International Journal of Forecasting 30 (2014) 20–29

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

# Forecasting with approximate dynamic factor models: The role of *non-pervasive* shocks

Matteo Luciani *

*F.R.S.-FNRS, 5 Rue d'Egmont, 1000, Brussels, Belgium*
*European Center for Advanced Research in Economics and Statistics (ECARES), Solvay Business School of Economics and Management, Université libre de Bruxelles, 50 av. F.D. Roosevelt, CP 114, B1050 Brussels, Belgium*

ARTICLE INFO

*Keywords:*
Dynamic factor models
Penalized regressions
Local factors
Bayesian shrinkage
Forecasting

ABSTRACT

This paper studies the role of *non-pervasive* shocks when forecasting with factor models. To this end, we first introduce a new model that incorporates the effects of *non-pervasive* shocks, an Approximate Dynamic Factor Model with a sparse model for the idiosyncratic component. Then, we test the forecasting performance of this model both in simulations, and on a large panel of US quarterly data. We find that, when the goal is to forecast a disaggregated variable, which is usually affected by regional or sectorial shocks, it is useful to capture the dynamics generated by *non-pervasive* shocks; however, when the goal is to forecast an aggregate variable, which responds primarily to macroeconomic, i.e. *pervasive*, shocks, accounting for *non-pervasive* shocks is not useful.


## 1. Introduction

In recent years, the literature has proposed two methods for coping with the curse of dimensionality problem, namely: factor models (Forni, Hallin, Lippi, & Reichlin, 2000; Stock & Watson, 2002a) and Bayesian shrinkage (De Mol, Giannone, & Reichlin, 2008). Roughly speaking, the main idea of factor models is to *summarize* the information content of a large number of predictors in a few factors, while the idea of Bayesian shrinkage is to limit the estimation uncertainty by *shrinking* the potentially complex model toward a simple *naïve* prior model.[1]

In factor models, each variable ($x_{it}$) can be decomposed into the sum of two mutually orthogonal components, one capturing the comovement among the data ($\chi_{it}$), which is assumed to be driven by a small number of *pervasive* shocks ($\mathbf{u}_t$); and one capturing the idiosyncratic dynamics ($\xi_{it}$): $x_{it} = \chi_{it} + \xi_{it}$. Due to the strong comovement among macroeconomic time series, these models offer a realistic (and parsimonious) representation of the data. Moreover, these models can be estimated easily using the method of principal components under the assumption of "weakly" cross-sectionally-dynamically correlated idiosyncratic components (Bai, 2003; Bai & Ng, 2002; Forni et al., 2000; Forni, Hallin, Lippi, & Reichlin, 2005; Stock & Watson, 2002a), a likely feature in large macroeconomic databases where *non-pervasive* (sectorial or regional) shocks might affect groups of variables (local factors).[2]

Factor models have proved to be successful in predicting economic activity. A large body of literature has shown

* Correspondence to: European Center for Advanced Research in Economics and Statistics (ECARES), Solvay Business School of Economics and Management, Université libre de Bruxelles, 50 av. F.D. Roosevelt, CP 114, B1050 Brussels, Belgium. Tel.: +32 (0) 2 650 33 66.
*E-mail address:* matteo.luciani@ulb.ac.be.

[1] Other methods which are not used in this paper, but which are also able to forecast with large numbers of predictors, include partial least squares (Groen & Kapetanios, 2008), forecast combination (Bates & Granger, 1969), the Bayesian model average (Leamer, 1978), and bagging (Breiman, 1996). For a review of forecasting with many predictors, see Stock and Watson (2006).

[2] The literature refers to these models as *approximate* factor models, as distinct from *exact* factor models, which are characterized by cross-sectionally-dynamically uncorrelated idiosyncratic components, i.e., $\xi_{it} \sim iid(0, 1)$. For the sake of simplicity, throughout this paper we will refer to *approximate* factor models simply as "factor models".

0169-2070/$ – see front matter 
http://dx.doi.org/10.1016/j.ijforecast.2013.05.001

how factor models can outperform common univariate benchmark forecasts by simply forecasting the common component (i.e., the one driven by *pervasive* shocks), while approximating with an autoregressive model the idiosyncratic component (i.e., the one driven by *non-pervasive* shocks). Specifically, this body of literature has shown that a forecast obtained as $x_{i,t+h|t} = \chi_{i,t+h|t} + \xi_{i,t+h|t}$, where $\chi_{i,t+h|t} = Proj\{x_{i,t+h}|\mathbf{u}_t, \mathbf{u}_{t-1}, \ldots\}$ and $\xi_{i,t+h|t} = Proj\{x_{i,t+h}|\xi_{it}, \xi_{it-1}, \ldots\}$, outperforms a univariate forecast such as $x_{i,t+h|t} = Proj\{x_{i,t+h}|x_{it}, , x_{it-1}, \ldots\}$. A (non-exhaustive) list of papers that use this approach is: Artis, Banerjee, and Marcellino (2005), Bańbura and Modugno (in press), Boivin and Ng (2005, 2006), Bulligan, Golinelli, and Parigi (2010), Camacho and Perez-Quiros (2010), D'Agostino and Giannone (2012), Forni, Hallin, Lippi, and Reichlin (2003), Forni et al. (2005), Marcellino, Stock, and Watson (2003), Schumacher (2007) and Stock and Watson (2002a,b). One limitation of this approach is that it uses factor models in forecasting as if the idiosyncratic components were mutually orthogonal. Consequently, the literature to date has not taken into account the correlation induced by *non-pervasive* shocks. This is a contradiction, though, since factor models are estimated under the hypothesis of "weakly" cross-sectionally-dynamically correlated errors. The questions then arise: does taking this correlation into account lead to any forecasting gains? Is it worth accounting for *non-pervasive* shocks when forecasting with factor models? This paper answers these questions.

In this paper, we study the role of *non-pervasive* shocks when forecasting with factor models. To this end, we first introduce a new model that incorporates the effects of *non-pervasive* shocks, then test its forecasting performance both in simulations, and on a large panel of US quarterly data.

Our model augments the factor model with a sparse model for the idiosyncratic component, thus taking into account, and exploiting in forecasting, the fact that, in approximate dynamic factor models, the idiosyncratic component is "weakly" cross-sectionally-dynamically correlated. Our model produces a forecast as $x_{i,t+h|t} = \chi_{i,t+h|t} + \xi_{i,t+h|t}$, where $\chi_{i,t+h|t} = Proj\{x_{i,t+h}|\mathbf{u}_t, \mathbf{u}_{t-1}, \ldots\}$ and $\xi_{i,t+h|t} = Proj\{x_{i,t+h}|\boldsymbol{\xi}_t, \boldsymbol{\xi}_{t-1}, \ldots\}$, with $\boldsymbol{\xi}_t = [\xi_{1,t}, \xi_{2,t}, \ldots, \xi_{N,t}]'$. This forecast is obtained by mixing factor models and $L_1$ penalized regressions, which are equivalent to Bayesian shrinkage with double exponential priors, or boosting. We choose $L_1$ penalized regressions and boosting because, by performing both shrinkage and variable selection, they impose a sparse structure on the idiosyncratic component. This sparse structure is particularly appropriate for our purpose, since we are interested in capturing *non-pervasive* shocks that, by definition, affect only a limited number of variables.

