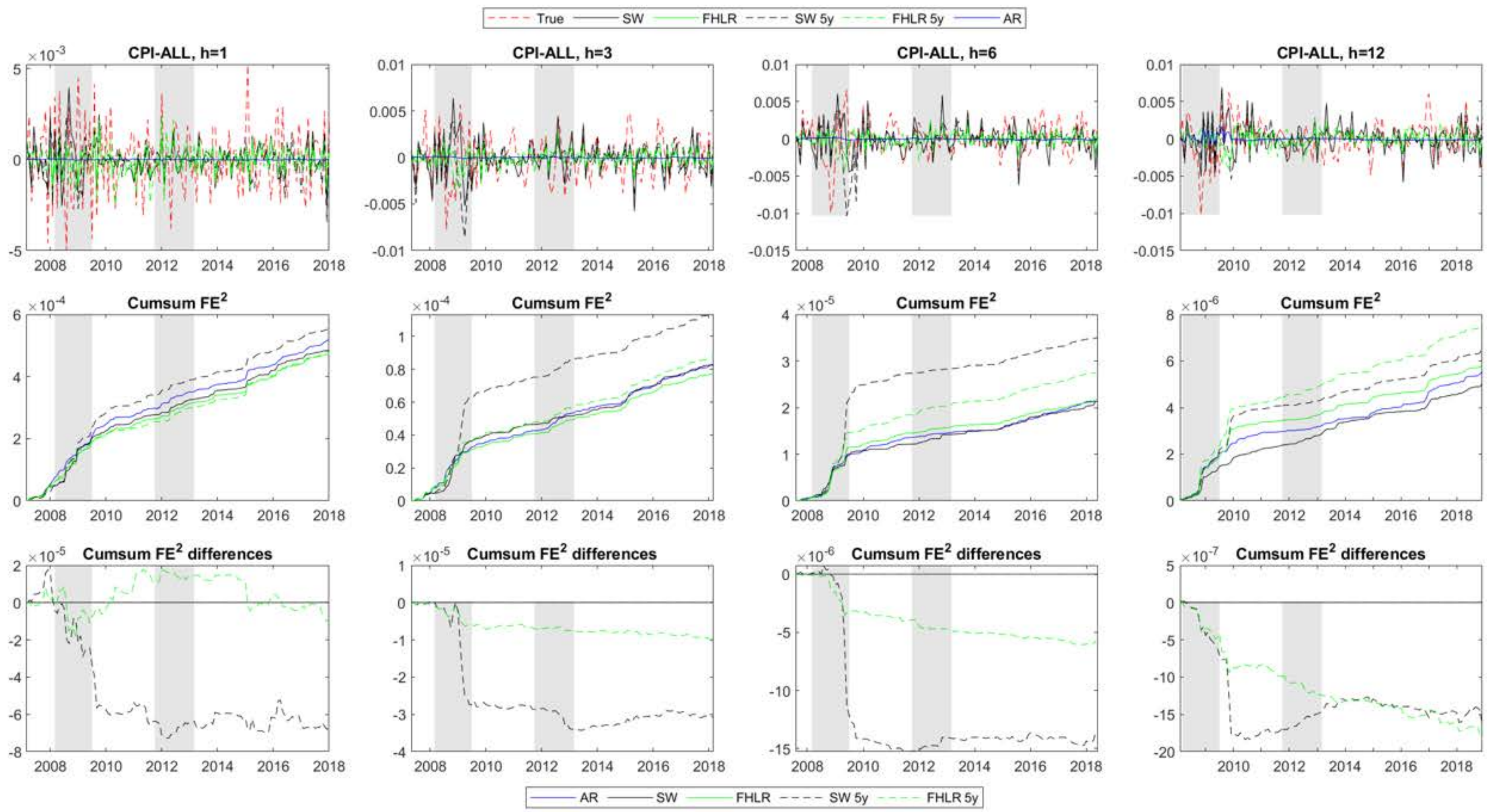

Figure A.10: Forecasts (panel above), cumulative sum of the squared forecasting errors (FE^2, center panel) and cumulative FE^2 differences with respect to the relative 10-year model (CSSFED, panel below): CPI.
Note: "SW-5y" and "FHLR-5y" use a 5-year window. The estimation sample at time t is $[t-59, t]$.

