

DISCUSSION PAPER SERIES

No. 6564

COMPARING ALTERNATIVE PREDICTORS BASED ON LARGE-PANEL FACTOR MODELS

Antonello D'Agostino and
Domenico Giannone

INTERNATIONAL MACROECONOMICS



Centre for Economic Policy Research

www.cepr.org

Available online at:

www.cepr.org/pubs/dps/DP6564.asp

COMPARING ALTERNATIVE PREDICTORS BASED ON LARGE-PANEL FACTOR MODELS

Antonello D'Agostino, Central Bank and Financial Services Authority of Ireland
Domenico Giannone, European Central Bank and CEPR

Discussion Paper No. 6564
November 2007

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **INTERNATIONAL MACROECONOMICS**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as a private educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions. Institutional (core) finance for the Centre has been provided through major grants from the Economic and Social Research Council, under which an ESRC Resource Centre operates within CEPR; the Esmée Fairbairn Charitable Trust; and the Bank of England. These organizations do not give prior review to the Centre's publications, nor do they necessarily endorse the views expressed therein.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Antonello D'Agostino and Domenico Giannone

CEPR Discussion Paper No. 6564

November 2007

ABSTRACT

Comparing Alternative Predictors Based on Large-Panel Factor Models

This paper compares the predictive ability of the factor models of Stock and Watson (2002) and Forni, Hallin, Lippi, and Reichlin (2005) using a large panel of US macroeconomic variables. We propose a nesting procedure of comparison that clarifies and partially overturns the results of similar exercises in the literature. Our main conclusion is that for the dataset at hand the two methods have a similar performance and produce highly collinear forecasts.

JEL Classification: C31, C52 and C53

Keywords: factor models, forecasting and large cross-section

Antonello D'Agostino
Central Bank and Financial Services
Authority of Ireland (CBFSAI)
PO 559
Dame Street
D2 Dublin
IRELAND
Email:
antonello.dagostino@centralbank.ie

Domenico Giannone
European Central Bank
DG Research
Kaiserstrasse 29
D-60311
Frankfurt am Main
GERMANY
Email: dgiannon@ulb.ac.be

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=156621

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=154334

Submitted 07 November 2007

Comparing Alternative Predictors Based on Large-Panel Factor Models

Antonello D'Agostino*

Central Bank and Financial Services Authority of Ireland

Domenico Giannone†

European Central Bank and CEPR

August 2006

Abstract

This paper compares the predictive ability of the factor models of Stock and Watson (2002) and Forni, Hallin, Lippi, and Reichlin (2005) using a “large” panel of US macroeconomic variables. We propose a nesting procedure of comparison that clarifies and partially overturns the results of similar exercises in the literature. Our main conclusion is that for the dataset at hand the two methods have a similar performance and produce highly collinear forecasts.

JEL Classification: C31, C52, C53.

Keywords: Factor Models, Forecasting, Large Cross-Section.

1 Introduction

This paper compares two standard methods for forecasting using factor models in large panels: Stock and Watson (2002)'s static principal components method (SW)

*Central Bank and Financial Services Authority of Ireland - Economic Analysis and Research Department, PO Box 559 - Dame Street, Dublin 2, Ireland. E-mail: antonello.dagostino@centralbank.ie.

†European Central Bank, Kaiserstrasse 29, D- 60311, Frankfurt am Main, Germany. E-mail: domenico.giannone@ecb.int

and Forni, Hallin, Lippi, and Reichlin (2005)'s two-step approach based on dynamic principal components (FHLR). We are not the first in performing such an evaluation. Boivin and Ng (2005) focus on a very similar problem, while Stock and Watson (2005) compare the performances of a larger class of predictors.

The SW and FHLR methods essentially differ in the computation of the forecast of the common component. In particular, they differ in the estimation of the factor space and in the way projections onto this space are performed. In SW, the factors are estimated by static principal components (*PC*) of the sample covariance matrix and the forecast of the common component is simply the projection of the predicted variable on the factors.

FHLR propose efficiency improvements in two directions. First, they estimate the common factors based on generalized principal components (*GPC*) in which observations are weighted according to their signal to noise ratio. Second, they impose the constraints implied by the dynamic factors structure when the variables of interest are projected on the common factors. Specifically, they take into account the leading and lagging relations across series by means of principal components in the frequency domain. This allows for an efficient aggregation of variables that may be out of phase. Whether these efficiency improvements are helpful to forecast in a finite sample is however an empirical question and depends on the dataset at hand.

Literature has not yet reached a consensus. Using a large panel of US macroeconomic variables, Stock and Watson (2004a) find that both methods perform similarly (although they focus on the weighting of the idiosyncratic and not on the dynamic restrictions), while Boivin and Ng (2005) find that SW's method largely outperforms

the FHLR's and, in particular, they conjecture that the dynamic restrictions implied by the latter method are harmful for the forecast accuracy of the model. Schumacher (2006) finds instead that the FHLR's method generally outperforms the SW's in forecasting German macroeconomic variables.

This paper tries to shed some new light on these conflicting results. We perform a simulated out-of sample exercise to evaluate and compare the performance of the two methods in forecasting industrial production index (IP) and the consumer price index (CPI). Our data set, borrowed from Stock and Watson (2002), consists of 146 variables for the US economy. The data spans from 1959 to 1999.

In order to isolate and evaluate specific characteristics of the methods, we design a procedure where the two non-parametric approaches are nested in a common framework. In addition, for both versions of the factor models, we study the contribution of the idiosyncratic component to the forecast. We also investigate other non-core aspects of the model: robustness with respect to the choice of the number of factors and variable transformations. Finally, we study the sub-sample performances of the factor based forecasts.

The purpose of the exercise is to design an experiment for assessing the contribution of the core characteristics of different models to the forecasting performance and discussing auxiliary issues. We hope that this may also serve as a guide for practitioners in the field.

The paper is organized as follows. Section 2 describes the different models. Section 3 gives an overview of the database employed and describes its main characteristics. Section 4 designs the empirical exercise and discusses two issues: (i) factors vs uni-

variate forecasts; (ii) comparison between factor based forecasts. Section 5 concludes.

2 Models

Consider an $(n \times 1)$ covariance stationary process $Y_t = (y_{1,t}, \dots, y_{n,t})'$. We are interested in forecasting some elements $y_{i,t+h}$ of Y_t , at time $t + h$, by using all the variables $(y_{1,t}, \dots, y_{n,t})'$ as predictors. The best linear forecast is defined by the following linear projection

$$y_{i,t+h|t} = \text{proj} \{y_t | \Omega_t\} \quad (2.1)$$

where $\Omega_t = \text{span} \{Y_{t-p}, p = 0, 1, 2, \dots\}$ is a potentially large information set at time t when the forecasts are made.

When the size of the information set, n , is large, the above projection is unfeasible in practice since it requires the estimation of a large number of parameters with a substantial loss of degrees of freedom (curse of dimensionality). If most of the interactions among variables in the information set are accounted by few underlying common factors, $F_t = \{F_{1t}, \dots, F_{rt}\}$ with $r \ll n$, while there is a limited interaction among variable-specific dynamics, the curse of dimensionality problem can be solved. In this case, the projection of a variable y_{it+h} on the whole information set Ω_t is well approximated by the projection on the smaller information set including common factors and past values of the variable, therefore:

$$y_{i,t+h|t} = \text{proj} \{y_{i,t+h|t} | \Omega_t\} \approx \text{proj} \{y_{i,t+h|t} | \Omega_t^i\} \quad (2.2)$$

where $\Omega_t^i = \text{span} \{F_t\} \cup \text{span} \{y_{it}, y_{it-1}, \dots\}$ is the parsimonious representation of the information set that exploits the factor structure. The parsimonious approximation

makes the projection feasible since it only requires the estimation of a limited number of parameters. It is worth stressing that this projection coincides with the optimal one if the idiosyncratic components are cross-sectionally orthogonal.