The literature recently suggested a different forecasting strategy which also involves $L_1$ penalized regressions. This method was used by Bai and Ng (2008a) and De Mol et al. (2008). The former suggest extracting the factors only from those variables that are really informative for forecasting the target variable. The latter suggest selecting the predictors and estimating the model using only the selected predictors.

Although our approach uses the same method as those of Bai and Ng (2008a) and De Mol et al. (2008), it is theoretically different: while they impose a sparse structure on the whole dataset, we impose a sparse structure only on the idiosyncratic component. That is, we begin by extracting what is common, then impose sparsity on what is left.

An alternative method of accounting for *non-pervasive* shocks, used in forecasting by Bańbura, Giannone, and Reichlin (2011), involves estimating a factor model with both global and local factors using either maximum likelihood techniques (Doz, Giannone, & Reichlin, 2012) or Bayesian methods (Kose, Otrok, & Whiteman, 2008; Moench, Ng, & Potter, in press).[3] However, we do not consider this approach here, since it requires *a priori* information on the structure of the economy in order to identify *non-pervasive* shocks. In contrast, our method identifies *non-pervasive* shocks automatically by performing variable selection.

The rest of the paper is organized as follows. We illustrate our model in Section 2. In Section 3, by means of a simulation exercise, we study whether and when it is useful to account for *non-pervasive* shocks when forecasting with factor models. We test our model in Section 4 by means of a pseudo real time forecasting exercise on US quarterly data against the factor model of Forni et al. (2005), and against the methods of Bai and Ng (2008a) and De Mol et al. (2008). In Section 5, we verify the robustness of our model to the composition of the database. Finally, we present our conclusions in Section 6.

## 2. Methodology

Let $\mathbf{x}_t$ be an $N \times 1$ vector of stationary variables. Suppose that we are interested in forecasting the $i$th variable $h$ steps ahead, $x_{i,t+h|t}$, by using all $N$ potential predictors. In this case, the best linear prediction, defined as

$$x_{i,t+h|t} = Proj\{x_{i,t+h}|\Omega_t\}, \tag{1}$$

where $\Omega_t = span\{\mathbf{x}_{t-p}, p = 0, 1, \ldots\}$, might be extremely inefficient, or even impossible, due to the lack of degrees of freedom. This is the well-known curse of dimensionality problem. In recent years, the literature has suggested two solutions: factor models (Forni et al., 2000; Stock & Watson, 2002a) and Bayesian shrinkage (De Mol et al., 2008).

If the comovement of the $N$ variables in $\mathbf{x}_t$ can be approximated well by a small number $q \ll N$ of *pervasive* (or common) shocks $\mathbf{u_t}$, while the variable specific dynamics ($\boldsymbol{\xi}_t$) are only mildly correlated, then the information set can be split into two orthogonal spaces: the space spanned by the common shocks and the space spanned by the idiosyncratic components ($\Omega_t = \Omega_t^u \oplus \Omega_t^{\xi}$, where $\Omega_t^u = span\{\mathbf{u}_{t-p}, p = 0, 1, \ldots\}$, and $\Omega_t^{\xi} = span\{\boldsymbol{\xi}_{t-p}, p = 0, 1, \ldots\}$, with $\Omega_t^F \cap \Omega_t^{\xi} = \{0\}$). The idea of factor models is to approximate the linear projection on the whole information set by the sum of the linear projection on the space

[3] Hallin and Liška (2011) suggest a method involving dynamic principal components, which is able to account for *non-pervasive* shocks but cannot be used in forecasting.

spanned by the common shocks, and of the linear projection on the space spanned by the present and past values of the variable-specific dynamic:

$$x_{i,t+h|t} = Proj\{x_{i,t+h}|\Omega_t\} \approx Proj\{x_{i,t+h}|\Omega_t^u\} + Proj\{x_{i,t+h}|\Omega_{it}^\xi\}, \quad (2)$$

where $\Omega_{it}^\xi = span\{\xi_{i,t-p}, p = 0, 1, \ldots\}$. That is, factor models solve the curse of dimensionality problem by *summarizing* the information content of a large number of predictors in a few common shocks that capture the comovement in the data, and by approximating $\Omega_t^\xi = span\{\boldsymbol{\xi}_{t-p}, p = 0, 1, \ldots\}$ by $\Omega_{it}^\xi = span\{\xi_{i,t-p}, p = 0, 1, \ldots\}$.

The idea of Bayesian shrinkage is to limit the estimation uncertainty by *shrinking* the potentially complex model toward a simple *naïve* prior model. Specifically, Bayesian shrinkage solves the curse of dimensionality by

$$x_{i,t+h|t} = Proj_s\{x_{i,t+h}|\Omega_t\}, \quad (3)$$

where $Proj_s$ means linear projection with *shrinkage* of the parameters, and $\Omega_t = span\{\mathbf{x}_{t-p}, p = 0, 1, \ldots\}$. Bayesian shrinkage is strictly related to penalized regressions. In particular, Bayesian shrinkage with Gaussian priors is equivalent to $L_2$ penalized regressions, also known as ridge regressions; while Bayesian shrinkage with double exponential priors is equivalent to $L_1$ penalized regressions, also known under the labels of Lasso Regressions, or Least Angle Regressions (LARS).[4] Moreover, De Mol et al. (2008) show that both $L_1$ and $L_2$ penalized regressions are intimately related to factor models. On the one hand, when the data are highly correlated, as is the case when they have a factor structure, a few variables selected with an $L_1$ regression are able to capture the bulk of the comovement in the data; on the other hand, $L_2$ penalized regressions and factor models both produce linear combinations of all of the variables in the dataset, and differ only in the weighting scheme.

Factor models have proved successful at predicting economic activity (for a review, see Eickmeier & Ziegler, 2008). However, these models are used in forecasting as if the idiosyncratic components were mutually orthogonal. That is, factor models ignore the correlations among idiosyncratic components which are induced by *non-pervasive* (i.e. local) shocks. This depends on Eq. (2), according to which the forecast of the idiosyncratic component is obtained by projecting only onto the space spanned by the *variable specific* dynamic ($\xi_{it}$). Let us give an example in order to clarify this point.

Suppose that we have a panel of $N$ variables. Suppose that each variable in the panel is driven by one common shock, $u_t \sim iid(0, 1)$, one variable-specific shock, $e_{it} \sim iid(0, 1)$, and (for only the first $n \ll N$) one *non-pervasive* shock, $v_t \sim iid(0, 1)$:

$$x_{it} = (1 + \lambda_i L)u_t + (1 + \gamma_i L)v_t + (1 + \rho_i L)e_{it}, \quad i = 1, \ldots, n,$$
$$x_{it} = (1 + \lambda_i L)u_t + (1 + \rho_i L)e_{it}, \quad i = n + 1, \ldots, N,$$

where $\chi_{it} = (1 + \lambda_i L)u_t$ is the common component and $\xi_{it} = x_{it} - \chi_{it} = (1 + \gamma_i L)v_t + (1 + \rho_i L)e_{it}$ for $i \le n$, while $\xi_{it} = (1 + \rho_i L)e_{it}$ for $i > n$ is the idiosyncratic component.