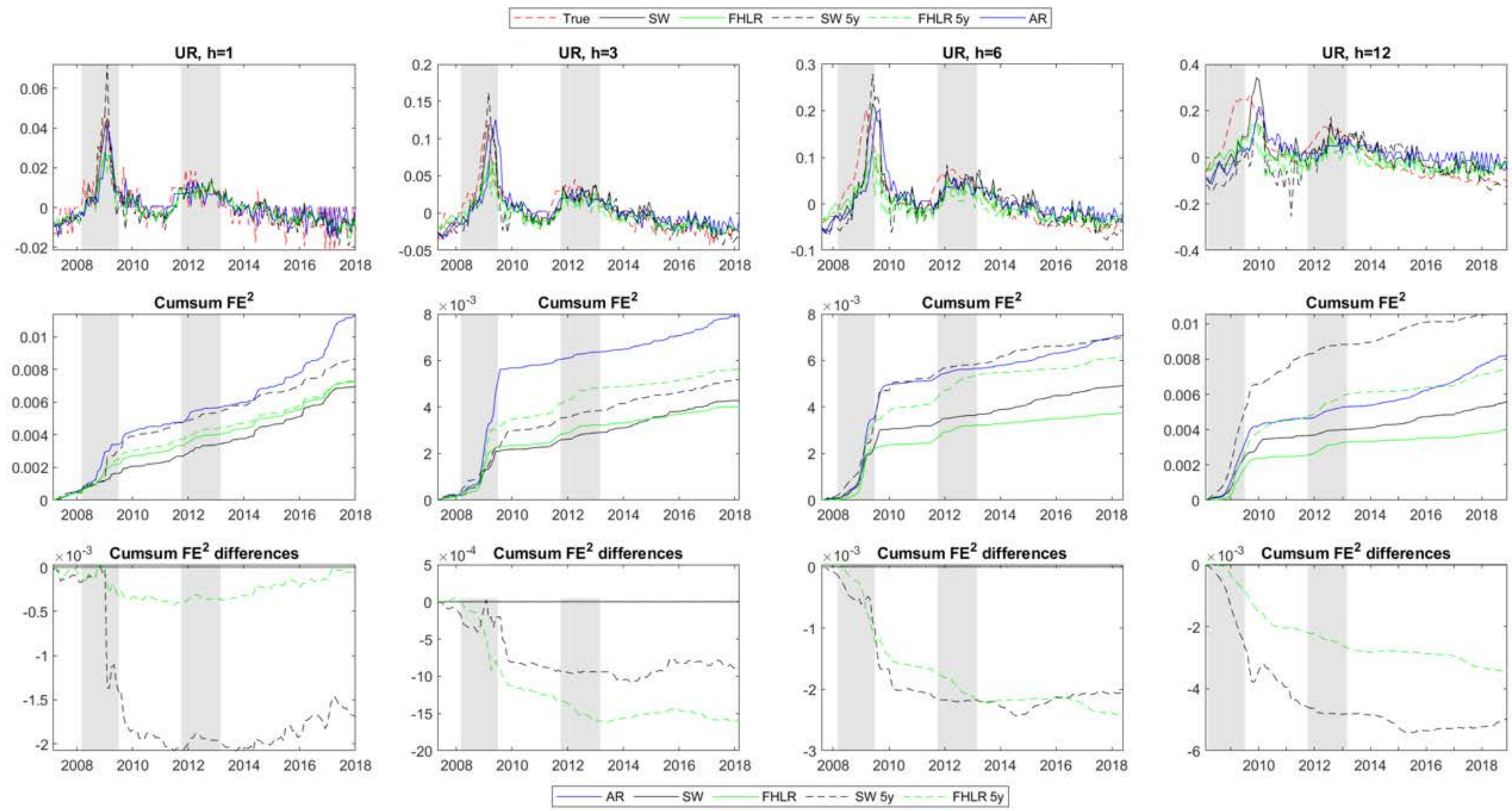

Figure A.11: Forecasts (panel above), cumulative sum of the squared forecasting errors (FE^2, center panel) and cumulative FE^2 differences with respect to the relative 10-year model (CSSFED, panel below): UR.
Note: "SW-5y" and "FHLR-5y" use a 5-year window. The estimation sample at time t is $[t-59, t]$.