2.1 Approximate Factor Structures

Suppose that $X_t = (x_{1,t}, \dots, x_{n,t})'$ is the standardized version of Y_t , i.e.

$$x_{i,t} = \frac{y_{i,t} - \mu_i}{\sigma_i}, \quad \mu_i = E[y_{i,t}], \quad \sigma_i = \sqrt{E[(y_{i,t} - \mu_i)^2]}$$

If X_t is described by a factor model, it can be written as the sum of two orthogonal components:

$$x_{i,t} = b_i(L)f_t + \xi_{it} = [b_{i0}, \dots, b_{is}] \begin{pmatrix} f_t \\ \dots \\ f_{t-s} \end{pmatrix} + \xi_{i,t} = \underline{b}_i F_t + \xi_{i,t} \quad (2.3)$$

or, in vector notation:

$$X_t = B(L)f_t + \xi_t = [B_0, \dots, B_s] \begin{pmatrix} f_t \\ \dots \\ f_{t-s} \end{pmatrix} + \xi_t = \underline{B}F_t + \xi_t \quad (2.4)$$

where f_t is a $(q \times 1)$ vector of dynamic factors, $B(L) = B_0 + B_1L + \dots + B_sL^s$ is an $(n \times q)$ matrix of filters of finite length s , ξ_t is the $(n \times 1)$ vector of idiosyncratic components, F_t is the $(r \times 1)$ vector of the stacked factors, with $r = q(s + 1)$. We assume that f_t and ξ_t are mutually orthogonal stationary processes and define $\chi_t = B(L)f_t$ as the common component.

We will refer to $X_t = B(L)f_t + \xi_t$ as the dynamic representation, and to $X_t = \underline{B}F_t + \xi_t$ as the static representation. Correspondingly, f_t will be defined as vector of the dynamic factors while F_t as the vector of the static factors.¹

¹The model presented above is a restricted version of the more general “dynamic factor model” proposed by Forni, Hallin, Lippi, and Reichlin (2000) in which the filter $B(L)$ may have an infinite length. For a detailed discussion of the relations between “restricted” and “general” dynamic factor model see Forni, Giannone, Lippi, and Reichlin (2005); Forni, Hallin, Lippi, and Reichlin (2005); Giannone, Reichlin, and Sala (2006).

Given the orthogonality assumption between common factors and idiosyncratic component, the spectral density matrix of X_t at each frequency $\theta \in [-\pi, \pi]$ can be decomposed into the sum of the spectral densities of the common and the idiosyncratic component²:

$$\Sigma(\theta) = \Sigma_\chi(\theta) + \Sigma_\xi(\theta) \quad (2.5)$$

where $\Sigma_\chi(\theta) = B(e^{-i\theta})\Sigma_f(\theta)B(e^{-i\theta})'$ is the spectral density matrix of the common component χ_t and $\Sigma_\xi(\theta)$ is the spectral density matrix of the idiosyncratic component ξ_t . Analogously, the covariance matrix of X_t can be decomposed as:

$$\Gamma_k = \Gamma_k^\chi + \Gamma_k^\xi \quad (2.6)$$

where $\Gamma_k^\chi = \underline{B}\Gamma_k^F\underline{B}'$, Γ_k^F is the covariance matrix of F_t at lag k and Γ_k^ξ is the covariance matrix of ξ_t at lag k .

Note that the rank of the spectral density of the common component, $\Sigma_\chi(\theta)$, is equal to q , the number of dynamic factors, while the rank of the covariance matrix Γ_k^χ is equal to r , the number of stacked (static) factors.

The forecast of the i th variable h -steps ahead can be decomposed into the sum of two components: the forecast of the common component and the forecast of the idiosyncratic component. Given the orthogonality assumption the forecast of the common component is obtained by projecting onto the space spanned by the common factors. Then, assuming that the bulk of dynamic interaction among variables is captured by the common factors, the forecast of the idiosyncratic component can be

²Given a matrix A with complex entries, its conjugate transpose will be denoted as A' .

approximated by using only the past values of the dependent variable:

$$\begin{aligned} x_{i,t+h|t} &\approx \text{proj} \{ \chi_{i,t+h} | \Omega_t^i \} + \text{proj} \{ \xi_{i,t+h} | \Omega_t^i \} \\ &= \text{proj} \{ \chi_{i,t+h|t} | F_t \} + \text{proj} \{ \xi_{i,t+h|t} | y_{i,t}, y_{i,t-1}, \dots \} \end{aligned} \quad (2.7)$$

The projection above is not feasible in practice since the common factors are unobserved. However, if data follow an approximate dynamic factor model, the set of common factors F_t can be consistently estimated by appropriate cross-sectional averages, or aggregators in the terminology of Forni and Reichlin (1998) and Forni and Lippi (2001). The intuition is that only the pervasive common sources survive the aggregation, since the weakly correlated idiosyncratic errors are averaged out. Building on Chamberlain and Rothschild (1983), Forni, Hallin, Lippi, and Reichlin (2000) and Stock and Watson (2002) have shown that principal components of the observed variables X_t , are appropriate averages. That is, the common component can be approximated by projecting either on the first r principal components of the covariance matrix (see Stock and Watson (2002)) or on the first q dynamic principal components (see Forni, Hallin, Lippi, and Reichlin (2000)).

For the empirical application we will consider the following consistent estimator of the autocovariance matrix of standardized data, $\hat{X}_t = (\hat{x}_{1t}, \dots, \hat{x}_{nt})'$:

$$\hat{\Gamma}_k = \frac{1}{T-k-1} \sum_{t=k}^T \hat{X}_t \hat{X}'_{t-k} \quad (2.8)$$

where

$$\hat{x}_{i,t} = \frac{y_{i,t} - \hat{\mu}_i}{\hat{\sigma}_i}$$

and $\hat{\mu}_i$ and $\hat{\sigma}_i$ denote the sample mean and standard deviation of the scalar process y_{it} respectively. T denotes the sample size. The spectral density matrix will be estimated

by averaging a given number m of autocovariances:

$$\hat{\Sigma}(\theta) = \frac{1}{2\pi} \sum_{k=-m}^m w_k \hat{\Gamma}_k e^{-i\theta k} \quad (2.9)$$

where w_k are weights satisfying the conditions: $w(0) = 1$ and $0 \leq w(k) \leq 1, \forall k \leq m$.

The estimates of the spectral density are consistent provided that $m \rightarrow \infty$ and $\frac{m}{T} \rightarrow 0$ as $T \rightarrow \infty$. In the empirical section we will use $m = \sqrt{T}$, which satisfies the above asymptotic requirements.

2.2 SW approach

The Stock and Watson (2002) approach to forecasting with factor models consists of estimating the common factors F_t as the sample principal components of the whole, standardized panel \hat{X}_t . The estimated factors are then used as predictors. Consider the following eigenvalue problem for the sample covariance matrix:

$$\hat{\Gamma}_0 V_r = V_r D_r \quad (2.10)$$

where D_r is a diagonal matrix having on the diagonal the first r largest eigenvalues of $\hat{\Gamma}_0$ and $V_r = [v_1, \dots, v_r]$ is the $(n \times r)$ matrix whose columns are the corresponding eigenvectors. The first r principal components (PC) are defined as:

$$\hat{F}_t = V_r' \hat{X}_t \quad (2.11)$$

and the covariance matrix of the common component is estimated by:

$$\hat{\Gamma}_0^x = V_r D_r V_r' \quad (2.12)$$

If the data follow an approximate dynamic factor model, \hat{F}_t are consistent estimates of the unknown common factors.