Let $k$ be an integer smaller than $n$, and suppose that the goal is to forecast $x_{k,t+1}$. With a factor model, the forecast is obtained as $x_{k,t+1|t} = Proj(x_{k,t+1}|u_t, u_{t-1}) + Proj(x_{k,t+1}|\xi_{kt}) = Proj(\chi_{k,t+1}|u_t, u_{t-1}) + Proj(\xi_{k,t+1}|\xi_{kt})$, which makes use of the approximation $E(\xi_{kt}\xi_{it-1}) = 0$. Note, however, that this approximation is not correct, since, when $i \le n$, $E(\xi_{kt}\xi_{it-1}) = \gamma_k$, due to the effect of the *non-pervasive* shock $v_t$. Therefore, a better forecast should be obtained as $x_{k,t+1|t} = Proj(x_{k,t+1}|u_t, u_{t-1}) + Proj(x_{k,t+1}|\xi_{jt}, j = 1, \ldots, n)$; that is, by projecting onto the space spanned by all of the idiosyncratic components that are affected by the shock $v_t$.

Practically, running such a linear projection involves two main problems. First, unless we have *a priori* information that helps to identify *non-pervasive* shocks, and unless we want to impose restrictions which are derived from this information, how can we determine which variables to include in the linear projection? In other words, how can we identify the variables that are driven by the same *non-pervasive* shock that drives the target variable? Second, suppose that $n$ (the number of variables affected by the *non-pervasive* shock) is large. In this case, the linear projection might be extremely inefficient, or even impossible, due to a lack of degrees of freedom.

In this section, we introduce a new model which is capable of solving these problems, and hence of incorporating the effects of *non-pervasive* shocks. In the following sections, we will test the forecasting performance of this model in order to understand whether and when it is useful to account for *non-pervasive* shocks when forecasting with factor models.

Our model augments the factor model with a sparse model for the idiosyncratic component. Thus, it takes into account, and exploits in forecasting, the fact that, in approximate dynamic factor models, the idiosyncratic component is "weakly" cross-sectionally-dynamically correlated. Our approach consists of combining factor models and shrinkage. Our idea is to capture the bulk of the comovement in the data via the factor model, and to capture the local correlation via shrinkage, thus:

$$x_{i,t+h|t} = Proj\{x_{i,t+h}|\Omega_t^u\} + Proj_s\{x_{i,t+h}|\Omega_t^\xi\}$$

where $Proj\{x_{i,t+h}|\Omega_t^u\}$ is the linear projection onto the space spanned by the common shocks, and $Proj_s\{x_{i,t+h}|\Omega_t^\xi\}$ is the linear projection with shrinkage onto the space spanned by the idiosyncratic components. In practice, we estimate $Proj\{x_{i,t+h}|\Omega_t^u\}$ by the two-step procedure of Forni et al. (2005), while we estimate $Proj_s\{x_{i,t+h}|\Omega_t^\xi\}$ by $L_1$ penalized regressions or boosting (Freund & Schapire, 1997).[5]

We suggest that the $L_1$ penalty and boosting represent the appropriate penalization. In contrast to the $L_2$ penalty,

[4] To clarify, Lasso (Least Absolute Shrinkage and Selection Operator) and LARS (Least Angle Regressions) are two algorithms proposed by Efron, Hastie, Johnstone, and Tibshirani (2004) and Tibshirani (1996) for estimating $L_1$ penalized regressions.

[5] In this paper, we use the LARS algorithm (Efron et al., 2004, LARS) to estimate $L_1$ penalized regressions, while we use the *component-wise* algorithm suggested by Bühlmann and Yu (2003) to estimate boosting.

which performs shrinkage and the aggregation of variables, the $L_1$ penalty and boosting perform both shrinkage and the selection of variables, thus imposing a sparse structure on the idiosyncratic component. This sparse structure is particularly appropriate for our purpose, since we are interested in capturing *non-pervasive* shocks, which, by definition, are related only to a limited number of variables.

Our approach is similar to factor models in that it summarizes the comovement in the data with a few common shocks; it differs from factor models in that it considers the whole space spanned by the idiosyncratic component, rather than approximating it by the present and past values of the variable-specific dynamics. In common with Bayesian shrinkage, our approach considers the whole information set spanned by $\mathbf{x}_{t-p}$ ($\Omega_t$), but, in contrast to Bayesian shrinkage, our approach exploits the fact that $\Omega_t$ can be split into two orthogonal spaces ($\Omega_t = \Omega_t^u \oplus \Omega_t^\xi$).

## 3. Simulations

In this section, we use a Monte Carlo experiment to investigate whether and when it is useful to account for *non-pervasive* shocks when forecasting with factor models. To this end, we will compare four models: (1) the original model of Forni et al. (2005), $FHLR_1$, which forecasts only the common component; (2) the extension of Forni et al. (2005), proposed by D'Agostino and Giannone (2012), $FHLR_2$, which forecasts the idiosyncratic component with an autoregressive model; (3) our first proposal, $ML_1$, which forecasts the idiosyncratic component with $L_1$ penalized regressions; and (4) our second proposal, $ML_2$, which forecasts the idiosyncratic component with boosting. In all four models, the forecast of the common component is obtained by the two-step method of Forni et al. (2005).

The model from which we simulate is a dynamic factor model with $q$ common shocks, $\mathbf{u}_t$; $r = q(s+1)$ common factors, $\mathbf{f}_t$; and cross-sectionally and serially correlated idiosyncratic components, $\boldsymbol{\xi}_t$.[6] Each idiosyncratic component is the sum of two shocks, a *non-pervasive* shock, $v_{lt}$, which is loaded according to an autoregressive scheme, and an idiosyncratic shock, $e_{it}$, which is cross-sectionally correlated, i.e., $E(e_{it}, e_{jt}) \neq 0$. There are $b$ *non-pervasive* shocks, one for each block of variables (all blocks are of equal size), and $N$ idiosyncratic shocks, one for each variable.

Let $t = 1, \ldots, T$, $i = 1, \ldots, N$, $j = 1, \ldots, q$, $k = 1, \ldots, s$, and $l = 1, \ldots, b$; then, the model used for simulations is defined as

$$\mathbf{x}_t = \boldsymbol{\chi}_t + \boldsymbol{\xi}_t$$

$$\boldsymbol{\chi}_t = \sum_{k=0}^{s} \boldsymbol{\Lambda}_k \mathbf{f}_{t-k}$$

$$f_{jt} = \phi_j f_{jt-1} + u_{jt} \quad u_{jt} \sim \mathcal{N}(0,1)$$

$$\xi_{it} = \boldsymbol{\Gamma}_{\mathbf{i}} \mathbf{g}_t + e_{it} \quad e_{it} \sim \mathcal{N}(0, \sigma_i^2),$$

$$\sigma_i^2 \sim \mathcal{U}(0.1, 1.1), \qquad E(e_{it}, e_{jt}) = \tau^{|i-j|}$$

$$g_{lt} = \psi_l g_{lt-1} + v_{lt} \quad v_{lt} \sim \mathcal{N}(0,1),$$

where $\Lambda_{i,j,k} \sim N(0,1)$, $\tau = 0.5$, $\Gamma_{il} \sim N(0,1)$ if the $i$th variable belongs to block $l$, and $\Gamma_{il} = 0$ otherwise, $q = 2$, $s = 1$ (hence $r = 4$), and $b = 10$. Moreover, although we set $\sigma_i^2 \sim \mathcal{U}(0.1, 1.1)$ to introduce some heterogeneity into the size of the idiosyncratic shocks, the common component is calibrated so that it accounts for 50% of the total variance on average, and, similarly, local factors, $\mathbf{g}_t$, are calibrated so that they account for 50% of the variance of the idiosyncratic component on average.