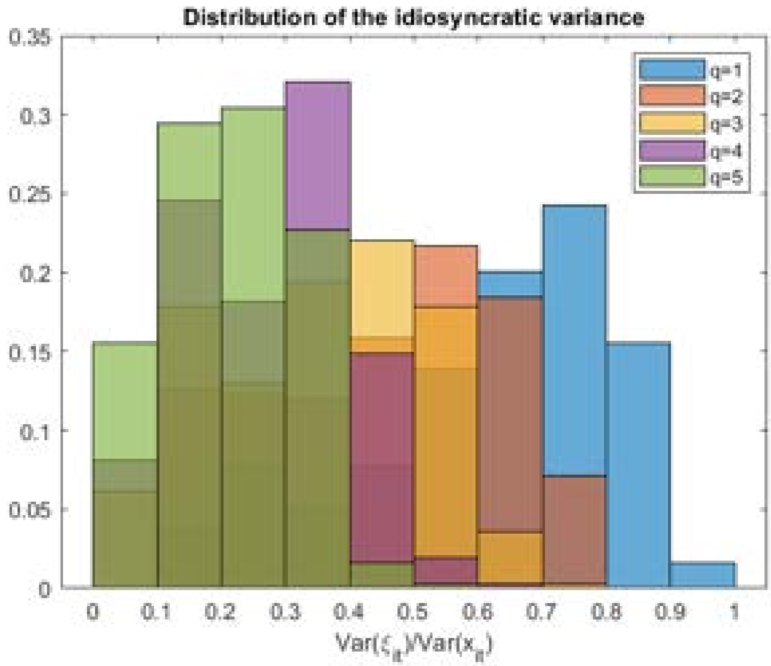

Figure A.12: Distribution share of variance explained by the idiosyncratic term

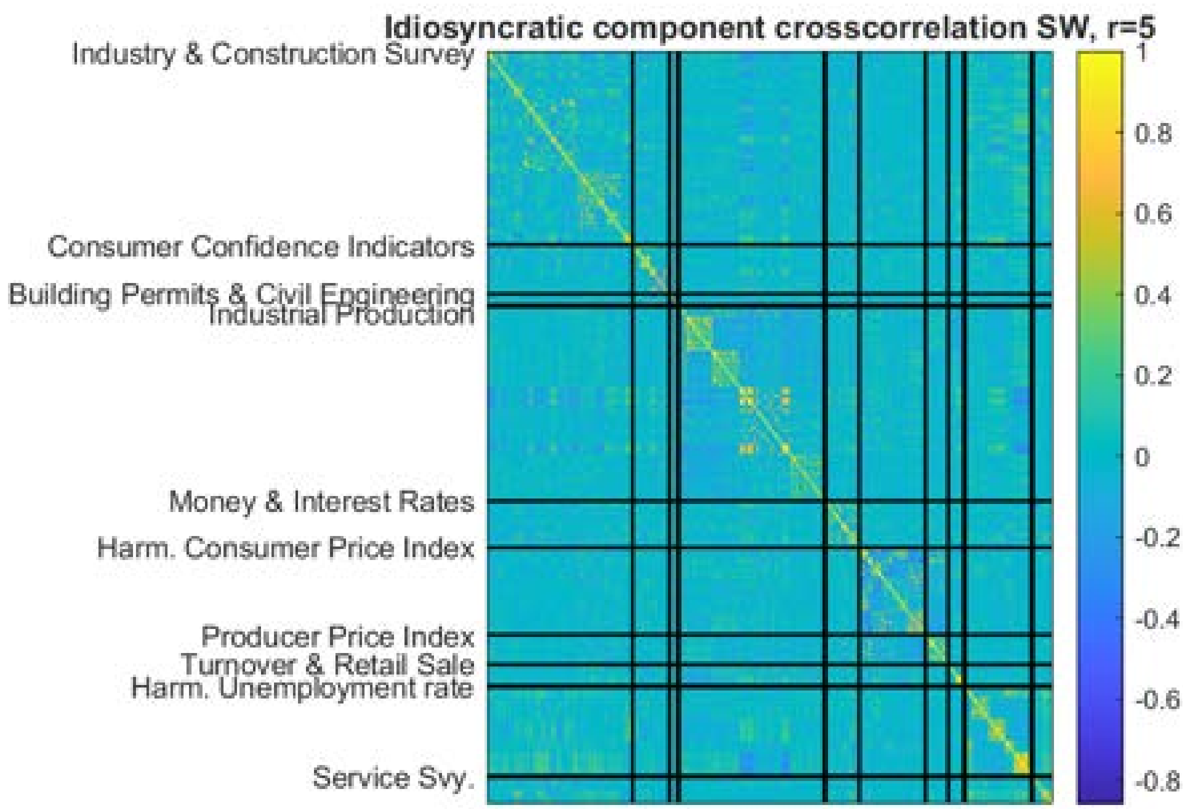

Figure A.13: Idiosyncratic cross correlation.
Note: $\hat{\Gamma}_0^{\xi}$ is obtained by the sample covariance