The estimated common factors are then used as predictors. Specifically,

$$\hat{y}_{i,t+h|t} = \text{proj} \left\{ y_{i,t+h} | \hat{\Omega}_t^i \right\}$$

where $\hat{\Omega}_t^i = \text{span} \left\{ \hat{F}_t \right\} \cup \text{span} \left\{ y_{i,t}, y_{i,t-1}, \dots \right\}$. This approach is implemented through the following forecasting equation:

$$y_{i,t+h}^{pc} = \alpha_{ih} + \beta_{ih} \hat{F}_t + \gamma_{ih}(L) y_{i,t} + \varepsilon_{it+h} \quad (2.13)$$

where the two sub-scripts i and $t+h$ refer to the variable i and to the forecast horizon h respectively, while the super-script pc indicates that factors are extracted via principal components. The lag polynomial $\gamma_{i,h}(L)$ is of length s and $\varepsilon_{i,t+h}$ is an error term. As stressed above, the general idea behind this approach is that the comovements among series can be captured by means of the common factors; since the interdependence among idiosyncratic components is assumed to be weak it can be captured by lagged values of the dependent variable.³

Denoting by $\hat{\alpha}_{ih}, \hat{\beta}_{ih}, \hat{\gamma}_{ih}$ the *OLS* estimates, the forecast of $y_{i,T+h|T}^{pc}$ is defined as:

$$\hat{y}_{i,T+h|T}^{pc,ols} = \hat{\alpha}_{ih} + \hat{\beta}_{ih} \hat{F}_T + \hat{\gamma}_{ih}(L) y_{iT} \quad (2.14)$$

Note that in this approach the factor structure assumption is exploited only for the extraction of the common factors. The forecasting projection, *eq.*(2.13), does not incorporate the restrictions implied by the dynamic factor structure. Indeed, the OLS projection does not exploit either the lead-lag structure among the series (which are

³This equation is a restricted version of that used by Stock and Watson (2002), since it allows only for dependence on contemporaneous factors. They call this a DI-AR (diffusion index-autoregressive) forecast. When the autoregressive part is dropped it is called simply DI forecast. SW (2002) show that, in most cases, DI-AR and DI forecasts are as good as or better than forecasts obtained including also lagged factors, labelled as DI-AR, Lag.

reflected in the rank of the spectral density matrix) or the orthogonality assumption between common and idiosyncratic component. These restrictions are instead embedded in the FHLR method.

2.3 FHLR Approach

The starting point of the method proposed by Forni, Hallin, Lippi, and Reichlin (2005) is the problem of the spectral density matrix, defined, at a given frequency θ , as:

$$\widehat{\Sigma}(\theta)\mathcal{V}_q(\theta) = \mathcal{V}_q(\theta)\mathcal{D}_q(\theta) \quad (2.15)$$

where $\mathcal{D}_q(\theta)$ is a diagonal matrix having the diagonal on the first q largest eigenvalues of $\widehat{\Sigma}(\theta)$ and $\mathcal{V}_r(\theta)$ is the $(n \times q)$ matrix whose columns are the corresponding eigenvectors.

If the panel of series X_t is driven by q dynamic factors, a consistent estimate of the spectral density matrix of the common component is given by:

$$\widehat{\Sigma}_\chi(\theta) = \mathcal{V}_q(\theta)\mathcal{D}_q(\theta)\mathcal{V}_q(\theta)' \quad (2.16)$$

The spectral density matrix of the idiosyncratic part, is estimated as a residual:

$$\widehat{\Sigma}_\xi(\theta) = \widehat{\Sigma}(\theta) - \widehat{\Sigma}_\chi(\theta).$$

The covariance matrices of common and idiosyncratic parts are computed by applying the inverse Fourier transforms to the respective estimated spectral density matrices:

$$\widehat{\Gamma}_k^\chi = \frac{2\pi}{2m+1} \sum_{j=-m}^m \widehat{\Sigma}_\chi(\theta_j) e^{ik\theta_j} \quad (2.17)$$

$$\widehat{\Gamma}_k^\xi = \frac{2\pi}{2m+1} \sum_{j=-m}^m \widehat{\Sigma}_\xi(\theta_j) e^{ik\theta_j} \quad (2.18)$$

where $\theta_j = \frac{2\pi}{2m+1}j$ and $j = -m, \dots, m$.

The auto-covariance matrices computed in this first step incorporate the restrictions implied by the dynamic properties of the common factors, since they are estimated by imposing the rank reduction on the spectral density matrix.

In a second step, the estimated covariance matrix of the common components is used to solve the generalized principal components (*GPC*) problem:

$$\widehat{\Gamma}_0^\chi V_{rg} = \widehat{\Gamma}_0^\xi V_{rg} D_{rg} \quad (2.19)$$

$$\text{s.t. } V_{rg}' \widehat{\Gamma}_0^\xi V_{rg} = I_r$$

where D_{rg} is a diagonal matrix having on the diagonal the first r largest generalized eigenvalues of the pair $(\widehat{\Gamma}_0^\chi, \widehat{\Gamma}_0^\xi)$ and V_{rg} is the $(n \times r)$ matrix whose columns are the corresponding eigenvectors.

The first r GPCs are defined as:

$$\widehat{F}_t^g = V_{rg}' \widehat{X}_t \quad (2.20)$$

In practice the estimate of the covariance matrix of the idiosyncratic component, $\widehat{\Gamma}_0^\xi$, is ill-conditioned when the cross-sectional dimension is large and this makes the generalized principal components unstable. To overcome this problem, the out off-diagonal elements of $\widehat{\Gamma}_0^\xi$ are set to zero. Consequently, the generalized principal components can be seen as static principal components computed on weighed data; weights are inversely proportional to the variance of the idiosyncratic components. Such weighting scheme should provide a more efficient estimates of the common factors.

Using the estimates of the auto-covariance matrices of the common and idiosyncratic components, we can compute the forecasts of both components separately as:

$$\hat{\chi}_{i,T+r|T} = \text{Proj}[\chi_{i,T+r} | \hat{F}_T^g] = \hat{\Gamma}_{i,r}^\chi V_{rg} (V_{rg}' \hat{\Gamma}_0 V_{rg})^{-1} V_{rg}' \hat{X}_T \quad (2.21)$$

and

$$\hat{\xi}_{i,T+k|T} = \text{Proj}[\xi_{i,T+k} | \hat{x}_{iT}, \dots, \hat{x}_{iT-p}] = [\hat{\Gamma}_{ii,r}^\xi, \dots, \hat{\Gamma}_{ii,r+p}^\xi] W_{i,k}^{-1} [\hat{x}_{iT}, \dots, \hat{x}_{iT-p}]' \quad (2.22)$$

where $\hat{\Gamma}_{i,r}$ denotes the i th row of $\hat{\Gamma}_r$ and

$$W_{i,k} = \begin{bmatrix} \hat{\Gamma}_{ii,0} & \dots & \hat{\Gamma}_{ii,-(k-1)} \\ \dots & \dots & \dots \\ \hat{\Gamma}_{ii,k-1} & \dots & \hat{\Gamma}_{ii,0} \end{bmatrix}$$

while $\hat{\Gamma}_{ij,r}$ denotes the entry of the i th row and the j th column of $\hat{\Gamma}_r$.

The forecast of $y_{i,T+h|T}$, labelled as $y_{i,T+k|T}^{fhlr}$, is computed as the sum of the common component forecast and idiosyncratic component forecast:

$$\hat{y}_{i,T+k|T}^{fhlr} = \hat{\sigma}_i \left(\hat{\chi}_{i,T+k|T} + \hat{\xi}_{i,T+k|T} \right) + \hat{\mu}_i$$

Since the procedure is run on standardized data, the forecast are then transformed to reattribute mean and variance.

This approach explicitly takes into account the distinction between dynamic factors and their lags. It is therefore possible to infer the dynamic structure of the panel, that is, the number of shocks driving the economy. This can be addressed by, for example, looking at the performance of the models across different specifications of r and q .

2.4 Nesting the Models

In order to compare the predicting performances of the two factor models, the two methodologies need to be nested.

There are three main differences between the SW and FHLR methods. First the procedures differ in terms of the weighting scheme adopted when computing the common factors. As discussed above, SW use standard *PCs* to extract factors, while FHLR propose an estimator based on the *GPC*. We recall that when extracting the common factors the *GPC* down-weight series with large idiosyncratic components.

The second difference relates to the projection of common factors. While SW perform the projection using ordinary least squares (OLS), FHLR perform a non-parametric regression that takes into account the restrictions implied by the dynamic factor structure, essentially imposing restrictions by the rank reduction of the spectral density matrix. We will denote such projection as dynamic least squares (*DLS*).

Finally, the methods differ in the way they forecast the idiosyncratic component: SW include lags of the dependent variable as additional predictors, while FHLR exploit the orthogonality between idiosyncratic and common components and hence forecast the two components separately. If the idiosyncratic component is unforecastable, as we will show in our empirical exercise, we can focus only on the first two differences; therefore, considering only factor estimations and projections, four combinations are possible: PC/OLS OLS regression on principal components (SW); GPC/OLS: OLS regression on generalized principal components; PC/DLS: DLS regression on principal components; GPC/DLS: DLS regression on generalized principal

components (FHLR).