We generate the data both for different values of the parameter governing the persistence of the common shocks ($\phi_j = \{0.2, 0.5, 0.8\}$), and for different values of the parameter governing the persistence of the *non-pervasive* shocks ($\psi_l = \{0.2, 0.4, 0.6, 0.8\}$). Therefore, when $\phi_j = \{0.5, 0.8\}$ and $\psi_l = \{0.2, 0.4\}$, both the bulk of the comovement and the bulk of the dynamics in the data are accounted for by common shocks; whereas when $\phi_j = 0.2$ and $\psi_l = \{0.6, 0.8\}$, the bulk of the dynamics is accounted for by *non-pervasive* shocks.[7]

Following Boivin and Ng (2005), we simulate 1000 panels of dimensions $T = 84$ and $N = 104$. For each panel, we take the first 80 observations, which corresponds to a rolling window of 20 years when considering quarterly data, and produce one-step-ahead forecasts. We always forecast the first three variables ($x_1, x_2, x_3$), for which the common components account for 75%, 50%, and 25% of the total variance, respectively; i.e., we always forecast a variable that is strongly affected by common shocks ($x_1$), a variable that is affected by common and idiosyncratic shocks equally ($x_2$), and a variable that is only slightly affected by common shocks ($x_3$).

We assume that both the number $q$ of common shocks and the number $r = q(s+1)$ of common factors is known. For model $FHLR_2$, the forecast of $\xi_{i,t+h|t}$ is obtained by projecting $\xi_{i,t+h}$ on $x_t$. For model $ML_1$, we stop the LARS algorithm at the $2r^{\text{th}}$ iteration, while for model $ML_2$ we determine which iteration to retain using the modified BIC suggested by Bai and Ng (2009). Throughout the paper, we use the method of direct forecasting (Stock & Watson, 2002b), meaning that $x_{i,t+h}^h = \sum_{j=1}^h x_{i,t+j}$.

Table 1 shows results of the simulation exercise. The first two columns report the values of $\psi$ and $\phi$ which are used to generate the data, while from the third column onward, each cell reports the mean squared error of model $m_1$ relative to that of model $m_2$ ($\frac{m_1}{m_2}$). An entry lower than 1 means that model $m_1$ beats model $m_2$, while an entry greater than 1 means that model $m_1$ performs worse than model $m_2$.

Columns 3–5 of Table 1 show the relative mean squared errors of the model originally proposed by Forni et al.

6 Similar models have been used by Boivin and Ng (2005), Doz et al. (2012), Forni et al. (2005) and Stock and Watson (2002a).

7 In a Supplementary Appendix (available upon request), we show simulations for different numbers of blocks, for other parameter configurations, and for the four-step-ahead forecasting horizon. Specifically, we show results when there are only two *non-pervasive* shocks, each affecting half of the variables. This is an interesting case, since the assumption of weakly cross-correlated idiosyncratic components is no longer satisfied when the number of variables affected by the *non-pervasive* shock either grows with $N$, or is very large. The results show that the performances of our models deteriorate in this case, becoming comparable to that of the model of Forni et al. (2005) with an extra factor.

**Table 1**
Relative mean squared errors of one-step-ahead forecasts from Monte Carlo simulations.

| $\psi$ | $\phi$ | $\frac{FHLR_1}{AR}$ | | | $\frac{FHLR_2}{FHLR_1}$ | | | $\frac{ML_1}{FHLR_1}$ | | | $\frac{ML_2}{FHLR_1}$ | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ |
| 0.20 | 0.20 | 0.71 | 0.78 | 0.92 | 1.00 | 1.00 | 1.01 | 1.01 | 1.02 | 1.02 | 1.03 | 1.05 | 1.05 |
| 0.20 | 0.50 | 0.70 | 0.80 | 0.94 | 1.00 | 1.00 | 0.99 | 1.01 | 1.01 | 1.02 | 1.01 | 1.03 | 1.03 |
| 0.20 | 0.80 | 0.76 | 0.80 | 0.89 | 1.00 | 1.00 | 1.01 | 1.02 | 1.00 | 1.02 | 1.01 | 1.05 | 1.07 |
| 0.40 | 0.20 | 0.70 | 0.81 | 0.95 | 0.99 | 0.99 | 1.00 | 1.01 | 1.02 | 1.03 | 1.02 | 1.03 | 1.03 |
| 0.40 | 0.50 | 0.71 | 0.84 | 0.93 | 1.00 | 0.99 | 0.98 | 1.01 | 1.03 | 1.02 | 1.01 | 1.03 | 1.05 |
| 0.40 | 0.80 | 0.76 | 0.86 | 0.93 | 1.00 | 1.00 | 0.98 | 1.01 | 1.03 | 1.02 | 1.02 | 1.00 | 1.04 |
| 0.60 | 0.20 | 0.71 | 0.85 | 0.96 | 0.99 | 0.99 | 0.97 | 1.02 | 1.01 | 1.01 | 0.98 | 0.96 | 0.95 |
| 0.60 | 0.50 | 0.69 | 0.86 | 1.02 | 0.99 | 0.97 | 0.96 | 1.02 | 1.00 | 1.01 | 1.00 | 0.97 | 0.99 |
| 0.60 | 0.80 | 0.80 | 1.00 | 0.97 | 0.98 | 0.96 | 0.99 | 1.00 | 1.02 | 1.02 | 0.99 | 0.98 | 0.99 |
| 0.80 | 0.20 | 0.69 | 0.91 | 1.06 | 0.98 | 0.95 | 0.94 | 1.00 | 1.01 | 1.01 | 0.94 | 0.89 | 0.88 |
| 0.80 | 0.50 | 0.79 | 0.90 | 1.08 | 0.98 | 0.97 | 0.94 | 1.01 | 1.02 | 1.00 | 0.98 | 0.94 | 0.85 |
| 0.80 | 0.80 | 0.98 | 1.00 | 1.05 | 0.97 | 0.96 | 0.94 | 1.02 | 0.98 | 1.01 | 0.95 | 0.93 | 0.91 |

Notes. Each cell is the mean squared error of model $m_1$ relative to that of model $m_2$ ($\frac{m1}{m2}$). An entry lower than 1 means that model $m_1$ beats model $m_2$, while an entry greater than 1 means that model $m_1$ does worse than model $m_2$.

(2005), $FHLR_1$, versus a benchmark AR(1) model. The results show that, with a few exceptions, the model of Forni et al. (2005) outperforms the benchmark AR. These exceptions occur either when common shocks account for a small share of the total variance (column 5), or when *non-pervasive* shocks are very persistent ($\psi = 0.8$).

Columns 6–8 show the relative mean squared errors of the model of Forni et al. (2005) augmented with the forecast of the idiosyncratic component obtained as suggested by D'Agostino and Giannone (2012), $FHLR_2$, versus the model originally proposed by Forni et al. (2005). As we can see, forecasting the idiosyncratic component using its own past provides no advantage, unless the *non-pervasive* shock is persistent ($\psi = \{0.6, 0.8\}$).