matrix of the residuals of SW estimated assuming $r = 5$ static factors.

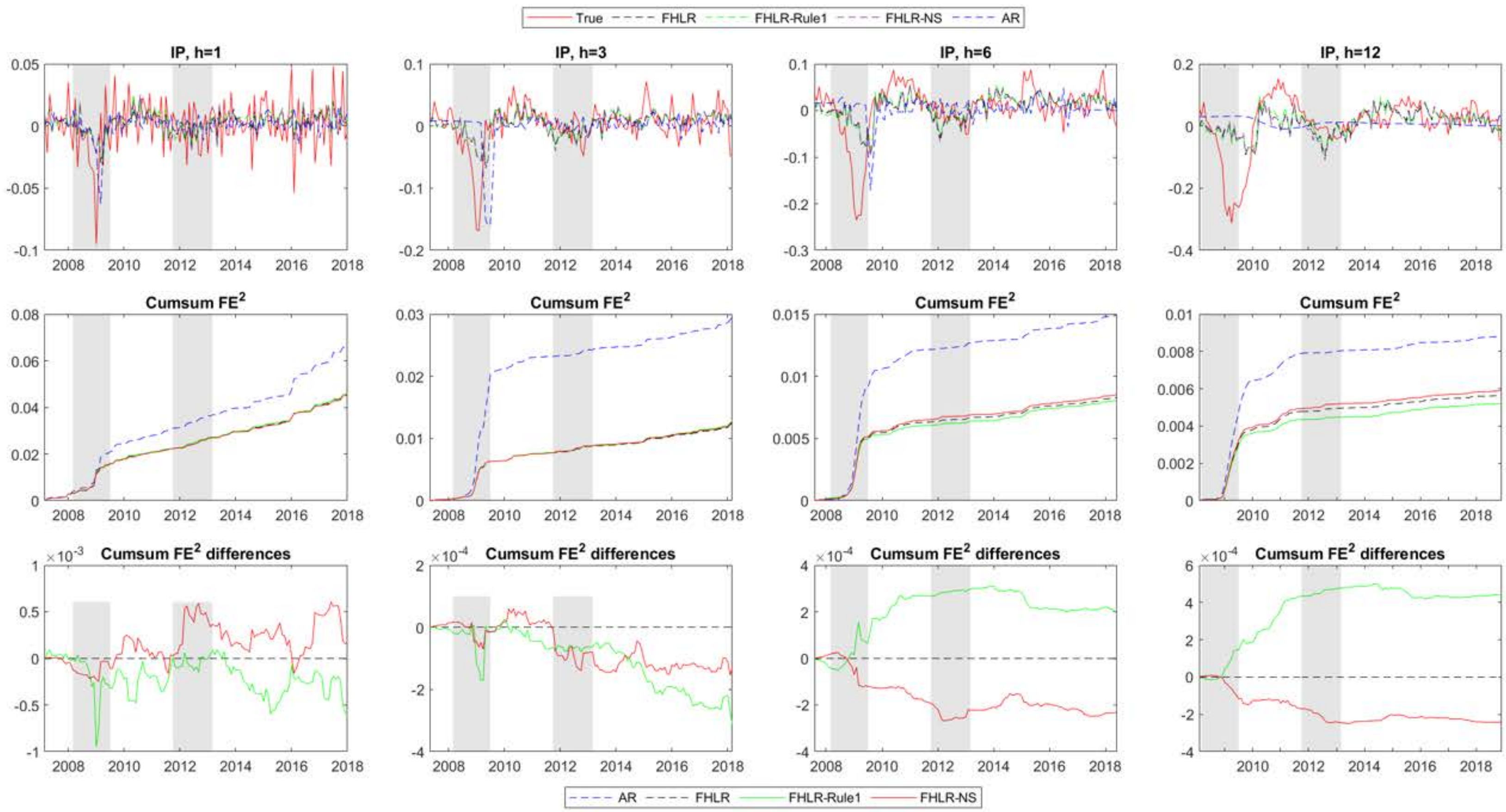

Figure A.14: Forecasts (panel above) and cumulative sum of the squared forecasting errors (FE^2, panel below) for the two new datasets (NS and Rule1): IP.