We can isolate and evaluate the relevance of the efficiency improvement associated with the FHLR's weighting scheme by comparing the forecasts obtained with PC/OLS and GPC/OLS estimators.

The GPC/OLS forecast is computed simply by replacing static principal component with generalized principal components in *eq.*(2.20):

$$\hat{x}_{i,T+h|T}^{gpc,ols} = \hat{\alpha}_{ih}^g + \hat{\beta}_{ih}^g \hat{F}_T^g \quad (2.23)$$

where the parameters are estimated by OLS.

The effect of the restrictions imposed with the dynamic projection can be instead evaluated by comparing PC/DLS and PC/OLS estimators. Thus, taking into account the restrictions implied by the factor structure and imposed through dynamic principal components (computed in the first step of the FHLR's procedure), we obtain the following forecasting equation:

$$\hat{x}_{i,T+h|T}^{pc,dls} = \hat{\Gamma}_{i,h}^\chi V_r (V_r' \hat{\Gamma}_0 V_r)^{-1} V_r' \hat{X}_T \quad (2.24)$$

and $\hat{y}_{i,T+h|T}^{pc,dls} = \hat{\sigma}_i \hat{x}_{i,T+r|T}^{pc} + \hat{\mu}_i$.

The unrestricted projection on simple principal components (PC/OLS) can be derived as a particular case of the forecasts in (*eq.* 2.24), when the number of dynamic factors, q , is equal to the cross-section dimension, n . In this case, the dynamic factor structure restrictions are not imposed, since the rank of the spectral density matrix is left unrestricted (we use all the dynamic principal components).⁴ Therefore, the

⁴It is worth stressing that we are not anymore in the framework of Section 2.3 since the number of dynamic factors, q , is allowed to be larger than the number of static factors, r .

estimated autocovariance matrix of the common component $\hat{\Gamma}_{i,r}^X$ is equal to the sample autocovariance matrix of the series themselves ($\hat{\Gamma}_{i,r}$) and the estimator is:

$$\hat{x}_{i,T+k|T}^{pc,ols} = w_k \hat{\Gamma}_{i,r} V_r (V_r' \hat{\Gamma}_0 V_r)^{-1} V_r' \hat{X}_T \quad (2.25)$$

Nevertheless, for $q = n$ the PC/DLS still does not nest the SW's method (PC/OLS) because of the presence of the weighting term w_k on the right-hand side. The presence of the weighting term is due to the fact that the spectral estimates are computed on the weighted auto-covariance function. For this reason, when computing the inverse Fourier transform of the spectral density matrix, $\hat{\Sigma}(\theta)$, we do not recover the covariance matrix of order k , $\hat{\Gamma}_k$, but $w_k \hat{\Gamma}_k$. Only when the weights are set equal to one, $w_k = 1, k = -m, \dots, m$ (rectangular window), the PC/DLS estimator, with $q = n$, matches the SW's estimator PC/OLS.

We will consider two cases; rectangular (rect) window ($w_k = 1, k = -m, \dots, m$) and Bartlett, triangular (triang) window ($w_k = 1 - \frac{|k|}{(2m+1)}, k = -m, \dots, m$). The latter was originally proposed by Forni, Hallin, Lippi and Reichlin (2005).

To understand how the weights w_k can affect forecasts, it is important to notice that the triangular window shrinks toward zero covariances at longer lags. This shrinking can damage the forecasts if the variables of interest are persistent, that is the covariances fade away slowly. On the other hand, if the persistence is low, covariances fade away much more rapidly and a shrinking term in the covariance function can help to reduce the sample variance of the forecasts, thus improving the forecast accuracy.

In our empirical exercise we will evaluate how different combination of: *PC* vs.

DPC (estimation of the common factors); *OLS* vs. *DLS* (projection on the common factors) and triang vs. rect widows (estimation of the spectral density matrix) affect the forecast precision.

3 Data and Basic Characteristics of the Panel

The data set employed for the out-of-sample forecasting analysis is the same as that adopted by Stock and Watson (2002), with the exception of a few series. The panel includes real variables (sectoral industrial production, employment and hours worked), nominal variables (consumer and producer price indices, wages, money aggregates), asset prices (stock prices and exchange rates), the yield curve and surveys.

Series are transformed to obtain stationarity. In general, for real variables such as employment, industrial production and sales, we use the monthly growth rate. We use first differences for series already expressed in rates: unemployment rate, capacity utilization, interest rate and some surveys. Prices and wages are transformed into first differences of annual inflation following Giannone, Reichlin, and Sala (2004) and Giannone, Reichlin, and Small (2005).

3.1 Descriptive Statistics

The main motivation behind the factor representation is the strong comovement observed in macroeconomic time series, which is possible only if there are few underlying common driving forces. The simplest statistic to describe comovements among series is the percentage of the variance of the panel accounted for by common factors estimated by *PCs*. If the series are characterized by strong comovements, then a small

number of principal components account for a relevant percentage of the overall panel variance while the remaining principal components have a small marginal contribution.⁵ On the other hand, a low degree of comovement is reflected in the fact that all principal components account for a small percentage of explained variance and hence a larger number of *PCs* are required to explain a non-trivial percentage of the overall variance. Table 1 reports the percentage of the total variance of the panel explained by $q = 1, \dots, 15$ dynamic common factors (second row) and $r = 1, \dots, 15$ static common factors (third row).⁶

Table 1: *Variance explained by the PCs (Commonality)*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
<i>n. of dynamic fac. (q)</i>	0.29	0.42	0.52	0.59	0.64	0.68	0.72	0.75	0.78	0.80	0.82	0.84	0.86	0.87	0.88
<i>n. of static fac. (r)</i>	0.16	0.26	0.31	0.36	0.40	0.44	0.48	0.51	0.54	0.56	0.59	0.60	0.62	0.64	0.66

Notes: Percentage of total variance explained by the first q (dynamic) and r (static) principal components.

Results show that comovements are relatively strong. A few number of dynamic principal components (three to four) capture most of the variance of the panel, while the marginal contribution of the remaining is quite small. There is however, a remarkable difference between the variance explained by static and dynamic *PCs*. Ten static factors, estimated by *PCs*, are needed to explain around the same amount of variance captured by three/four dynamic factors, estimated by dynamic *PCs*. This

⁵The marginal contribution of each principal component in explaining the panel variance (at each frequency) is equal to the eigenvalues of the covariance matrix (spectral density matrix at a given frequency). The approximate factor structure assumption requires that few of those eigenvalues become large while the remaining stay small, when the cross-sectional dimension becomes larger and larger.

⁶Our measure is given by: $\text{trace}\hat{\Gamma}_0^\chi/\text{trace}\hat{\Gamma}_0$ where $\hat{\Gamma}_0^\chi$ is estimated by using the first q dynamic principal components (cfr. eq.(2.17)) and the first r static principal components (cfr. eq.(2.12)). For dynamic principal components, the spectral density matrix is estimated using a Bartlett window.

finding suggests that there are substantial dynamics in our panel. In fact, lagged dynamic factors are counted as additional static factors, and hence their presence increase the asymptotic rank of the covariance matrix r . On the other hand, the asymptotic rank of the spectral density matrix, q , is invariant with respect to the presence of lagged factors. Therefore, a big gap in the variance accounted for the static and dynamic PCs indicate that the panel is characterized by a rich lead-lag structure. In fact, in *eq. (2.4)* the difference between the number of dynamic factors, q , and the number of static factors, r , reflects the length, s , of the filters $B(L)$. If we select q and r so as to explain around 50 percent of the total variance, then $q \sim 3$ and $r \sim 10$. Indeed, given that $r = q(s + 1)$, then $s \sim 3, 4$, suggesting the presence of a quite relevant dynamic structure. The restrictions implied by these rich dynamics are imposed by the FHLR's approach when computing the forecast by projecting on the common factors.