Finally, columns 9–14 show the relative mean squared errors of our models, $ML_1$ and $ML_2$, against the original proposal of Forni et al. (2005). Overall, accounting for *non-pervasive* shocks provides no forecasting gains unless these shocks are strong enough to produce predictable dynamics ($\psi = \{0.6, 0.8\}$). In addition, $ML_2$ performs better than $ML_1$, suggesting that boosting is capable of capturing the effects of *non-pervasive* shocks better than $L_1$ penalized regressions can.[8]

In summary, the simulation exercise shows that if there are *non-pervasive* shocks that affect a block of variables, and if these shocks induce sufficiently strong dynamics, then our model can improve on the original proposal of Forni et al. (2005). Moreover, we show that, in this case, our model also performs better than the model of D'Agostino and Giannone (2012), meaning that the usual strategy of not modeling the idiosyncratic component, or simply forecasting it using an AR model, is dominated by the strategy suggested in this paper.

In light of these results, we expect that the usefulness of the model proposed here will depend on which variable the researcher is interested in forecasting. If the target variable is likely to be affected by *non-pervasive* (such as sectorial or regional) shocks which affect a group of variables in the database, then our model can provide gains in terms of forecasting performances; otherwise, we expect it to perform as well as the original proposal of Forni et al. (2005).

## 4. Empirics

In this section, we use a large panel of US quarterly data to study the question of whether it is useful to account for *non-pervasive* shocks when forecasting with factor models.[9] We evaluate the forecasting performance of our model by comparing four forecasting approaches: (1) factor models (Forni et al., 2005), $FHLR_1$;[10] (2) factor models and selection of predictors (this paper), $ML_1$, and $ML_2$; (3) factor models estimated on selected predictors (Bai & Ng, 2008a), $BN_1$ and $BN_2$; and (4) selection of predictors (De Mol et al., 2008), $DGR_1$ and $DGR_2$.

Of these, models $FHLR_1$, $ML_1$, and $ML_2$ were introduced in the previous section.

Models $BN_1$ and $BN_2$ are equivalent to model $FHLR_1$, except that the factors are extracted from selected predictors. These two models differ simply in the way in which

8 As a robustness check, we investigate whether our model fails to improve on that of Forni et al. (2005) simply because the idiosyncratic component is estimated poorly. For this purpose, we compute the unfeasible forecast, where the term "unfeasible" is used to mean that we forecast the common and idiosyncratic components by using the simulated variables. The results are provided in the Supplementary Appendix but can be summarized as follows: the unfeasible forecast of $FHLR_1$ is much better than the feasible one; however, the ratios $\frac{ML_1}{FHLR_1}$ and $\frac{ML_2}{FHLR_1}$ are the same as those in Table 1. This suggests that the problem lies not in the estimation of the idiosyncratic component, but rather in the persistence of *non-pervasive* shocks.

9 The database consists of 104 quarterly series describing the US economy. The variables cover 12 different categories: Industrial Production, Consumer Price Indexes, Producer Price Indexes, Monetary Aggregates, Banking, GDP and Components, Housing Sector, Productivity and Cost, Interest Rates, Employment and Population, Survey, and Financial Markets. All of the variables are transformed to reach stationarity and then standardized to have a zero mean and unit variance, thus preventing possible scale effects when extracting the factors. The complete list of the variables, sources, and transformations used is available upon request.

10 Stock and Watson (2002a) suggest a method for forecasting with factor models which differs slightly from that of Forni et al. (2005) (for a comparison of the two methods, see D'Agostino & Giannone, 2012). The results obtained using the method of Stock and Watson (2002a) are not significantly different from those obtained using the method of Forni et al. (2005), and are available upon request.

the predictors are selected: for model $BN_1$, the predictors are selected with $L_1$ penalized regressions (52 variables); while for model $BN_2$, the predictors are selected via "economic judgment" (25 variables), an approach which has also been adopted by Bańbura, Giannone, and Reichlin (2010) and Bańbura et al. (2011), both in forecasting with large Bayesian VARs and in nowcasting with factor models.[11]

Models $DGR_1$ and $DGR_2$ differ from the other models. Both of these models select predictors and estimate the forecast only on the selected predictors, without assuming any factor structure. For model $DGR_1$, the forecast of the $i$th variable is computed using $L_1$ penalized regressions, while in model $DGR_2$, the forecast of the $i$th variable is computed using boosting.

For each model, we select a benchmark specification that we will use for comparisons throughout the paper. For models $FHLR_1$, $ML_1$, $ML_2$, and $BN_1$, our benchmark specification includes two common shocks ($q = 2$) and four common factors ($r = 4$), while for model $BN_2$ it includes one common shock and two common factors. Finally, for models $ML_1$ and $DGR_1$, we stop the LARS algorithm at the $2r^{\text{th}}$ iteration, while for models $ML_2$ and $DGR_2$ we determine which iteration to retain using the modified BIC suggested by Bai and Ng (2009).[12]

The forecasting exercise is performed on six variables: GDP (*GDPC*), Residential Investment (*PRFIC*), Non-Residential Investment (*PNFIC*), Consumption of Non-Durable Goods (*PCNDGC*), Consumption of Durable Goods (*PCDGCC*), and Consumption of Services (*PCESVC*).

Forecasts are produced according to a rolling scheme by using observations for the last 20 years at each point in time (80 observations): the first estimation is carried out on a sample from 1960:3 to 1980:2, while the last estimation uses a sample from 1990:4 to 2010:3. For each variable and each method, we produced one- and four-step-ahead forecasts. Overall, we produced 122 one-step-ahead forecasts (1980:3–2010:4) and 119 four-step-ahead forecasts (1981:2–2010:4).

Table 2 reports relative mean squared errors, which are computed relative to a benchmark AR(1) model. Overall, all models beat the benchmark forecast, particularly when forecasting *GDPC* and *PNFIC*; when forecasting consumption variables, they do nearly as well as the benchmark AR.

With respect to our research question, the results in Table 2 confirm those of the simulation exercise: $ML_2$ performs better than $FHLR_1$ when forecasting *PRFIC*, while it does as well as $FHLR_1$ when forecasting other variables. In other words, Table 2 shows that, when forecasting GDP or consumption, i.e., aggregate variables which respond primarily to macroeconomic (*pervasive*) shocks, forecasting the idiosyncratic component is not useful; however, it also shows that our model improves on the forecasts from a factor model when forecasting residential investment (a variable that is strongly affected by regional and sectorial shocks; see Luciani, in press; Moench & Ng, 2011; Stock & Watson, 2009), thus showing that capturing the dynamics generated by *non-pervasive* shocks is useful.

These results show that the model introduced in this paper may be useful for practical purposes; for example, when forecasting many variables at the same time, of which some are more dependent on *non-pervasive* shocks and others are less so. In this case, our model does as well as a factor model when forecasting aggregate variables, while it does better than a factor model when forecasting non-aggregated variables.