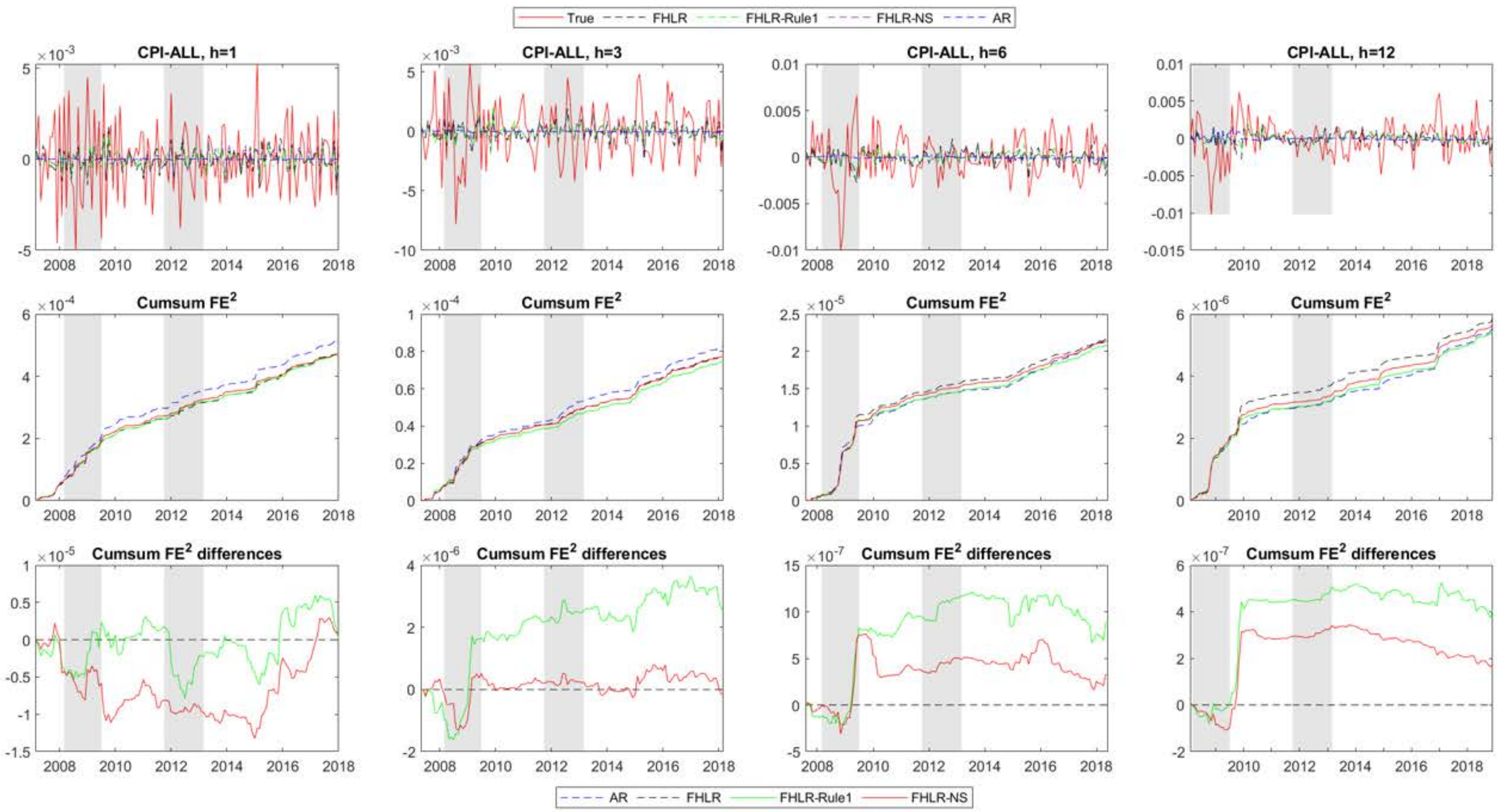

Figure A.15: Forecasts (panel above) and cumulative sum of the squared forecasting errors (FE^2, panel below) for the two new datasets (NS and Rule1): CPI.