In order to evaluate the importance of idiosyncratic dynamics, Table 2 shows the distribution of the percentage of variance left unexplained when projecting on two dynamic common factors (noise to signal). The estimates are computed using dynamic principal components (see *eq. 2.18*).

It is evident that the distribution of the idiosyncratic variance is quite dispersed: some variables are largely driven by common sources (around 50 percent of the series have the variance of the idiosyncratic component that accounts for a percentage of total variance between 20 and 50 percent, *cfr. columns 4 to 6*), while other variables are driven mainly by idiosyncratic forces. This result suggests that weighting

Table 2: *Distribution of the idiosyncratic variance*

$\frac{\text{Var}(\xi_{it})}{\text{Var}(x_{it})}$	(.0-.1)	(.1-.2)	(.2-.3)	(.3-.4)	(.4-.5)	(.5-.6)	(.6-.7)	(.7-.8)
% of series	0.68	4.11	14.38	15.75	19.18	16.44	17.81	11.64

Note: Frequency distribution of the size of the idiosyncratic component. First line: percentage of variance not explained by the common component. Second line: percentage of series with idiosyncratic variance in a fixed range.

the variables according to their signal to noise ratio, that is, putting less weight on variables with large idiosyncratic variance, should provide an efficiency improvement in the extraction of the common factors. Therefore, *GPCs* should be more effective than *PCs* to recover the factor space. However, the empirical importance of such restrictions will be evaluated in an out-of-sample forecasting exercise.

In conclusion, these findings (strong comovements, rich dynamic structure and heterogeneity in the degree of commonality) suggest that factor models provide a reasonable representation of our panel and indicate that there is room for improving the SW simple principal components forecasts by exploiting the restrictions implied by the factor structure and by weighting appropriately the data when extracting the common factors.

To find out if such efficiency improvements are helpful in forecasting, an empirical evaluation is necessary. This will be carried out using a simulated pseudo out-of-sample exercise.

4 Design of the Forecasting Comparison and Empirical Results

Let us define IP as the monthly industrial production index and CPI as the consumer price index. The variables we forecast are y_{t+h} and π_{t+h} , where $y_t = 100 \times \log IP_t$ is the (rescaled) logarithm of IP and $\pi_t = (p_t - p_{t-12})/p_{t-12} \times 100$ is the annual CPI inflation.

Since IP enters in the panel in monthly growth rates, while CPI enters as monthly differences of annual inflation, we first compute the forecasts $\widehat{\Delta y}_{T+1|T}, \dots, \widehat{\Delta y}_{T+h|T}$ and $\widehat{\Delta \pi}_{T+1|T}, \dots, \widehat{\Delta \pi}_{T+h|T}$, then the forecasts for the (log) IP and the level of inflation are computed as:

$$\widehat{y}_{T+h|T} = y_T + \widehat{\Delta y}_{T+1|T} + \dots + \widehat{\Delta y}_{T+h|T} \quad (4.26)$$

and

$$\widehat{\pi}_{T+h|T} = \pi_T + \widehat{\Delta \pi}_{T+1|T} + \dots + \widehat{\Delta \pi}_{T+h|T} \quad (4.27)$$

The prediction accuracy is evaluated at a given horizon h , using the mean squared forecast error (*MSFE*) metric, given by:

$$MSFE_{\pi}^h = \frac{1}{T_1 - T_0 - h + 1} \sum_{T=T_0}^{T_1-h} (\widehat{\pi}_{T+h|T} - \pi_{T+h}) \quad (4.28)$$

and

$$MSFE_y^h = \frac{1}{T_1 - T_0 - h + 1} \sum_{T=T_0}^{T_1-h} (\widehat{y}_{T+h|T} - y_{T+h}) \quad (4.29)$$

The sample has a monthly frequency and ranges from January 1959 (1959:1) to February 1999 (1999:2). The evaluation period is 1970:01 to 1999:02. $T_1=1999:02$ is the last available point in time, $T_0=1969:12$ and $h=12$.

We perform a preliminary forecast analysis to understand the broad predicting characteristics of the models and in particular the role of the idiosyncratic component in forecasting.

We compute MSFEs for IP and CPI using the two factor models, for different parameter specifications and for a naive model, the random walk with drift. Table 3 below reports the MSFEs of the factor based forecasts relative to those of the naive model (RMSFE). A number below one indicates that the factor model specifications deliver, on average, more accurate forecasts than the naive model.

We report results for a wide range of parameter specifications.⁷ The static rank r ranges from zero (no factors) up to a maximum of fifteen. The lags s of the dependent variable in SW, and the lags of the idiosyncratic component itself in FHLR, included to forecast the idiosyncratic component, range from zero (no lagged dependent variable) up to six. The FHLR method is estimated using a triangular window of size \sqrt{T} and with three dynamic factors, $q = 3$. When $r = 0$, the SW forecasts correspond to those of an autoregressive (AR) model. The entries with both $s = 0$ and $r = 0$ correspond to the benchmark naive model (random walk with drift) for which we report the MSFE.

The two factor models appear to clearly outperform the AR model (specification with $r=0$ in Table 3) in forecasting IP and CPI. Moreover, the common factors capture the bulk of the dynamics of the two variables of interest. In fact, once controlling

⁷For each specification, the lag length and number of common factors are kept fixed for the whole out-of-sample simulation exercise. All the exercises in the paper have also been performed by using information criteria to select the parameterizations in real-time, as in Boivin and Ng (2005). Results, available upon request, show that this alternative strategy does not affect the qualitative conclusions of the paper.

Table 3: *Relative MSFEs of 12-steps ahead predictions, SW and FHLR methods*

<i>IP</i>	<i>r</i>						<i>r</i>					
	<i>s</i>	<i>0</i>	<i>1</i>	<i>3</i>	<i>5</i>	<i>10</i>	<i>15</i>	<i>0</i>	<i>1</i>	<i>3</i>	<i>5</i>	<i>10</i>
<i>0</i>	24.37	0.96	0.56	0.57	0.57	0.61	0.65	0.60	0.61	0.64
<i>1</i>	0.99	0.91	0.55	0.56	0.57	0.61	0.65	0.60	0.60	0.64
<i>2</i>	1.00	0.90	0.55	0.56	0.56	0.60	0.65	0.61	0.61	0.64
<i>3</i>	1.03	0.91	0.55	0.56	0.56	0.58	0.65	0.61	0.61	0.65
<i>6</i>	1.07	0.95	0.57	0.59	0.58	0.59	0.64	0.59	0.60	0.64
<i>CPI</i>	<i>r</i>						<i>r</i>					
<i>s</i>	<i>0</i>	<i>1</i>	<i>3</i>	<i>5</i>	<i>10</i>	<i>15</i>	<i>0</i>	<i>1</i>	<i>3</i>	<i>5</i>	<i>10</i>	<i>15</i>
<i>0</i>	4.91	0.69	0.68	0.58	0.61	0.69	0.76	0.68	0.69	0.73
<i>1</i>	0.95	0.67	0.69	0.58	0.61	0.68	0.77	0.69	0.69	0.72
<i>2</i>	0.96	0.67	0.70	0.59	0.62	0.69	0.78	0.70	0.70	0.73
<i>3</i>	0.97	0.67	0.70	0.59	0.63	0.69	0.78	0.70	0.71	0.73
<i>6</i>	0.98	0.66	0.72	0.61	0.64	0.71	0.79	0.70	0.71	0.74

Notes: Relative MSFEs for different model specifications. s and r denote the number of lags for the idiosyncratic component and the number of static factors respectively. In the first cell, on the left hand side, there is the MSFE of the naive model ($s = 0$ and $r = 0$). The number of dynamic factors q for the FHLR's specification is equal to three.

for common factors, the lags of the dependent variables do not help to forecast. This suggests that the idiosyncratic component is unforecastable for the transformations we adopted and only the common components of industrial production and inflation are predictable. In summary, the common factors, constructed to explain the maximum amount of cross-sectional variance of the panel, are also able to capture all the predictable dynamics of the key aggregated variables.⁸

Table 4 below compares the relative performances of the SW and FHLR forecasts for different specifications of the the dynamic rank q and for $s = 0$ (given the

⁸This result is in line with that of Giannone, Reichlin, and Sala (2004). Stock and Watson (2002) obtained a similar result for IP, but not for CPI. What characterizes our forecast is the different transformation used for prices: Stock and Watson (2002) use monthly changes of monthly CPI inflation, while we use monthly changes of yearly inflation. The transformations in Stock and Watson (2002) are such that prices display dynamic properties that are different from those of the rest of the panel. Indeed their spectrum peaks at the high frequencies. In this paper, the transformed series have, as most of the other series included in the panel, a power spectrum localized on the business cycle frequencies. This makes the use of the lags of the prices themselves redundant.

unforecastability of the idiosyncratic component). The static rank r is reported in the first column. The FHLR's method is evaluated for a dynamic rank (reported in parenthesis in the first row) ranging from 1 to 5. The SW forecasts are reported for comparison. The spectral density matrix is estimated with a triangular window of size $m = \sqrt{T}$.