Three additional results can be inferred from Table 2. First, the forecasting performances of $ML_1$ and $ML_2$ are nearly identical to that of $BN_1$. Recall that $BN_1$ first selects predictors and then extracts factors, while we first extract factors and then select predictors. Table 2 suggests that these two procedures seem to be equivalent, in terms of forecasting performances. Second, $BN_2$ does very well in forecasting *GDPC*, while it does particularly badly when forecasting other variables. This result confirms those of Bańbura et al. (2010, 2011), who show that, in forecasting GDP, medium-size models (i.e., including 10–30 variables) perform as well as large models (about 100 variables). Moreover, this result also suggests that aggregate variables may suffice to produce a very good forecast when forecasting GDP, while disaggregated information is fairly important when forecasting more disaggregated variables. Third, although $DGR_1$ and $DGR_2$ perform particularly well when forecasting *PRFIC* and *GDPC*, they do as well as factor models overall. This result is a consequence of the factor structure in the data, and indicates that all of these models essentially capture the same information.[13] This result is in line with those of De Mol et al. (2008), who show that, if data are highly collinear, as is the case when they have a factor structure, the factors and a few appropriately selected variables (such as those selected with $L_1$ regressions or boosting) contain essentially the same information.

All in all, these results indicate that, although there is some cross-correlation due to *non-pervasive* shocks in the idiosyncratic component, accounting for this rarely boosts the forecasting accuracy. This indicates that the factor structure is particularly strong in the data and that common factors capture not only the comovement, but also the bulk of the dynamics in the data.

## 5. Are more data always harmful for factor forecasting?

To test the robustness of our model, we investigate whether its performance depends on the composition of the dataset. That is, we evaluate how our model is affected by increased cross-correlations in the idiosyncratic component.

11 Selection using "economic judgment" essentially amounts to the inclusion of only aggregate variables in the model.

12 We report the results of the tests and criteria used to determine the number of factors for the benchmark specifications in the Supplementary Appendix, together with the results obtained with different parameter configurations.

13 This conclusion is also confirmed when looking at the correlations between the forecasts and at the forecasting performances of the estimated models over time.

**Table 2**
Relative mean squared errors.

| | $h$ | $FHLR_1$ | $ML_1$ | $ML_2$ | $BN_1$ | $BN_2$ | $DGR_1$ | $DGR_2$ |
|---|---|---|---|---|---|---|---|---|
| *GDPC* | 1 | 0.83 | 0.81 | 0.8 | 0.81* | 0.64** | 0.81** | 0.78** |
| | 4 | 0.86 | 0.85 | 0.88 | 0.87 | 0.85 | 0.9 | 0.84 |
| *PRFIC* | 1 | 1.07 | 0.9 | 0.87 | 0.86 | 1.23** | 0.60** | 0.56** |
| | 4 | 1.03 | 1 | 0.97 | 0.98 | 1.15 | 1.09 | 0.99 |
| *PNFIC* | 1 | 0.83* | 0.85 | 0.88 | 0.96 | 0.94 | 0.96 | 0.93 |
| | 4 | 0.71** | 0.74** | 0.75** | 0.78 | 0.91 | 0.98 | 0.9 |
| *PCNDGC* | 1 | 0.92 | 0.91 | 0.92 | 0.94 | 1.05 | 0.92 | 0.94 |
| | 4 | 0.92 | 0.89 | 0.89 | 0.96 | 1.06 | 0.95 | 0.97 |
| *PCESVC* | 1 | 1.07 | 1.07 | 1.13 | 0.93 | 1.31** | 1.12 | 1.02 |
| | 4 | 1.03 | 0.98 | 1.01 | 0.93 | 1.21 | 1.02 | 0.89 |
| *PCDGCC* | 1 | 0.96 | 0.9 | 0.89 | 0.94 | 1.17 | 1.03 | 1.12 |
| | 4 | 0.96 | 0.94 | 0.95 | 0.87 | 1.18 | 0.84 | 0.8 |

Notes. Each cell reports the relative mean squared error of the $m$th model compared to an AR(1) benchmark forecast. An entry lower than 1 means that the $m$th model beats the benchmark forecast, while an entry greater than 1 means that model $m$ does worse than the benchmark model. Asterisks indicate a rejection of the test of equal predictive accuracy between each model and the benchmark model at 5% (**) or 10% (*) significance levels. The test used here is the test of "unconditional predictive ability" of Giacomini and White (2006), which is equivalent to the Diebold and Mariano (1995) test statistic. When using rolling window estimations, this test statistics has a standard normal limit distribution.

This robustness check also links this paper to another open issue with factor models, namely the question of whether excess cross-correlation might harm the estimation of the factors (Bai & Ng, 2008b). Indeed, although the literature has shown that asymptotically "weakly" cross-sectionally-dynamically correlated errors do not affect the factor estimates, this will not necessarily be true for empirical applications. The simulations of Boivin and Ng (2006) show that, as the cross-correlation among the idiosyncratic errors increases, the estimation and forecasting performance of the model deteriorates. Onatski (2012) shows that if the explanatory power of the factors does not strongly dominate the explanatory power of the idiosyncratic terms, meaning that *pervasive* and *non-pervasive* shocks cannot be distinguished clearly, then the principal component estimator is inconsistent. In addition, the simulation results of Bai and Ng (2008b) confirm that the factor estimates can be severely compromised in such situations.

The construction of a database is a practical problem for which there is no recipe, and one which practitioners unavoidably face when forecasting with factor models. In principle, given that factor estimates are consistent as $N \to \infty$, including all available variables is a natural choice. However, as was pointed out by Boivin and Ng (2006), if adding an extra variable does not add information about the factors, but instead it simply adds an extra cross-correlation among idiosyncratic errors, the estimates of the factors will deteriorate. This is, in fact, a concrete possibility. Consider, for example, Producer Price Indexes (PPI). Our dataset includes six PPIs, but we have the option to retreive data for more than fifteen PPIs from the FRED database. We include only six PPIs because including them all would increase the cross-correlation among the idiosyncratic errors without adding any information regarding the factors that is not already contained in the six selected PPIs.

To perform this exercise, we mix up the dataset by adding many artificial variables, with the goal of mimicking a practical situation like the one just described. We increase the number of variables in the dataset by copying all producer price indexes (6 variables), and all consumer price indexes (7 variables). To avoid perfect collinearity, every time we copy one variable, we add some noise to this variable: $x^a_{it} = x_{it} + v_{it}$, where the superscript $a$ stands for "artificial", and $v_{it}$ is an error term. We do this four times, meaning that we add 52 variables overall, and specify a different error term each time:

1. $v_{it} \sim N(0, \frac{1}{4}\sigma_i^2)$;
2. $v_{it} = 0.2v_{it-1} + \epsilon_{it}$, with $\epsilon_{it} \sim N(0, \frac{1}{5}\sigma_i^2)$;
3. $v_{it} = \epsilon_{it} + 0.2\epsilon_{it-1}$, with $\epsilon_{it} \sim N(0, \frac{1}{5}\sigma_i^2)$;
4. $v_{it} = 0.2v_{it-1} + \epsilon_{it} + 0.2\epsilon_{it-1}$, with $\epsilon_{it} \sim N(0, \frac{1}{10}\sigma_i^2)$;

where $\sigma_i^2 = var(x_{it})$.

By running this procedure, we are sure that we are not adding any information, only noise.[14] As a result of the increased noise, the cross-correlation between the idiosyncratic components is much higher in the augmented database than in the benchmark database.[15] According to Boivin and Ng (2006), the performance of our model on the augmented database should deteriorate consistently.