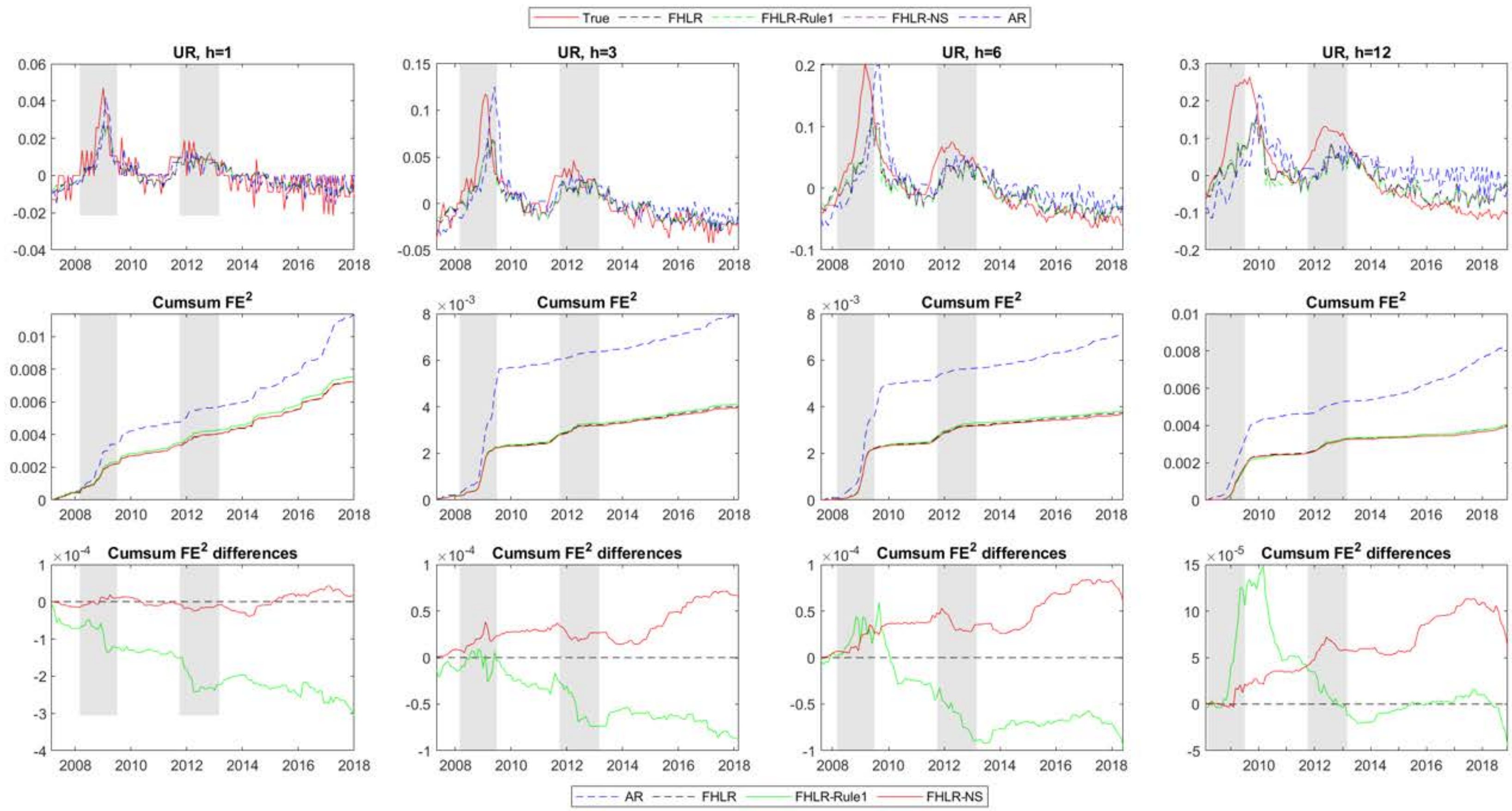

Figure A.16: Forecasts (panel above) and cumulative sum of the squared forecasting errors (FE^2, panel below) for the two new datasets (NS and Rule1): UR.