Results show that for all the specifications, factor forecasts are more precise than the simple univariate forecasts (value less than one) and, when we include at least three dynamic factors ($q = 3$), FHLR factor-based forecasts are very close to the SW forecasts. However, the RMSFEs associated to the FHLR's forecasting method are slightly higher than those obtained with the SW principal components forecasts. This is particularly true for CPI. These results could be interpreted as evidence that the restrictions implied by the dynamic factor structure and imposed by the FHLR method are not satisfied by the data. Nevertheless, there are further qualifications to this result. The magnitude of the improvements in forecasting accuracy of the two factor models over the univariate model is of first order importance, while the differences between factor models are quite small or at least of second order. The two models not only have similar performances but also produce highly correlated forecasts. Indeed the correlation between forecasts obtained with the two factor models is around 0.9 both for IP and CPI.⁹ However, as stressed in section 2.3, at this stage, the two methods are not directly comparable and, apart from the dynamic factor structure restrictions, there are still many other differences between the two

⁹Correlation is computed between forecasts obtained with a specification of 3 dynamic factors, 10 static factors and without idiosyncratic component.

techniques. Therefore it is impossible to draw a conclusion on the usefulness of the factor restrictions on the basis of these results. We will come back to this point in the next sections.

Table 4: *The performance of the SW and FHLR methods (RMSFE)*

r	<i>IP</i>						<i>CPI</i>					
	<i>SW</i>	<i>fhlr(1)</i>	<i>fhlr(2)</i>	<i>fhlr(3)</i>	<i>fhlr(4)</i>	<i>fhlr(5)</i>	<i>SW</i>	<i>fhlr(1)</i>	<i>fhlr(2)</i>	<i>fhlr(3)</i>	<i>fhlr(4)</i>	<i>fhlr(5)</i>
1	0.96	0.80					0.69	0.87				
2	0.64	0.71	0.63				0.64	0.78	0.73			
3	0.56	0.66	0.63	0.65			0.68	0.72	0.67	0.76		
5	0.57	0.66	0.69	0.60	0.61	0.62	0.58	0.68	0.68	0.68	0.68	0.69
10	0.57	0.71	0.69	0.61	0.60	0.60	0.61	0.73	0.68	0.69	0.70	0.71
15	0.61	0.73	0.72	0.64	0.63	0.64	0.69	0.74	0.70	0.73	0.74	0.75

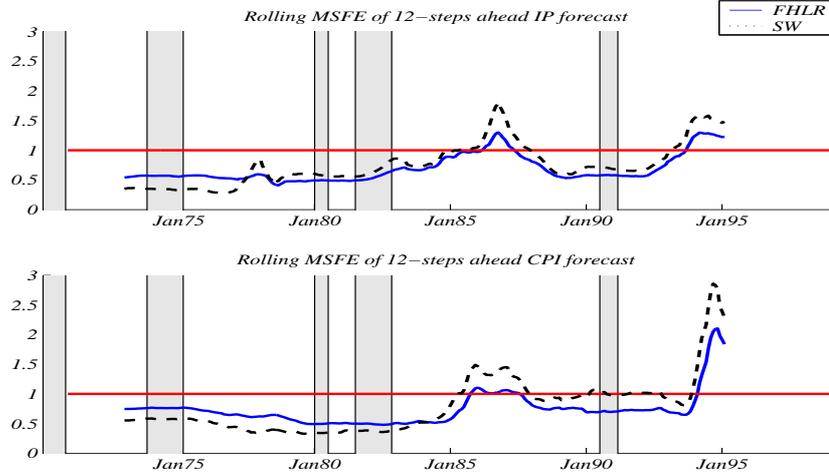
Notes: Relative MSFEs for different model specifications. r denotes the number of static factors, q , indicated in brackets, denotes the number of dynamic factors. The idiosyncratic component s is equal to zero.

4.1 When Does the Large Cross-Section Help?

In this section we study how the forecast accuracy of the two multivariate models has evolved over time. Figure 1 reports the MSFEs of the two factor models, relative to the naive benchmark, smoothed over time with a centered moving window spanning six years. The shaded areas denote the US recessions as dated by the NBER. The factor forecasts are computed using a specification with ten static factors for the SW's model and with three dynamic factors and ten static factors for the FHLR's model. The benchmark is the random walk with drift, specification with $r = 0$ and $s = 0$ in eq.(2.13).

The picture shows that the advantages of factor models on the AR forecast come, almost entirely, from the first part of the sample until 1985 (value of the RMSFE less than one). Improvements are quite remarkable over this period. This is not

Figure 1: *Time Varying Performance*



Note: Relative MSFE. Forecasts of dynamic specification obtained with: three dynamic, ten static factors and without idio. Forecasts of static specification obtained with: ten static factors and without idio. Moving window of six years.

surprising because at that time the series are characterized by strong comovements. This is a situation in which common factors describe the data particularly well. From 1985 onwards the picture is very different. The two factor models lose most of their advantage over the simple AR model, that is the RMSFEs are, on average, around one. The post 1985 period, known as the "Great Moderation", is characterized by a sizable decline in the volatility of output and price measures (see Stock and Watson, 2004b, and references cited there). Our result of declined forecast accuracy of factor models relative to simple univariate models indicates that the great moderation has also been associated with an important structural break in the relations between IP, CPI and the common factors extracted from our large panel. A slight increase in predictability is nonetheless observed in concomitance with the recession in the 90s. One possible interpretation is that downturn periods are characterized by increased comovements, a

situation in which factor based forecasts are likely to be more accurate.¹⁰ The picture also shows that the SW’s method performs better than the FHLR’s one on the pre-85 sample. On the other hand, during the post-1985 sample, the SW simple principal components forecasts become less accurate than the simple random walk, while the FHLR model still retains some advantage over the naive model in forecasting CPI inflation.

To shed more light on the sub-samples performances we split the out-of-sample evaluation period in two parts according to the “Great Moderation” breakdown of mid-1980s. We study the performances of the factor based forecasts for different parameterizations of the static and the dynamic rank.

Table 5 below reports the results in terms of RMSFE. The features emerging from Figure 1 are confirmed. During the first period, forecasts obtained with the factor models are remarkably accurate both for IP and CPI. In the second sub-sample the factor models lose the forecast accuracy that they display in the first sub-sample. Performance, relative to the naive benchmark is particularly poor for the SW method. These results hold for all specification of the static (r) and dynamic (q) rank. D’Agostino, Giannone, and Surico (2006) have shown that the decline of the predictive accuracy relative to naive forecasts is not limited to factor-based forecasts but is a general feature of a wide class of model-based and institutional forecasts.

Results on the relative performances of the two factor models are also confirmed. During the first period the SW forecasts are more accurate than those of the FHLR method and both the factor based forecasts are more accurate than the naive ones.

¹⁰Comovement is indeed one of the main features of the recessions.

This result holds for both IP and CPI inflation and for all specifications of the static and the dynamic rank. However, in the post-1985 sample, SW forecasts are less accurate than the naive forecasts for both the variables and for all the specifications, while, FHLR forecasts are less accurate than the naive forecasts for IP, but are more accurate than the random walk for CPI inflation, provided that the number of common shocks is kept small.