A first consequence of the increased cross-correlation is that the tests and criteria for determining the number of factors detect higher numbers of factors. In the benchmark database, the Onatski (2009) test and the $IC_1$ criteria of Bai and Ng (2002) detect two common shocks and four common factors, respectively. On the other hand, in the augmented database, the Onatski (2009) test does not find any common shocks, while $IC_1$ detects six common

[14] In addition, this method also skews the database heavily towards nominal variables: in this augmented database, 41% of the variables are price indexes.

[15] In the Supplementary Appendix we show that more than 25% of the variables in the augmented database have the highest (in absolute value) cross-correlation coefficient, which is greater than 0.87, while less than 1% of the variables in the benchmark database do so. Furthermore, more than 25% of the idiosyncratic components in the augmented database have the second highest (in absolute value) cross-correlation coefficient, which is greater than 0.8, while less than 1% in the benchmark database do so.

**Table 3**
Relative mean squared error augmented database.

| | $h$ | $FHLR_1$ | $ML_1$ | $ML_2$ | $BN_1$ | $BN_2$ | $DGR_1$ | $DGR_2$ |
|---|---|---|---|---|---|---|---|---|
| $q = 4, r = 6$ | | | | | | | | |
| *GDPC* | 1 | 0.90 | 0.87 | 0.88 | 0.95 | 0.64 | 0.79 | 0.78 |
| | 4 | 0.96 | 0.96 | 0.97 | 0.82 | 0.85 | 0.87 | 0.84 |
| *PRFIC* | 1 | 1.24 | 0.97 | 0.96 | 0.92 | 1.23 | 0.57 | 0.56 |
| | 4 | 1.20 | 1.09 | 1.07 | 0.91 | 1.15 | 1.01 | 0.99 |
| *PNFIC* | 1 | 0.94 | 0.90 | 0.94 | 1.01 | 0.94 | 0.95 | 0.94 |
| | 4 | 0.88 | 0.85 | 0.86 | 0.83 | 0.91 | 0.95 | 0.90 |
| *PCNDGC* | 1 | 1.04 | 1.02 | 1.03 | 1.05 | 1.05 | 0.90 | 0.93 |
| | 4 | 1.02 | 0.95 | 0.93 | 0.88 | 1.06 | 0.95 | 0.97 |
| *PCESVC* | 1 | 1.19 | 1.17 | 1.22 | 1.14 | 1.31 | 1.09 | 1.05 |
| | 4 | 1.09 | 1.06 | 1.06 | 1.03 | 1.21 | 0.95 | 0.90 |
| *PCDGCC* | 1 | 1.04 | 1.04 | 1.07 | 1.10 | 1.17 | 1.05 | 1.12 |
| | 4 | 1.01 | 0.98 | 0.98 | 0.88 | 1.18 | 0.83 | 0.81 |
| $q = 2, r = 4$ | | | | | | | | |
| *GDPC* | 1 | 0.89 | 0.84 | 0.88 | 0.80 | 0.64 | 0.83 | 0.78 |
| | 4 | 1.02 | 0.99 | 1.00 | 0.80 | 0.85 | 0.92 | 0.84 |
| *PRFIC* | 1 | 1.33 | 0.94 | 0.94 | 0.82 | 1.23 | 0.60 | 0.56 |
| | 4 | 1.24 | 1.11 | 1.11 | 0.97 | 1.15 | 1.09 | 0.99 |
| *PNFIC* | 1 | 0.80 | 0.81 | 0.84 | 0.90 | 0.94 | 0.96 | 0.94 |
| | 4 | 0.78 | 0.78 | 0.78 | 0.84 | 0.91 | 0.98 | 0.90 |
| *PCNDGC* | 1 | 1.04 | 1.06 | 1.11 | 0.94 | 1.05 | 0.90 | 0.93 |
| | 4 | 1.08 | 1.02 | 0.98 | 0.98 | 1.06 | 0.95 | 0.97 |
| *PCESVC* | 1 | 1.17 | 1.14 | 1.20 | 1.03 | 1.31 | 1.12 | 1.05 |
| | 4 | 1.09 | 1.06 | 1.08 | 0.92 | 1.21 | 1.02 | 0.90 |
| *PCDGCC* | 1 | 1.01 | 0.98 | 1.00 | 1.03 | 1.17 | 1.03 | 1.12 |
| | 4 | 1.07 | 1.04 | 1.11 | 0.98 | 1.18 | 0.85 | 0.81 |

Notes. Each cell reports the relative mean squared error of the $m$th model, compared to an AR(1) benchmark forecast. An entry lower than 1 means that the $m$th model beats the benchmark forecast, while an entry greater than 1 means that model $m$ does worse than the benchmark model.

factors.[16] However, we can be certain that the number of factors has not increased, since we have artificially added only noise.

Table 3 shows results of the forecasting exercise performed on the augmented database. Although we know that this is wrong, in the top panel of Table 3 we assume that there are four common shocks and six common factors, while in the bottom panel we *cheat* by exploiting our superior information and set $q = 2$ and $r = 4$. As we can see, the MSEs of $FHLR_1$, $ML_1$, and $ML_2$ increase slightly, confirming that, when there is noise in the data, meaning excess correlation between the idiosyncratic components, the factors are poorly estimated (Boivin & Ng, 2006). However, the MSE of $ML_1$ and $ML_2$ increases by less than that of $FHLR_1$ (particularly when $q = 4$ and $r = 6$, which is the relevant empirical case). This result indicates that our model is more robust both to the composition of the database, and to mis-specification of the number of factors, than the model of Forni et al. (2005). Moreover, mis-specification of the number of factors is problematic, but not as problematic as the noise in the data. Indeed, the MSEs reported in the bottom panel of Table 3 are lower than those reported in the top panel, but very similar. Finally, the forecasting performances of $BN_1$, $DGR_1$ and $DGR_2$ are unchanged, indicating that selection of predictors is not affected by noise.

In summary, in this section we have performed a robustness check of our model with respect to the composition of the database. Database construction is a practical problem for researchers. Due to inexperience, or a lack of data availability, he may construct an unbalanced database in which one sector, or one category, is over-represented. We have shown that this is a critical situation, since tests and criteria for determining the numbers of factors are extremely unreliable when the database is poorly constructed, often resulting in either an underestimation or an overestimation of the number of factors. The results of our robustness check showed that our model is more robust, both to the composition of the database and to mis-specification of the number of factors, than a factor model.

## 6. Conclusions

This paper studies the role of *non-pervasive* shocks when forecasting using factor models. To this end, we first augment the approximate dynamic factor model with a sparse model for the idiosyncratic component, thus enabling us to capture the effects of *non-pervasive* shocks. Then, we test the forecasting performance of this new model. The model is tested by means of both a simulation exercise and a pseudo real-time exercise performed on US quarterly data for six real variables: GDP, durables/non-durables/services consumption, and residential/non-residential investment.

The results show that it is not always necessary to account for *non-pervasive* shocks when forecasting with factor models. More specifically, in the simulation exercise we

16 In order to determine the number of common shocks in the augmented database, we used the criterion of Hallin and Liška (2007), which indicates the presence of four common shocks. The same criterion applied to the benchmark database indicates two common shocks, as per the Onatski (2009) test.

show that it clearly depends on the characteristics of the target variable whether our model leads to any forecasting gain: if a variable is strongly affected by *non-pervasive* shocks, then our model, which takes these kinds of shocks into account, is useful; otherwise, it is not. In other words, if a variable has a large, persistent idiosyncratic component, then it is worth forecasting it, otherwise it is unnecessary. The empirical analysis confirms this result: when forecasting GDP or consumption, i.e., aggregate variables which respond primarily to macroeconomic (*pervasive*) shocks, accounting for *non-pervasive* shocks is not useful. However, when forecasting residential investment, i.e., a variable that is also strongly affected by regional and sectorial shocks, capturing the dynamics generated by *non-pervasive* shocks is useful.