Table 5: *The Performance of the SW and 2S methods over the two sub-samples (RMSFE)*

r	<i>IP 1970:1-1984:12</i>						<i>CPI 1970:1-1984:12</i>					
	<i>SW</i>	<i>fhlr(1)</i>	<i>fhlr(2)</i>	<i>fhlr(3)</i>	<i>fhlr(4)</i>	<i>fhlr(5)</i>	<i>SW</i>	<i>fhlr(1)</i>	<i>fhlr(2)</i>	<i>fhlr(3)</i>	<i>fhlr(4)</i>	<i>fhlr(5)</i>
1	0.95	0.77					0.62	0.86				
2	0.53	0.67	0.57				0.56	0.75	0.69			
3	0.44	0.62	0.56	0.60			0.61	0.68	0.62	0.73		
5	0.43	0.62	0.63	0.53	0.54	0.56	0.46	0.65	0.63	0.64	0.63	0.63
10	0.46	0.68	0.65	0.55	0.54	0.54	0.50	0.70	0.64	0.66	0.66	0.67
15	0.51	0.70	0.68	0.59	0.57	0.58	0.59	0.72	0.68	0.69	0.69	0.70
r	<i>IP 1985:1-1999:2</i>						<i>CPI 1985:1-1999:2</i>					
	<i>SW</i>	<i>fhlr(1)</i>	<i>fhlr(2)</i>	<i>fhlr(3)</i>	<i>fhlr(4)</i>	<i>fhlr(5)</i>	<i>SW</i>	<i>fhlr(1)</i>	<i>fhlr(2)</i>	<i>fhlr(3)</i>	<i>fhlr(4)</i>	<i>fhlr(5)</i>
1	0.99	0.97					1.09	0.95				
2	1.37	0.95	1.05				1.10	0.94	0.92			
3	1.38	0.93	1.11	1.00			1.06	0.96	0.91	0.93		
5	1.55	0.91	1.05	1.10	1.08	1.10	1.24	0.90	0.91	0.95	0.98	1.02
10	1.32	0.91	1.00	0.97	1.07	1.02	1.24	0.90	0.86	0.90	0.94	0.93
15	1.27	0.90	0.99	1.01	1.05	1.01	1.24	0.87	0.85	0.91	1.01	1.02

Notes: Relative MSFEs for different model specifications and for two sub-samples. r denotes the number of static factors, q , indicated in brackets, denotes the number of dynamic factors. The idiosyncratic component s is equal to zero.

Another important feature observed from the results is that over the first sample the most accurate forecasts computed using the FHLR's method are obtained using a specification with at least three dynamic factors, $q > 3$. After the mid-eighties, instead, a specification with one or two dynamic factors, $q = 1, 2$, provide more

accurate forecasts.

In conclusion, four main results emerge from our exercise. First, if variables are transformed appropriately, idiosyncratic dynamics are irrelevant for forecasting. Second, both the SW and FHLR factor model outperform a simple AR model. The advantages of factor models over a simple univariate model are substantial during the pre-1985 period, which is characterized by strong comovements and high volatility. Third, factor based forecasts are highly collinear and have similar forecasting accuracy. The SW forecast are, however, slightly more accurate than those obtained via the FHLR in the pre-1985 period, but in the post-85 they become less accurate than both FHLR and random walk forecasts. Finally, the number of common shocks that generate predictable fluctuations of IP and CPI has declined in the “Great Moderation” period.

Although the results of this and the previous sections have highlighted important features of both the data panel and the models, they still cannot be used to evaluate the relevance of the dynamic restrictions and weighting scheme imposed by the FHLR. As stressed in section 2, the two factor-based approaches differ in a number of core and non-core characteristics. To understand the rationales behind the different performances it is necessary to isolate these different features.

4.2 SW and FHLR

In this section the two factor models are nested in a unified comparable framework. This allows us to evaluate the importance of the core (weighting schemes (*GPC* vs. *PC*) and dynamic restrictions) and non-core (estimation window of the spectral

density matrix) differences.

As stressed in Section 2, the models can be compared in two ways. First FHLR's approach can be nested in the SW approach by computing *OLS* projection on the generalized principal components instead of simple principal components, see *eq.* (2.23). This allows us to isolate the importance of the weighting scheme in the forecast accuracy. Second, the SW method can be obtained as a particular case of the FHLR method, with no restrictions on the spectral density matrix ($q = n$) and rectangular window for the estimation of the spectral density (see *eq.* (2.25)). In what follows we evaluate the relative performances of these nested models. We also evaluate the impact of alternative spectral density window estimators (triangular and rectangular) on the predictive performance.

Results for the whole sample and the two sub-samples are summarized in table 6. The static rank r is fixed as equal to ten.¹¹ We report the MSFEs relative to the random walk for six different forecast specifications, both for IP (left hand side) and CPI (right hand side).

For *OLS* forecasts based on generalized principal components (see *eq.* (2.23)), the covariance matrices of the common and idiosyncratic components are estimated by running dynamic principal components on the spectral density matrix that is computed with rectangular window (column 2, gpc,ols (r)) and triangular window (column 3, gps,ols (t)).

Dynamic forecast regression on simple (non-weighted) principal components (see

¹¹Qualitative results do not change for other specifications of the static rank. They are fully described in the working paper version, available at: <http://www.ecb.int/pub/pdf/scpwps/ecbwp680.pdf>.

eq. 2.24) are computed on the basis of two estimates of the spectral density matrix: using a rectangular (column 4, pc,dls (r)) and triangular window (column 5, pc,dls (t)). Results for the same exercise performed on weighted principal components are displayed in columns 6 (gpc,dls (r)) and 7 (gpc,dls (t)).¹²

The first column reports the static rank, q , ranging from 1 to n (no dynamic restrictions). The case with the spectral density matrix computed with a rectangular window and $q = n$ nests the SW's approach.

The table is rich in information. We now proceed to summarize it below.

First, forecasts based on simple *OLS* projections, computed on weighted and non-weighted principal components, are very similar (cfr. columns 2, 3). This is irrespective of the number of dynamic factors and the spectral density estimator (triangular vs rectangular window) used to compute the covariance matrix. Results are robust across samples and forecasted variables. This suggests that weighting for the signal to noise ratio when extracting the common factors does not have a major impact on forecasting accuracy.

Second, when focusing on the whole sample, forecasts based on dynamic factor regression, computed on both weighted (cfr. columns 4, 5) and non-weighted (cfr. columns 6, 7) principal components, are quite accurate and stable across specifications provided the number of common shocks q is larger than two. For $q = n$ results are very similar to those for $3 \leq q \leq 5$, suggesting that imposing the restrictions of the dynamic factor models has small effect on the forecasting accuracy. This stands in stark contrast to the conjecture of Boivin and Ng (2005), who claim that imposing

¹²Results in columns 5 are the same of the previous section.

Table 6: *The performances of nested models (RMSFE)*

	<i>Whole Sample</i>											
	<i>IP</i>						<i>CPI</i>					
	GPC, OLS		PC, DLS		GPC, DLS		GPC, OLS		PC, DLS		GPC, DLS	
	(r)	(t)	(r)	(t)	(r)	(t)	(r)	(t)	(r)	(t)	(r)	(t)
<i>q=1</i>	0.55	0.52	0.70	0.70	0.69	0.71	0.57	0.59	0.77	0.74	0.71	0.73
<i>q=2</i>	0.55	0.52	0.76	0.70	0.77	0.69	0.61	0.63	0.65	0.68	0.66	0.68
<i>q=3</i>	0.56	0.54	0.56	0.61	0.57	0.61	0.61	0.60	0.63	0.70	0.64	0.69
<i>q=4</i>	0.62	0.56	0.55	0.60	0.59	0.60	0.62	0.61	0.65	0.70	0.68	0.70
<i>q=5</i>	0.62	0.58	0.53	0.61	0.55	0.60	0.64	0.61	0.63	0.71	0.67	0.71
<i>q=n</i>			0.55	0.61					0.64	0.70		