All in all, these results indicate that, although there is some cross-correlation due to *non-pervasive* shocks in the idiosyncratic component, accounting for this rarely boosts the forecasting accuracy. This indicates that the factor structure is particularly strong in the data, and that common factors capture not only the comovement, but also the bulk of the dynamics in the data.

## Acknowledgments

I am grateful to two anonymous referees and an associate editor for excellent, helpful comments. I am particularly indebted to Matteo Barigozzi, Domenico Giannone, and Marco Lippi for useful suggestions and comments. I have also benefited from discussions with Cecilia Frale and Libero Monteforte. This paper is a revised version of the fourth chapter of my Ph.D. thesis at the University of Rome, "La Sapienza". I wrote a large part of this paper while I was visiting the Office for Economic and Financial Analysis of Italy's Ministry of Economy and Finance, Treasury Department, whose hospitality is gratefully acknowledged. I also gratefully acknowledge financial support from the Belgian National Bank and the IAP P6/07 contract, from the IAP program (Belgian Scientific Policy), "Economic policy and finance in the global economy". I am *Chargé de recherches* F.R.S.–F.N.R.S., and gratefully acknowledge their financial support. Of course, any errors are my responsibility.

## References

Artis, M. J., Banerjee, A., & Marcellino, M. (2005). Factor forecasts for the UK. *Journal of Forecasting*, *24*, 279–298.
Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, *71*, 135–171.
Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, *70*, 191–221.
Bai, J., & Ng, S. (2008a). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, *146*, 304–317.
Bai, J., & Ng, S. (2008b). Large dimensional factor analysis. *Foundations and Trends in Econometrics*, *3*, 89–163.
Bai, J., & Ng, S. (2009). Boosting diffusion indices. *Journal of Applied Econometrics*, *24*, 607–629.
Bańbura, M., Giannone, D., & Reichlin, L. (2010). Large Bayesian vector autoregressions. *Journal of Applied Econometrics*, *25*, 71–92.
Bańbura, M., Giannone, D., & Reichlin, L. (2011). Nowcasting. In M. P. Clements, & D. F. Hendry (Eds.), *Oxford handbook on economic forecasting*. New York: Oxford University Press.
Bańbura, M., & Modugno, M. Maximum likelihood estimation of factor models on data sets with arbitrary pattern of missing data, *Journal of Applied Econometrics*, in press.
Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *Operations Research Quarterly*, *20*, 451–468.
Boivin, J., & Ng, S. (2005). Understanding and comparing factor-based forecasts. *International Journal of Central Banking*, *1*, 117–151.
Boivin, J., & Ng, S. (2006). Are more data always better for factor analysis? *Journal of Econometrics*, *127*, 169–194.
Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*, 123–140.
Bühlmann, P., & Yu, B. (2003). Boosting with the $L_2$ loss: regression and classification. *Journal of the American Statistical Association*, *98*, 324–339.
Bulligan, G., Golinelli, R., & Parigi, G. (2010). Forecasting monthly industrial production in real-time: from single equations to factor-based models. *Empirical Economics*, *39*, 303–336.
Camacho, M., & Perez-Quiros, G. (2010). Introducing the Euro-sting: short-term indicator of Euro area growth. *Journal of Applied Econometrics*, *25*, 663–694.
D'Agostino, A., & Giannone, D. (2012). Comparing alternative predictors based on large-panel factor models. *Oxford Bulletin of Economics and Statistics*, *74*, 306–326.
De Mol, C., Giannone, D., & Reichlin, L. (2008). Forecasting using a large number of predictors: is Bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics*, *146*, 318–328.
Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, *13*, 253–263.
Doz, C., Giannone, D., & Reichlin, L. (2012). A quasi maximum likelihood approach for large approximate dynamic factor models. *Review of Economics and Statistics*, *94*, 1014–1024.
Efron, B., Hastie, T., Johnstone, L., & Tibshirani, R. (2004). Least angle regression. *Annals of Statistics*, *32*, 407–499.
Eickmeier, S., & Ziegler, C. (2008). How successful are dynamic factor models at forecasting output and inflation? A meta-analytic approach. *Journal of Forecasting*, *27*, 237–265.
Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2000). The generalized dynamic factor model: identification and estimation. *The Review of Economics and Statistics*, *82*, 540–554.
Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2003). Do financial variables help forecasting inflation and real activity in the Euro area? *Journal of Monetary Economics*, *50*, 1243–1255.
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.
Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, *55*, 119–139.
Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. *Econometrica*, *74*, 1545–1578.
Groen, J. J., & Kapetanios, G. (2008). *Revisiting useful approaches to data-rich macroeconomic forecasting*. Working Papers 624, Queen Mary, University of London, School of Economics and Finance.
Hallin, M., & Liška, R. (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association*, *102*, 603–617.
Hallin, M., & Liška, R. (2011). Dynamic factors in the presence of blocks. *Journal of Econometrics*, *163*, 29–41.
Kose, A. M., Otrok, C., & Whiteman, C. H. (2008). Understanding the evolution of world business cycles. *Journal of International Economics*, *75*, 110–130.
Leamer, E. E. (1978). *Specification searches*. New York: Wiley.
Luciani, M. (2013). Monetary Policy and the housing market: a structural factor analysis. *Journal of Applied Econometrics* (in press).
Marcellino, M., Stock, J. H., & Watson, M. W. (2003). Macroeconomic forecasting in the Euro area: country specific versus area-wide information. *European Economic Review*, *47*, 1–18.
Moench, E., & Ng, S. (2011). A hierarchical factor analysis of U.S. housing market dynamics. *Econometrics Journal*, *14*, C1–C24.
Moench, E., Ng, S., & Potter, S. (2013). Dynamic hierarchical factor models. *Review of Economics and Statistics*, in press.
Onatski, A. (2012). Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics*, *168*, 244–258.
Onatski, A. (2009). Testing hypotheses about the number of factors in large factor models. *Econometrica*, *77*, 1447–1479.
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. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, *97*, 1167–1179.
Stock, J. H., & Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics*, *20*, 147–162.
Stock, J. H., & Watson, M. W. (2006). Forecasting with many predictors. In G. Elliott, C. Granger, & A. Timmermann (Eds.), *Handbook of economic forecasting*. Elsevier.

Stock, J. H., & Watson, M. W. (2009). The evolution of national and regional factors in U.S. housing construction. In T. Bollerslev, J. Russell, & M. Watson (Eds.), *Volatility and time series econometrics: essays in honour of Robert F. Engle*. Oxford: Oxford University Press.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B*, *58*, 267–288.

**Matteo Luciani** is FNRS Post-Doctoral Researcher (Chargé de recherches FNRS) at the European Center for Advanced Research in Economics and Statistics (ECARES), Université libre de Bruxelles, Belgium. He holds a Ph.D. in Economics from University of Rome "La Sapienza" (2010), Italy, and an MA in Economics from the University of Southern California (2005), Los Angeles, USA.