	<i>I Sub-Sample (Pre-85)</i>											
	<i>IP</i>						<i>CPI</i>					
	GPC, OLS		PC, DLS		GPC, DLS		GPC, OLS		PC, DLS		GPC, DLS	
	(r)	(t)	(r)	(t)	(r)	(t)	(r)	(t)	(r)	(t)	(r)	(t)
<i>q=1</i>	0.45	0.42	0.66	0.68	0.64	0.68	0.48	0.49	0.74	0.71	0.67	0.70
<i>q=2</i>	0.41	0.41	0.71	0.66	0.70	0.65	0.53	0.53	0.60	0.64	0.62	0.64
<i>q=3</i>	0.46	0.43	0.48	0.55	0.49	0.55	0.50	0.50	0.54	0.66	0.55	0.66
<i>q=4</i>	0.49	0.44	0.45	0.54	0.46	0.54	0.49	0.50	0.57	0.66	0.58	0.66
<i>q=5</i>	0.54	0.48	0.44	0.55	0.46	0.54	0.51	0.50	0.54	0.66	0.57	0.67
<i>q=n</i>			0.46	0.56					0.54	0.65		

	<i>II Sub-Sample (Post-85)</i>											
	<i>IP</i>						<i>CPI</i>					
	GPC, OLS		PC, DLS		GPC, DLS		GPC, OLS		PC, DLS		GPC, DLS	
	(r)	(t)	(r)	(t)	(r)	(t)	(r)	(t)	(r)	(t)	(r)	(t)
<i>q=1</i>	1.20	1.20	0.97	0.90	1.02	0.91	1.09	1.13	0.94	0.92	0.91	0.90
<i>q=2</i>	1.45	1.32	1.07	0.96	1.27	1.00	1.07	1.15	0.92	0.87	0.92	0.86
<i>q=3</i>	1.30	1.30	1.15	0.98	1.11	0.97	1.23	1.18	1.14	0.91	1.12	0.90
<i>q=4</i>	1.52	1.39	1.26	1.04	1.49	1.07	1.35	1.24	1.16	0.95	1.24	0.94
<i>q=5</i>	1.21	1.32	1.18	1.01	1.17	1.02	1.43	1.25	1.17	0.95	1.26	0.93
<i>q=n</i>			1.17	0.98					1.16	0.96		

Notes: Exercise performed with r , the number of static factors, equal to ten and without idiosyncratic; q denotes the number of dynamic factors; (t) and (r) indicate the triangular and rectangular lag window for spectral density estimation. PC and GPC refers to simple and generalized principal components and OLS and DLS indicate ordinary and dynamic least square projections.

restrictions on the dynamic structure is harmful for forecasting accuracy.

The crucial “parameter” that explains the differences between the forecasts is rather the window used for the estimation of the spectral density matrix (column

4 vs column 5 and column 6 vs and 7). The small differences in forecast accuracy between the SW and FHLR methods are hence due to a “non-core” characteristic of the procedures, rather than to the fact that the “tight factor structure” imposed by FHLR is not satisfied by the data, as claimed by Boivin and Ng (2005).

In this context it is important to assess why the choice of the window is so crucial. To address this question we examined the sub-sample performances. On the first sub-sample, forecasts obtained using the rectangular window estimator outperform those computed using the triangular window. The reverse is true for the second sub-sample. The two periods are characterized by very different persistence of the series; high during 1970s to mid-1980s, and low during mid-1980s to the end of 1990s. The triangular window, used in the estimation, gives less weight to covariances associated with longer lags. This choice is, of course, not suitable when samples are characterized by long lasting time dependence, as observed over the first sub-sample. In this case, the rectangular window, that equally weights upon covariances of all orders, is the more appropriate choice. In the second period, characterized by less persistence, putting less weight upon covariances of higher order helps, since it aligns the forecasts toward those of a random walk, which has a very good performance in this low persistent sample (see Atkeson and Ohanian, 2001; D’Agostino, Giannone, and Surico, 2006)

5 Conclusions

This paper compares alternative forecasting methodologies based on large-panel factor models. We compare the static principal component approach of Stock and Watson

(2002) and the two-step approach of Forni, Hallin, Lippi, and Reichlin (2005) in forecasting Industrial Production (IP) and Consumer Price (CPI) inflation by using a large macroeconomic dataset constructed by Stock and Watson (2002). The main results can be summarized as follows: both approaches outperform the simple univariate autoregressive model. The gain from factor based predictions is substantial, especially in periods of high comovements.

Few factors capture all the predictable components of CPI inflation and IP, while idiosyncratic dynamics are negligible.

Even when factors are estimated by putting less weight to series with larger idiosyncratic variance, there is no evident improvement in the forecast accuracy.

Lastly, taking into account the heterogeneity in the lead-lag relations among series, analyzed by nesting the Stock and Watson (2002) approach (no dynamic heterogeneity) in the Forni, Hallin, Lippi, and Reichlin (2005) approach (dynamic heterogeneity), appears irrelevant for the predictive performances.

We conclude that although there is a significant heterogeneity in the signal-to-noise ratio and in the dynamic effects of the common shocks across macroeconomic variables, accounting for this heterogeneity using the Forni, Hallin, Lippi, and Reichlin (2005) approach does not help improving the forecasting accuracy of factor based forecasts. In the last period, however, the Forni, Hallin, Lippi, and Reichlin (2005) approach slightly improves on the naive model in particular for inflation. In this case, however, the degree of shrinkage of autocovariances at longer lags is crucial for forecast precision.

References

- ATKESON, A., AND L. E. OHANIAN (2001): “Are Phillips Curves Useful for Forecasting Inflation?,” *Quarterly Review*, pp. 2–11.
- BOIVIN, J., AND S. NG (2005): “Understanding and Comparing Factor-Based Forecasts,” *International Journal of Central Banking*, 3, 117–151.
- CHAMBERLAIN, G., AND M. ROTHSCHILD (1983): “Arbitrage, Factor Structure and Mean-Variance Analysis in Large Asset Markets.,” *Econometrica*, 51, 1305–1324.
- D’AGOSTINO, A., D. GIANNONE, AND P. SURICO (2006): “(Un)Predictability and Macroeconomic Stability,” Working Paper Series 605, European Central Bank.
- FORNI, M., D. GIANNONE, M. LIPPI, AND L. REICHLIN (2005): “Opening the Black Box: Structural Factor Models with large cross-sections,” Manuscript, Université Libre de Bruxelles.
- FORNI, M., M. HALLIN, M. LIPPI, AND L. REICHLIN (2000): “The Generalized Dynamic Factor Model: identification and estimation,” *Review of Economics and Statistics*, 82, 540–554.
- (2005): “The Generalized Dynamic Factor Model: one-sided estimation and forecasting,” *Journal of the American Statistical Association*, 100, 830–840.
- FORNI, M., AND M. LIPPI (2001): “The Generalized Dynamic Factor Model: representation theory,” *Econometric Theory*, 17, 1113–1141.

- FORNI, M., AND L. REICHLIN (1998): “Let’s Get Real: A Factor Analytical Approach to Disaggregated Business Cycle Dynamics,” *Review of Economic Studies*, 65, 453–73.
- GIANNONE, D., L. REICHLIN, AND L. SALA (2004): “Monetary Policy in Real Time,” in *NBER Macroeconomics Annual*, ed. by M. Gertler, and K. Rogoff, pp. 161–200. MIT Press.
- (2006): “VARs, Factor Models and the Empirical Validation of Equilibrium Business Cycle Models,” *Journal of Econometrics*, 132, 257–279.
- GIANNONE, D., L. REICHLIN, AND D. SMALL (2005): “Nowcasting GDP and Inflation: the real-time informational content of macroeconomic data releases,” Finance and Economics Discussion Series 2005-42, Board of Governors of the Federal Reserve System (U.S.).
- SCHUMACHER, C. (2006): “Forecasting German GDP Using Alternative Factor Models Based on Large Dataset,” *Journal of Forecasting*, forthcoming.
- STOCK, J. H., AND M. W. WATSON (2002): “Forecasting Using Principal Components from a Large Number of Predictors,” *Journal of the American Statistical Association*, 97, 147–162.
- STOCK, J. H., AND M. W. WATSON (2004a): “Forecasting with Many Predictors,” Unpublished manuscript, Princeton University.
- (2004b): “Has the Business Cycle Changed and Why?,” in *NBER Macroeconomics Annual 2003*, ed. by M. Gertler, and K. Rogoff. MIT Press.